8. Baskıdan Çeviri

# Psikoloji İçin İstatistiksel Metotlar

David C. Howell

Çeviri Editörü:
Prof. Dr. Yaşar Baykul

PEGEM AKADEMİ

David C. Howell

Çeviri Editörü: Prof. Dr. Yaşar BAYKUL

# PSİKOLOJİ İÇİN İSTATİKSEL METOTLAR

ISBN 978-605-318-569-7
DOI 10.14527/9786053185697

Kitabın orijinal adı: Statistical Methods for Psychology
Baskı sayısı: 8. Baskı
ISBN: 9781111835484
WADSWORTH Cengage Learning

Kitap içeriğinin tüm sorumluluğu yazarlarına aittir.

© 2017, PEGEM AKADEMİ

Bu kitabın basım, yayım ve satış hakları Pegem Akademi Yay. Eğt. Dan. Hizm. Tic. Ltd. Şti.ye aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri, kapak tasarımı; mekanik, elektronik, fotokopi, manyetik, kayıt ya da başka yöntemlerle çoğaltılamaz, basılamaz, dağıtılamaz. Bu kitap T.C. Kültür Bakanlığı bandrolü ile satılmaktadır. Okuyucularımızın bandrolü olmayan kitaplar hakkında yayınevimize bilgi vermesini ve bandrolsüz yayınları satın almamasını diliyoruz.

Pegem Akademi Yayıncılık, 1998 yılından bugüne uluslararası düzeyde düzenli faaliyet yürüten **uluslararası akademik bir yayınevi**dir. Yayımladığı kitaplar; Yükseköğretim Kurulunca tanınan yükseköğretim kurumlarının kataloglarında yer almaktadır. Dünyadaki en büyük çevrimiçi kamu erişim kataloğu olan **WorldCat** ve ayrıca Türkiye'de kurulan **Turcademy.com** ve **Pegemindeks.net** tarafından yayınları taranmaktadır, indekslenmektedir. Aynı alanda farklı yazarlara ait 1000'in üzerinde yayını bulunmaktadır. Pegem Akademi Yayınları ile ilgili detaylı bilgilere http://pegem.net adresinden ulaşılabilmektedir.

1. Baskı: Ağustos 2017, Ankara

Yayın-Proje: Özlem Sağlam
Dizgi-Grafik Tasarım: Tuğba Kuşcuoğlu
Kapak Tasarımı: Pegem Akademi

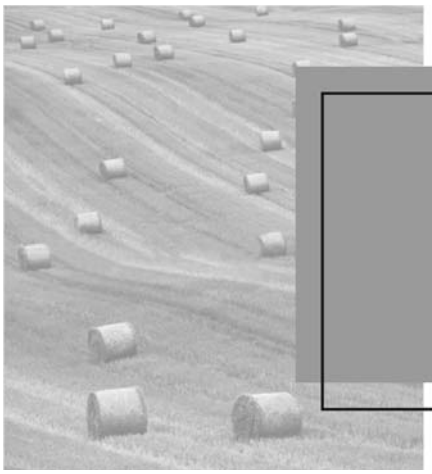
Vadi Grafik Tasarım ve Reklamcılık Ltd.Şti.
İvedik Osb Oto 3. Bölge 2284 Sokak No:101
Yenimahalle/Cebeci
0312 395 85 71-72

Yayıncı Sertifika No: 14749
Matbaa Sertifika No: 33748

**İletişim**

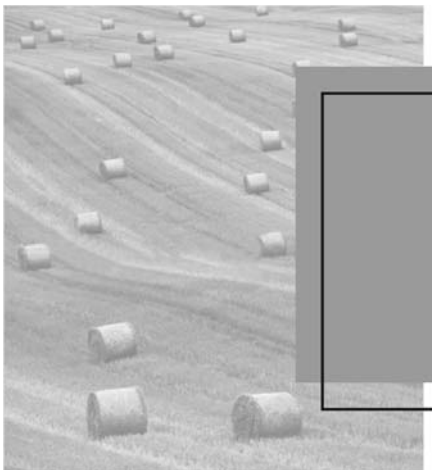
Karanfil 2 Sokak No: 45 Kızılay/ANKARA
Yayınevi: 0312 430 67 50 - 430 67 51
Yayınevi Belgeç: 0312 435 44 60
Dağıtım: 0312 434 54 24 - 434 54 08
Dağıtım Belgeç: 0312 431 37 38
Hazırlık Kursları: 0312 419 05 60
İnternet: www.pegem.net
E-ileti: pegem@pegem.net

***Donna'ya***



# Kısa İçerik


**Bölüm 1** Temel Kavramlar 1
**Bölüm 2** Verinin Tanımlanması ve Keşfedilmesi 15
**Bölüm 3** Normal Dağılım 63
**Bölüm 4** Örneklem Dağılımları ve Hipotez Testi 83
**Bölüm 5** Olasılığın Temel Kavramları 107
**Bölüm 6** Kategorik Veri ve Ki-Kare 137
**Bölüm 7** Ortalamalara Uygulanan Hipotez Testleri 177
**Bölüm 8** Güç 229
**Bölüm 9** Korelasyon ve Regresyon 251
**Bölüm 10** Diğer Korelasyon Teknikleri 303
**Bölüm 11** Basit Varyans Analizi 325
**Bölüm 12** İşlem Ortalamaları Arasında Çoklu Karşılaştırmalar 369
**Bölüm 13** Faktöriyel Varyans Analizi 411
**Bölüm 14** Tekrarlı-Ölçümler Desenleri 457
**Bölüm 15** Çoklu Regresyon 507
**Bölüm 16** Genel Doğrusal Modeller Olarak Varyans Analizi ve Kovaryans Analizi 573
**Bölüm 17** Meta Analizi ve Tek-Durum Desenleri 623
**Bölüm 18** Yeniden Örnekleme ve Parametrik Olmayan Testler 657



# İçindekiler

**Ön Söz** xv
**Çeviri Editörünün Ön Sözü** xxi
**Tercüme Edenler Listesi** xxv
**Yazar Hakkında** xvii

**BÖLÜM 1 Temel Kavramlar 1**

1.1 Önemli Terimler 2
1.2 Betimsel ve Vardamsal İstatistik 5
1.3 Ölçmede Ölçekler 6
1.4 Bilgisayar Kullanmak 8
1.5 Bu Basım Hakkında Bilinmesi Gerekenler 10

**BÖLÜM 2 Verinin Tanımlanması ve Keşfedilmesi 15**

2.1 Verinin Grafiğinin Çizilmesi 16
2.2 Histogramlar 18
2.3 Veri Kümesine Ait Grafiğin Sürekli Hale Getirilmesi 21
2.4 Gövde-ve-Yaprak Gösterimi 24
2.5 Dağılımların Tamamlanması 27
2.6 Sembol ile Gösterim (Notation) 30
2.7 Merkeze Eğilme Ölçüleri 32
2.8 Değişme (Yayılma) Ölçüleri 35
2.9 Kutu Grafikleri: Dağılımın ve Uç Değerlerin Grafik Gösterimleri 47
2.10 Değişme Ölçülerinin SPSS Yardımıyla Elde Edilmesi 51
2.11 Yüzde Birlikler, Çeyrekler ve Onda Birlikler 51
2.12 Doğrusal Dönüşümlerin Veri Üzerindeki Etkisi 52

**BÖLÜM 3 Normal Dağılım 63**

3.1 Normal Dağılım 66
3.2 Standart Normal Dağılım 69
3.3 Normal Dağılım Tablosunun Kullanılması 71
3.4 Bir Gözleme Ait Olası Sınırların Belirlenmesi 74
3.5 Veri Kümesinin Dağılımının Normal Olup Olmadığının Değerlendirilmesi 75
3.6 z ile İlgili Ölçüler 78

**BÖLÜM 4 Örneklem Dağılımı ve Hipotez Testi 83**

4.1 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dair İki Basit Örnek 84
4.2 Örneklem Dağılımları 86
4.3 Hipotez Test Etmenin Teorisi 88
4.4 Sıfır Hipotezi 90
4.5 Test İstatistikleri ve Bunların Örneklem Dağılımları 93
4.6 Sıfır Hipotezi Hakkında Karar Verme 93
4.7 I. ve II. Tip Hatalar 94
4.8 Tek ve Çift Yönlü Testler 97
4.9 Sıfır Hipotezini Reddetmek Ne Anlama Gelir? 99
4.10 Hipotez Test Etmeye Alternatif Bir Bakış 99
4.11 Etki Büyüklüğü 101
4.12 Son Bir İşlenmiş Örnek 102
4.13 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dönüş 103

**BÖLÜM 5 Olasılığın Temel Kavramları 107**

5.1 Olasılık 108
5.2 Temel Terimler ve Kurallar 110
5.3 Sürekli ve Kesikli Değişkenler 114
5.4 Kesikli Değişkenler İçin Olasılık Dağılımı 115
5.5 Sürekli Değişkenler İçin Olasılık Dağılımı 115
5.6 Permütasyonlar ve Kombinasyonlar 117
5.7 Bayes Teoremi 120
5.8 Binom Dağılımı 124
5.9 Hipotez Test Etmede Binom Dağılımının Kullanılması 129
5.10 Çok Terimli Dağılım 131

**BÖLÜM 6 Kategorik ve Ki-Kare 137**

6.1 Ki-Kare Dağılımı 138
6.2 Ki-Kare Uyum İyiliği Testi – Tek Yönlü Sınıflama 139
6.3 İki Sınıflamalı Değişken: Çapraz Tablo Analizleri 144

6.4 Ek Bir Örnek – 4 x 2'lik Bir Desen 148
6.5 Sıralama Düzeyindeki Veri Kümeleri İçin Ki-Kare 152
6.6 Ki-Kare Varsayımlarının Özeti 153
6.7 Bağımlı veya Tekrarlı Ölçümler 154
6.8 Tek ve İki Yönlü Testler 156
6.9 Olabilirlik Oranı Testleri 157
6.10 Mantel-Haenszel İstatistiği 158
6.11 Etki Büyüklükleri 160
6.12 Uyum Ölçüsü 166
6.13 Sonuçların Yazılması 167

**BÖLÜM 7 Ortalamalara Uygulanan Hipotez Testleri 177**

7.1 Ortalamanın Örneklem Dağılımı 178
7.2 Ortalamalar ile İlgili Hipotezleri Test Etme – $\sigma$ Bilindiğinde 181
7.3 $\sigma$ Bilinmediğinde Ortalamayı Test Etme - Tek Örneklem *t* Testi 183
7.4 Ortalamalara Uygulanan Hipotez Testleri – Eşleştirilmiş İki Örneklem 197
7.5 Ortalamalara Uygulanan Hipotez Testleri – İki Bağımsız Örneklem 206
7.6 Varyansın Heterojenliği: Behrens-Fisher Problemi 217
7.7 Hipotez Testlerinin Tekrar Gözden Geçirilmesi 220

**BÖLÜM 8 Güç 229**

8.1 Temel Güç Kavramı 231
8.2 Bir Testin Gücünü Etkileyen Faktörler 232
8.3 Gücün Geleneksel Yolla Hesaplanması 234
8.4 Tek-Örneklem t İçin Güç Hesaplamaları 236
8.5 Birbirinden Bağımsız İki Ortalama Arasındaki Fark İçin Güç Hesaplamaları 238
8.6 Eşleştirilmiş Örneklemler t Testi İçin Güç Hesaplamaları 241
8.7 Güç Üzerine Olan Durumu Tersine Çevirme 242
8.8 Daha Kompleks Desenlerde Güç Düşünceleri 243
8.9 Hesaplamaları Basitleştirmek İçin G*Power'ın Kullanılması 243
8.10 Geriye Dönük Güç 245
8.11 Güç Analizinin Sonuçlarını Raporlama 247

**BÖLÜM 9 Korelasyon ve Regresyon 251**

9.1 Dağılımın Grafiği 253
9.2 Yaşam Temposu ile Kalp Hastalığı Arasındaki İlişki 255
9.3 Stres ve Sağlık Arasındaki İlişki 257
9.4 Kovaryans 258
9.5 Pearson Momentler Çarpımı Korelasyon Katsayısı (*r*) 260

9.6 Regresyon Doğrusu 261
9.7 Veriye Uygun Bir Doğrunun Çizilmesinin Diğer Yolları 266
9.8 Yordamanın Doğruluğu 266
9.9 Regresyon ve Korelasyon Altındaki Varsayımlar 272
9.10 Ŷ'ya Ait Güven Sınırları 274
9.11 Test Alma Becerilerinin Rolünü Gösteren Bir Bilgisayar Örneği 277
9.12 Hipotez Testi 280
9.13 Son Bir Örnek 288
9.14 Korelasyon ve Regresyonda Varsayımların Rolü 290
9.15 Korelasyonu Etkileyen Faktörler 291
9.16 Pearson'ın r'si İçin Güç Hesaplaması 293

**BÖLÜM 10 Diğer Korelasyon Teknikleri 303**

10.1 Nokta-Çiftserili Korelasyon ve Fi (Phi): Pearson Korelasyonlarının Diğer Adı 304
10.2 Çift Serili ve Tetrakorik Korelasyon: Pearson-Dışı Korelasyon Katsayıları 313
10.3 Sıralı Veri Kümeleri İçin Korelasyon Katsayıları 313
10.4 Sıralı Veriden Oluşan Çapraz Tabloların Analizi 317
10.5 Kendall'ın Konkordans Katsayısı (*W*) 320

**BÖLÜM 11 Basit Varyans Analizi 325**

11.1 Bir Örnek 326
11.2 Altta Yatan Model 327
11.3 Varyans Analizinin Mantığı 329
11.4 Varyans Analizinde Hesaplamalar 332
11.5 Sonuçların Raporlanması 338
11.6 Bilgisayar Çözümleri 339
11.7 Eşit Olmayan Örneklem Büyüklükleri 341
11.8 Varsayımlara Uyulmaması 343
11.9 Dönüştürmeler 346
11.10 Sabite Karşı Tesadüfî Modeller 353
11.11 Deneysel Etkinin Büyüklüğü 353
11.12 Güç 357
11.13 Bilgisayar Analizleri 361

**BÖLÜM 12 İşlem Ortalamaları Arasında Çoklu Karşılaştırmalar 369**

12.1 Hata Oranları 370
12.2 Morfin Toleransı Üzerine Yürütülen Basit Bir Deneydeki Çoklu Karşılaştırmalar 373
12.3 Önsel Karşılaştırmalar 376
12.4 Karşılaştırmalar İçin Güven Aralıkları ve Etki Büyüklükleri 388

12.5 Sonuçların Rapor Edilmesi 391
12.6 Planlı-Çoklu Doğrusal Karşılaştırmalar 391
12.7 Tukey Testi 393
12.8 Hangi Test? 398
12.9 Bilgisayar Çözümlemeleri 398
12.10 Eğilim Analizi 401

**BÖLÜM 13 Faktöriyel Varyans Analizi 411**

13.1 Eysenck'ın Çalışmasının Genişletilmesi 414
13.2 Yapısal Model ve Beklenen Kareler Ortalamaları 418
13.3 Etkileşimler 419
13.4 Basit Etkiler 420
13.5 Varyans Analizinin Sigara İçmenin Etkilerine Uygulanması 423
13.6 Ortalamalar Arası Karşılaştırmalar 426
13.7 Faktöriyel Deneyler İçin Güç Analizi 427
13.8 Alternatif Deneysel Desenler 430
13.9 İlişki ve Etki Büyüklüğü Ölçümleri 437
13.10 Sonuçların Rapor Edilmesi 443
13.11 Eşit Olmayan Örneklem Büyüklükleri 444
13.12 Yüksek Mertebeden Faktöriyel Desenler 446
13.13 Bir Bilgisayar Örneği 451

**BÖLÜM 14 Tekrarlı-Ölçümler Desenleri 457**

14.1 Yapısal Model 460
14.2 *F* Oranları 460
14.3 Kovaryans Matrisi 461
14.4 Varyans Analizinin Rahatlama Terapisine Uygulanması 462
14.5 Tekrarlı-Ölçümler Desenlerinde Karşılaştırmalar ve Etki Büyüklüğü 465
14.6 Sonuçların Raporlanması 466
14.7 Bir Denekler-Arası Değişken ve Bir Denekler-İçi Değişken 467
14.8 İki Denekler-Arası Değişken ve Bir Denekler-İçi Değişken 478
14.9 İki Denekler-İçi Değişken ve Bir Denekler-Arası Değişken 484
14.10 Sınıflar İçi Korelasyon 489
14.11 Tekrarlı Ölçüm Analizlerine Dair Diğer Düşünceler 491
14.12 Tekrarlı Ölçüm Desenleri İçin Karma Modeller 492

**BÖLÜM 15 Çoklu Regresyon 507**

15.1 Çoklu Doğrusal Regresyon 508
15.2 Ek Yordayıcıların Kullanılması 519

15.3 Regresyon Katsayılarının Standart Hatası ve Test Edilmesi 521
15.4 Yeniden Örnekleme Yaklaşımı 522
15.5 Artık Varyans 524
15.6 Dağılım Varsayımları 524
15.7 Çoklu Korelasyon Katsayısı 525
15.8 Kısmi ve Yarı Kısmi Korelasyon 527
15.9 Bastırıcı Değişkenler 531
15.10 Regresyona Hazırlık 532
15.11 Regresyon Eşitliğinin Oluşturulması 539
15.12 Her Bir Değişkenin "Önemi" 543
15.13 Yaklaşık Regresyon Katsayılarının Kullanılması 545
15.14 Aracı Değişken ve Şartlı Değişken İlişkileri 546
15.15 Lojistik Regresyon 556

**BÖLÜM 16 Genel Doğrusal Modeller Olarak Varyans Analizi ve Kovaryans Analizi 573**

16.1 Genel Doğrusal Model 574
16.2 Tek-Yönlü Varyans Analizi 577
16.3 Faktöriyel Desenler 580
16.4 Eşit Olmayan Örneklem Büyüklükleri ile Varyans Analizi 587
16.5 Tek Yönlü Kovaryans Analizi 594
16.6 Kovaryans Analizinde Etki Büyüklüğünün Hesaplanması 604
16.7 Kovaryans Analizi Sonuçlarının Yorumlanması 606
16.8 Kovaryans Analizi Sonuçlarının Raporlaştırılması 607
16.9 Faktöriyel Kovaryans Analizi 607
16.10 Çoklu Ortak Değişkenlerin Kullanılması 615
16.11 Alternatif Deneysel Desenler 616

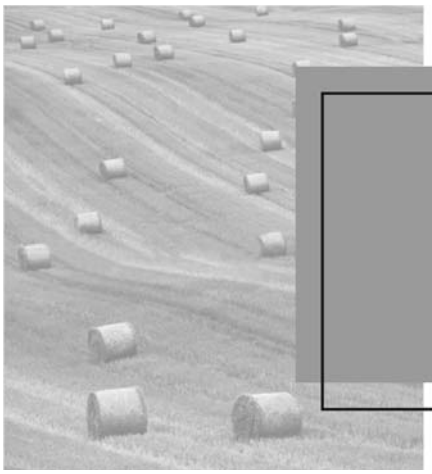
**BÖLÜM 17 Meta Analizi ve Tek-Durum Desenleri 623**

Meta-Analiz 624
17.1 Etki Büyüklüğü Ölçülerinin Kısa Özeti 625
17.2 Bir Örnek - Çocuk ve Ergenlik Depresyonu 628
17.3 İkinci Bir Örnek - Nikotin Sakızı ve Sigarayı Bırakma 638
Tek Durum Desenleri 641
17.4 Standartlaştırılmış Ortalama Farkların Araştırıldığı Analizler 641
17.5 Depresyon İçin Bir Durum Çalışması 642
17.6 Tek-Durum Desenine İkinci Bir Yaklaşım - Parça Parça Regresyon Kullanımı 646

## BÖLÜM 18 Yeniden Örnekleme ve Parametrik Olmayan Testler 657

18.1 Genel Bir Yaklaşım Olarak Özyükleme 659
18.2 Bir Örneklem ile Özyükleme 661
18.3 Bir Korelasyon Katsayısında Özyükleme Güvenlik Sınırları 662
18.4 Eşleştirilmiş İki Örneklem ile Yeniden Örnekleme Testleri 665
18.5 Bağımsız İki Örneklem ile Yeniden Örnekleme Testleri 667
18.6 Wilcoxon Sıra-Toplamları Testi 668
18.7 Wilcoxon Eşleştirilmiş-Çiftler İşaretli-Sıralar Testi 673
18.8 İşaret Testi 677
18.9 Kruskall-Wallis Tek-Yönlü Varyans Analizi 678
18.10 *k* Sayıda İlişkili Örneklem İçin Friedman Sıra Testi 679

**Ekler 685**
**Kaynakça 719**
**Alıştırmalara Cevaplar 733**
**Dizin 757**
**Semboller Sözlüğü 771**
**Tercümede Türkçeleştirilen Sembol ve Kısaltmalar 775**



# Ön Söz

Psikoloji İçin İstatistiksel Metotlar eserinin sekizinci basımı da, daha önceki basımlarda olduğu gibi davranışsal ve sosyal bilimlerde, özellikle de psikoloji ve eğitim alanlarında yaygın biçimde kullanılan istatistiksel teknikleri incelemektedir. Kitap, üst seviyedeki lisans ve yüksek lisans öğrencileri için tasarlanmış olsa da, öğrencilerin daha önce bir istatistik dersi veya lise matematiğinin ötesinde bir matematik dersi aldıklarını varsaymaz. Daha önce giriş niteliğinde bir ders almış olan öğrenciler, önceden gördükleri materyalin faydalı bir gözden geçirme olduğunu göreceklerdir. Bu kitap, bir dönemlik veya tüm yılı kapsayan bir ders için uygundur; ben de kitabı hem bir dönemlik hem de bir yıllık derslerde başarılı bir şekilde kullandım. Öğrencilerin ve aynı zamanda fakültenin, istatistiksel bir problemle karşılaştıklarında istatistiği ilk öğrendiği kitaba sıkça başvurduğunu gördüm, bu yüzden de kitaba, onu gelecekteki kullanımlar için faydalı bir referans kitabı yapacak materyalleri ekledim. Bu materyali dersine dâhil etmemeyi isteyen bir öğretim görevlisi bununla ilgili bir zorluk yaşamayacaktır. Bununla beraber, yalnızca faydalı olması muhtemel materyali dâhil edebilmek için yukarıda bahsettiğim materyalde bir miktar kesinti yaptım. İlginç olan her fikri kitaba dâhil etme düşüncesi, göz korkutucu olmaya başlayan bir kitaba sebep olabilirdi.

Bazı yönlerden bu basım, geçmiş basımlardan sonra bir kırılmayı temsil etmektedir. Yıllar içinde her basım belli ölçüde yeni materyal içeriyordu ve yazı, teori ve pratikteki değişiklikleri yansıtmak için değiştiriliyordu. Bu basım da aynı yolu izlemektedir fakat değişikliklerden bazıları normalden daha kapsamlıdır. Bu basımda, öğrencilerin anlaması gereken belli sayıdaki ek konudan bahsetmeyi düşündüm ve bunu da, bazı eski materyalleri kaldırmadan yapamazdım. Buna karşılık internet sayesinde materyal silinmeyecek, yalnızca yer değiştirecek. Çoklu karşılaştırma testini veya log-doğrusal modelleri içeren bölümün eski ve tam bir versiyonunu görmek isterseniz benim web sitemden eski versiyonları indirebilirsiniz.

Bu kitabın ilk basımı 1973-1974 yıllarında yazılmıştı; o sıralarda ben İngiltere'deki Durham Üniversitesi'nde uzun dönemli akademik izindeydim. O dönemde, psikoloji dalındaki istatistik derslerinin çoğu deneysel psikologları hedefliyordu ve bu çalışmaların büyük bir yüzdesi de varyans analizi ile ilişkiliydi. B. J. Winer tarafından yazılan Deneysel Tasarımda İstatistik İlkeleri (1962, 1971) adlı eser, istatistik alanının büyük bir kısmına hâkim oldu ve deneylerin tasarlanma, yürütülme ve analiz edilme şekillerini etkiledi. Benim kitabım büyük ölçüde bu trendi takip etti ve bunu da hakkıyla yaptığını düşünüyo-

rum. Verilerin istatistiksel analizi o zamandan beri oldukça fazla değişikliğe tanık oldu. Varyans analizi hâlâ hâkim teknik olsa da ve çoklu regresyon onu yakından takip etse de, birçok başka konu da listeye girmiştir. Örneğin etki büyüklüğü ölçüleri şu anda, o zaman olduğundan çok daha önemli durumdadır. Meta çözümlemesi önceleri adeta hiçbir şey iken, bir alanı özetlemenin önemli bir yolu haline gelmiştir. Kayıp verilere yapılacak işlemler de, lineer karma modeller ve veri değerlendirme yöntemleri sayesinde önemli ölçüde değişmiştir. Yüksek kalitede araştırma yapmak isteyen öğrencilere bu konuları anlatmadan onları çalışmaya başlatamayız.

Fakat eğer metin yönetilebilir bir büyüklükte kalacaksa, bu konuları ve buna benzer konuları kitaba dâhil etmek için bazı şeylerin de gitmesi gerekiyor. Geçmişte, örneğin çoklu karşılaştırma yöntemleri, birbiriyle rekabet halinde olan birçok yaklaşımın kayda değer dikkat gösterdiği bir konu olmuştu. Fakat eğer insanların literatürde gerçekten ne yaptıklarına bakarsanız, birbiriyle rekabet halindeki yaklaşımların çoğu uygulayıcılardan çok teorisyenlerin ilgisini çekmektedir. Benim düşünceme göre kayıp verilere yapılacak işlemler hakkında yazmak, ortalamalar arasında karşılaştırma yapmak konusundaki tüm alternatifleri listelemekten çok daha faydalıdır. Bu durum özellikle de, eski materyali okumak isteyen herkes için onu kolayca bulunabilir hale getirmek kolay olduğu için doğrudur.

Bu kitabı yazmaktaki amacım materyali sezgisel bir seviyede açıklamaktı. Bu durum, materyalin "sulandırıldığı" şeklinde anlaşılmamalı; vurgunun kavramsal anlayışta olduğu şeklinde anlaşılmalı. Örneğin, t'nin örneklem dağılımını başarılı bir şekilde çıkarabilen bir öğrenci, o dağılımın nasıl kullanılacağını hiç anlamamış olabilir.

Bu örnekle ilgili olarak amacım, örnekleme dağılımının anlamına odaklanmak ve onun hipotez testinin genel teorisinde oynadığı rolü göstermekti. Benim düşünceme göre bu yaklaşım öğrencilerin, belli bir testin çalışma şekline ve testlerin arasındaki karşılıklı ilişkiye dair teknik bir yaklaşımın sağlayacağından daha iyi bir anlayış edinmelerini sağlamaktadır.

Popüler görüşün aksine istatistiksel yöntemler sürekli olarak evrim geçirmektedir. Bunun nedeni kısmen davranışsal bilimlerin birçok yeni kola ayrılması ve kısmen de, verilerimiz hakkında sorular sormanın daha iyi yollarını bulmamızdır. Hiçbir kitap, işlenilmesi gereken tüm materyali içeremez fakat ihtiyaç duydukları zaman o materyali almaları için öğrencileri ve profesyonelleri hazırlamak önemlidir. Örneğin çok düzeyli/hiyerarşik modeller, araştırma literatüründe çok daha yaygın hale gelmektedir. Bu modellerin anlaşılması uzmanlaşmış metinler gerektirir fakat söz konusu literatürün içinden seçim yapmaya başlamak bile, rastgele değişkenlere karşı sabit değişkenlerin ve iç içe tasarımların anlaşılmasını gerektirir. Bu kitap birincisini, yani gerekli modelleri çıkarmayı üstlenemez fakat hem sabit hem de rastgele tasarımların ve iç içe girmenin altyapısını yaparak ikincisine değinebilir ve değinmektedir. Başka konular için de benzer altyapılar oluşturmaya çalıştım. Bunu, örneğin daha modern grafik cihazları ve yeniden örnekleme istatistikleri için yaptım ve bunu yaparken de okuyucuyu bataklığın içine çekmedim. Benim sorumluluğum bir açıdan, nereye gittiğimizi görmeye çalışmak ve okuyucuya o yönde gitmek konusunda bir temel sağlamaktır.

## 8. Basımdaki Değişiklikler

Bu sekizinci basım, kitabı öğrenci için daha cazip ve yöntembilim ve veri analizinin gerçek sürecine daha uygun hale getiren yeni veya genişletilmiş özellikleri içermektedir:

- Amerikan Psikoloji Derneği komitesi tarafından sıfır hipotezinin testi konusunda karşılaşılan meseleye cevap vermeye devam ettim ayrıca etki büyüklüğü ve etkinin önemi konusunda da daha fazla içeriği kitaba dâhil ettim. Bu basımdaki kapsam, önceki basımlardaki kapsamın ötesine geçmektedir ve materyale detaylı bir giriş görevi görmektedir.
- Yedinci basımda, aralarında olasılık çizişlerinin, çekirdek yoğunluk çizişlerinin ve artık çizişlerinin de olduğu grafiksel gösterimler konusunda yeni materyal eklemiştim. Bunların her biri, verilerimizi daha iyi anlamamıza ve yaptığımız varsayımların ne kadar mantıklı olduğunu değerlendirmemize yardımcı olur. Bu basımda grafiksel gösterimlerin kullanımını arttırdım.
- Daha geleneksel yaklaşımların, temeldeki varsayımlara dayanarak neler yapmaya çalıştığını göstermek için yeniden örnekleme kavramını daha da geliştirdim. Yeniden örneklemenin yönündeki değişme hızlanmıştır ve birçok bilgisayar çözümü "önyükleme", "benzetme" ve "yeniden örnekleme" gibi kullanıcı seçenekleri sunmaktadır ki bunların hepsi de o yaklaşıma dayanır.
- Cochran-Mantel-Haenszel olumsallık çizelgeleri analizinin içeriği, Berkeley yüksek lisans kabul verilerine de uygulandığı gibi Simpson Paradoksu klasik örneği ile ilişkilidir. Bu konu öğrencilerin, verilerinin ne anlama geldiğini derin bir şekilde düşünmelerini sağlayacak temel motivasyonla ilişkilidir.
- Çoklu karşılaştırma tekniklerini anlatan Bölüm 12'yi biraz daha değiştirerek, daha önce bahsetmiş olduğum testlerin geniş kapsamını daralttım ve Benjamini ve Hochberg'in Yanlış Keşif Oranı konusunu ekledim. Dikkatimizi, hata oranlarıyla ilgili çıkarım türünden yanlış keşif oranına kaydırdıkça analizlerimizin gücünü, Tip I hata açısından görece küçük bir maliyetle arttırırız.
- Tekrarlayan ölçümlerde varyans analizi konusundaki bölümün bir kısmı, karma modeller konusunu genişletmektedir. Bu yaklaşım, kayıp verilere yapılacak işlemin çok daha iyi olmasını sağlar ve bileşik simetri konusundaki mantıksız varsayımları zayıflatır. Bu da, bir anda tamamen yeni bir alana girmeye çalışmadan karma modellere yapılan bir giriş görevi görür.
- Eski basımların birkaç tanesi, log-doğrusal modeller konusundaki bir bölümü içeriyordu. Bu bölümün yerine, hem meta çözümlemesi hem de tek denekli tasarımlar konusunu içeren materyali dâhil ettim. Meta çözümlemelerinin davranışsal bilimlerdeki önemi artmıştır ve bunun temelinde de "belirti tabanlı tıp" konusuna yapılan vurgu vardır. Bir metin sınırlı ölçüde materyali kapsayabildiği için bu iki konuya yapılan girişin, log-doğrusal modelleri içeren bir bölümden daha değerli olduğunu düşünüyorum. Fakat ikinci konuyu kaçıranlar için bölüm, internetteki ek materyal kısmında halen mevcuttur.
- R adı verilen serbest işlem ortamı, ticari paketi S-PLUS ile birlikte istatistiksel analiz üzerindeki etkisini her ay arttırmaktadır. R'yi her şeyin üzerinde tutarak onunla ilgili şeyler öğretmeye çalışmak gibi bir niyetim yok fakat kitabın web

sitesinde R kodunun çok miktarda programını veya seçimini sunuyorum. Bu da, R'yi kullanmakla ilgilenenler için bir başlangıç noktası oluşturur ve metinde geçen noktaların ilginç gösterimlerinin yapılmasını sağlar.

- Öğretim görevlileri ve öğrenciler için materyal toparlamak ve bu materyali internetteki web sayfalarına yerleştirmek için çok zaman harcadım. Kullanıcılar, ilave (ve karmaşık) örneklere, metinde yer almayan konuların içeriklerine, internetteki diğer kaynaklara, bir sınıf veya laboratuar için uygun olacak gösterimlere vs. kolayca erişebilir. Eğer öğrenci bir konuyu daha derinlemesine incelemek isterse kitaptaki birçok bölüm özellikle bu materyale atıfta bulunmaktadır. Tüm bunlar, internet bağlantısı olan herkesin kolaylıkla erişebileceği şeylerdir. Bu materyale ekleme yapmaya da devam ediyorum; insanları da onu kullanmaları ve eleştirmeleri için teşvik ediyorum. Kitabın web sitesi http://www.uvm.edu/~dhowell/methods8/ adresidir. Daha kapsamlı bir içeriğe sahip olan ilave bir web sitesi de http://www.uvm.edu/~dhowell/StatPages/StatHomePage.html adresidir (bu adres büyük-küçük harfe duyarlıdır). Kullanıcıların her iki siteyi de keşfetmelerini öneririm.

Bu basım önceki basımlarla, kitabın ana içeriği olan istatistiksel hipotez testlerinden az veya çok bağımsız olan iki temel temayı paylaşmaktadır.

- İlk tema, bir hipotez testi ile olaya ani bir giriş yapmadan önce verilere bakmanın önemidir. Bunu bir kenarda tutarak, verilerin çiziminin yapılması, aykırı değerlerin aranması ve varsayımların kontrol edilmesi konularından detaylı olarak bahsettim. (Grafik gösterimler geniş ölçüde kullanılmıştır.) Veri setinin, gelişmiş bir istatistiksel tekniğin örneği olması amaçlansa da bu detaylı ifadeleri her bir veri setini sunduğum noktada kullanmaya çalıştım.
- İkinci tema ise, kullanılacak istatistik test ile deney tarafından ortaya konulan teorik sorular arasındaki ilişkinin önemidir. Bu ilişkiyi vurgulamak için gerçek örnekler kullanırım böylece öğrencinin deneyin gerçek amacını ve teori tarafından yapılan tahminleri anlamasını sağlamaya çalışırım. Bu sebeple bazen tüm bir kısmın ve hatta tüm bir kitap bölümünün odağı olarak tek bir temel örnek kullanırım. Örneğin Kaufman ve Rock (1962) tarafından yapılmış olan ay yanılsamasıyla ilgili çalışmanın ilginç verileri t testinin çeşitli şekillerinde kullanılmıştır (sayfalar 191-213). Bölüm 12'nin büyük bir kısmı ise Siegel (1975) tarafından morfin bağımlılığı konusundaki önemli bir çalışmanın etrafında düzenlenmiştir. Tüm bu örneklerin her biri öğrencileri direkt olarak ilgilendirmektedir. Bu basımda etki büyüklüklerine daha fazla vurgunun yapılması kişinin, verilerini ve sorunsallarını dikkatli bir şekilde düşünmesini gerektiren noktayı benimsemesini sağlar.

Kimsenin bu kitabı tartışmalı olarak nitelemeyeceği büyük bir olasılık olsa da birtakım tartışmalı konularda görüşlerimi ifade etmenin önemli olduğunu düşündüm. Sonuçta, istatistik alanında yapılan tartışmalar onu ilginç yapan şeyin bir bölümüdür. Örneğin temel ölçüm yapma ölçeğinin, bazı insanların iddia ettiği kadar önemli olmadığı görüşünü savundum. Ayrıca, eşit olmayan grup büyüklüklerine sahip varyans analizlerine belli bir türde işlem yapılmasını savundum (diğer türlü yapmak için zorlayıcı bir sebep olmadığı sürece). Her bir öğretim görevlisinin benimle aynı fikirde olmasını beklemiyorum; aslında bazılarının benimle aynı fikirde olmamasını umuyorum. Bu durum, öğrencilere karşıt görüşleri verme ve onların konuları anlamalarına yardımcı olma fırsatını sunar.

Öğrenciye çeşitli farklı çoklu karşılaştırma yöntemlerini sunduktan sonra (ki ben öyle yapıyorum), öğrencinin problemi için hangi yöntemin en iyi yöntem olduğu konusunda ona herhangi bir tavsiye vermeden öğrenciyi bırakmanın adil olmayan ve faydasız bir durum olduğunu düşünüyorum. Bahsettiğim yöntemlerin sayısını azaltmış olsam da öğrenciye, bir yaklaşımı seçerken nelerin gerektiği konusunu açıklamak için her türlü girişimi yaparım.

Öğrenciler için bir Çözüm Kitabı mevcuttur. Bu kitapta, kitabın web sitesinde bulunan tek sayılı alıştırmaların ayrıntılı çözümleri vardır. Ayrıca yayıncıdan gelen bir Öğretmen Kitabı da mevcuttur ve bu kitapta da tüm problemlerin çözümü vardır.

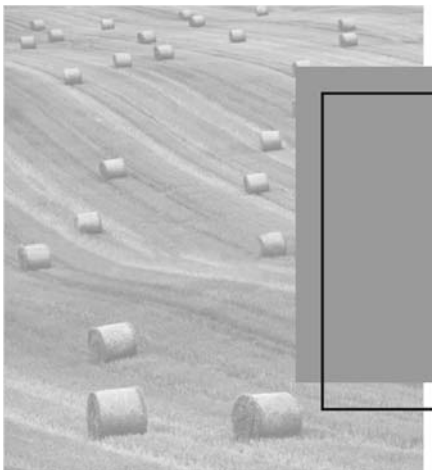
## Teşekkür

Taslak metni okuyup değerli geri bildirimlerini sunan aşağıdaki eleştirmenlere teşekkürlerimi sunarım: Angus MacDonald, University of Minnesota; William Smith, California State University–Fullerton; Carl Scott, University of St. Thomas–Houston; Jamison Fargo, Utah State University; Susan Cashin, University of Wisconsin–Milwaukee ve birçok basımda değerli rehberlik sunan Karl Wuensch, East Carolina University. Önceki basımlarda yararlı yorumlar ve öneriler aldığım insanlar ise: Kenneth J. Berry, Colorado State University; Tim Bockes, Nazareth College; Richard Lehman, Franklin and Marshall College; Tim Robinson, Virginia Tech; Paul R. Shirley, University of California–Irvine; Mathew Spackman, Brigham Young University; Mary Uley, Lindenwood University ve Christy Witt, Louisiana State University. Bu insanların katkıları bu basımda da belirgindir.

Bu basım için teşekkür etmek istediğim kişiler ise: Deborah A. Carroll, Southern Connecticut University; Paul Chang, Edith Cowen University; Ann Huffman, Northern Arizona University; Samuel Moulton, Harvard University; Therese Pigott, Loyola University Chicago; Lucy Troup, Colorado State University ve Meng-Jia Wu, Loyola University Chicago.

Tüm süreç boyunca yayın ekibi de çok yardımcı oldu. Bu konuda teşekkür etmek istediğim insanlar: Jessica Egbert, Pazarlama Müdürü; Vernon Boes ve Pamela Galbreath, Sanat Yönetmenleri; Mary Noel, Medya Editörü; Kristin Ruscetta, Yapım Proje Müdürü; Lauren Moody, Editör Asistanı ve Timothy Matray, Editör.

David C. Howell
Onursal Profesör
Vermont Üniversitesi
5/6/2011



# Çeviri Editörünün Ön Sözü

Elinizdeki kitap, Üniversity of Vermont'tan emekli Prof. Dr. Davit C. Howel tarafından yazılan "Statistical Methods for Psychology" adlı kitabın 8. baskısının Türkçeye tercümesidir. Kitabın aslı; 684 sayfası metin, 33 sayfası ekler ve geri kalan kısmı indeks, referanslar ve diğer kısımlar olmak üzeri 795 sayfadır. Kitap, üst düzeydeki lisans, yüksek lisans ve doktora öğrencileri için tasarlanmış olmakla birlikte, kitapta yer alan konuların anlaşılabilmesi için daha önce bir istatistik dersi ve lise matematiğinin ötesinde bir matematik dersi almış olunması gerekmez. Ayrıca kitap, davranış bilimlerinde ve sosyal bilimlerde, özellikle psikoloji ve eğitimde yaygın olarak kullanılan istatistiksel tenkitleri incelemektedir. Ayrcıa, güç, tekrar örnekleme (resampling), özyükleme (bootstraping) ve hiyerarşik modeller gibi son yıllarda araştırma literatüründe yaygın hale gelen konuları da içermektedir. Bu nedenlerle eldeki kitap, öğretimde ve araştırmalarda kullanılabilecek olduğu kadar davranış bilimlerinde ve sosyal bilimlerde gerekli olan istatistikle ilgili konularda bir referans kitabı olma özelliği taşımaktadır.

Kitabın metin kısmında bölümlere göre aşağıdaki kavramlar ve teknikler yer almaktadır: 1. Bölüm'de istatistikle ilgili temel kavramlar; 2. Bölüm'de verinin tanımlanması düzenlenmesi, grafikler, merkeze eğilme ve dağılma ölçüleri, doğrusal dönüştürmeler; 3. Bölüm'de normal dağılım, 4. Bölüm'de hipotez testi ile ilgli temel kavramlar, örneklem dağılımları, hipotez testine alternatif bir bakış ve etki büyüklüğü; 5. Bölüm'de olasılık ve ilgili temel kavramlar, değişken ve türleri, permütasyon ve kombinasyonlar, Bayes Teoremi, Binom ve çok değişkenli Binom dağılımları ve hipotez test etmede kullanılmaları; 6. Bölüm'de kategorik veri, ki-kare dağılım, hata oranları, basit deneysel çoklu karşılaştırma testlerinde kullanılması, Mantel-Haenszel testi, 7. Bölüm'de ortalamalara uygulanan hipotest testleri; 8. Bölüm'de güç kavramı ve gücü etkileyen faktörler, çeşitli hipotez testlerinde gücün hesaplanması, güç hesaplamalarının sadeleştirilimesinde G*Power'ın kullanılması; 9. Bölüm'de kovaryans, korelasyon, regresyon ve ilgili kavramlar, regrasyonla ilgili varsayımlar ve ilgili hipotez testleri, Pearson *r* korelasyonunun gücü; 10. Bölüm'de alternatif korelasyon teknikleri, 11. Bölüm'de basit varyans analizi, hesapların yapılması ve bilgisayar çözümleri, varsayımları ve varsayımların ihlâli durumları, transformasyonlar, deneysel etkinin büyüklüğü, güç ve bilgisayar çözümleri; 12. Bölüm'de hata oranları basit deneylerde çoklu karşılaştırmalar, güven aralıkları ve karşılaştırmaların etki büyükleri, tekrar örnekleme yaklaşımı, eğilim analizi; 13. Bölüm'de yapısal model ve beklenen kareler ortalaması, etkileşim, basit etkiler, ortalamalar arasında karşılaştırmalar, faktöriyel desenler ve güç analizleri, yüksek mertebeden faktöriyel desenler, 14. Bölüm'de tekrarlı ölçümler için yapısal model, kovaryans analizi, tekrarlı ölçüm desenlerinde karşılaştırmalar ve etki

büyüklükleri, bir ve iki değişkenli denekler-içi ve denekler-arası tekrarlı ölçüm desenleri, sınıflar-içi korelasyon, tekrarlı-ölçümler için karma modeller; 15. Bölüm'de çoklu doğrusal regresyon, regresyon katsayılarının test edilmesi, tekrar örnekleme yaklaşımı, dağılım varsayımları, çoklu kısmî ve yarı kısmî korelasyon katsayıları, baskıcı değişkenler, regresyon denkleminin oluşturulması, lojistik regresyon; 16. Bölüm'de genel doğrusal model, faktöriyel desenler, kovaryans ve faktöriyel kovaryans analizleri, kovaryans analizinde etki büyüklüğü, alternatif deneysel desenler; 17. Bölüm'de meta-analizi ve ilgili temel kavramlar, tek durum desenleri ve ilgili temel kavramlar, standartlaştırılmış ortalama farkların analizi, tek-durum desenleri ikinci bir yaklaşım; 18. Bölüm'de genel bir yaklaşım olarak özyükleme, tek örneklem ile özyükleme korelasyon katsayısı üzerine özyükleme, güven sınırları, eşleştirilmiş iki örneklem ve ile yeniden örnekleme, bağımsız iki örneklem ile yeniden örnekleme testleri, Wilcoxon sıra-toplamları testi, Wilcoxon eşleştirilmiş-çiftler işaretli- sıralar testi, işaret testi, Kruskal Wallis tek yönlü varyans analizi.

Kitapta, her bölümün sonunda çok sayıda sorudan oluşan alıştırmalara yer verilmiştir. Her bölüme ait alıştırmalardan tek numaralı olanların cevapları kitabın sonuna konulmuş (s. 733-755), çift numaralı olanların cevapları okuyucuya bırakılmıştır.

Kitaba ait bir Web sayfası bulunmaktadır, bu sayfada yardımcı materyallere yer verilmiştir. Bu materyaller şunlardır: R programı ve çalıştırması (yazar tarafından yazılmış), kitap içindeki örnekler için R programları. 2. Bölüm için: Çekirdek yoğunluk fonksiyonları için R ve SPSS kullanılarak hesaplama ve grafik çizme örnekleri. 3. Bölüm için: Normal dağılım için olasılık hesaplayıcısı, iPhone için normal dağılım hesaplayıcısı, Q-Q grafikleri üzerine kitaptakilerin dışında açıklamalar. 4. Bölüm için: Örneklem dağılımları, hipotez testleri, karar kuralları, Johns ve Tukey Tablosu. 5. Bölüm için: Binom ve çok değişkenli dağılımlarına ait hesaplayıcılar. 6. Bölüm için: Kitaptakiler dışında alternatif araştırma desenleri, sıralı kategorik veri kümeleri, bağımsız iki grubun ölçümlerinin sıralarının yer değiştirmesi durumunda hipotez testleri. 7. Bölüm için: Ortalamanın örneklem dağılımı, etki büyüklüğünün güven aralığı. 8. Bölüm için: G*Power'dan başka bir hesaplayıcı, anlamlı farklar ile ilgilenildiğinde güç hakkında düşünceler. 9. Bölüm için: Regresyon katsayısının örneklem dağılımı, Regresyon ve korelasyon katsayısı için faydalı bir kaynak. 10. Bölüm için: Sınıf-içi korelasyon katsayıları ve türleri, güvenirlik. 11. Bölüm için: Basit varyans analizi. 12. Bölüm için: Çoklu karşılaştırmalara ait kitap içindekilere ek açıklamalar, yanlış buluş oranına (false discovery rate) başka bir bakış. 13. Bölüm için: Faktöriyel varyans analizi (eşit olmayan örneklemler durumu, tesadüfî etkiler ve iç-içe desenler için SAS programı). 14. Bölüm için: Tekrarlı ölçümler için çoklu karşılaştırmalar, karma bir model olarak tekrarlı ölçümlerin analizi, eğilim analizi için bir döküman. 15. Bölüm için: Korelasyonların özel yapıları ile veri oluşturma. Çoklu regresyonun başka bir örneği, çok değişkenli lojistik regresyon, mediation ve moderation. 16. Bölüm için: Kovaryans analizinin başka bir örneği, genel doğrusal modellerin SPS yardımıyla çözümlenmesi. 17. Bölüm için: Meta-analize genel bakış, meta-analizi neye benzer, tek-durum desenleri hakkında daha fazla açıklama, tek-denek desenleri, Wohn Crawford'un makalesi. 18. Bölün için: Yeniden örnekleme, özyükleme tesadifîleşme testleri üzerine Prof. Dr. David C. Howell'ın bir yazısı.

Kitabın aslında kullanılan istatistik sembolleri ve bunların tanımlarına ait bir liste vardır. Bu liste tercümeye de konulmuştur. Ancak kitapta, bunların dışında İngilizce kısaltmalar da kullanılmıştır. Tercüme sırasında İngilizce kısaltmalar da Türkçeye çevrilmiş ve İngilizceleri de belirtilerek Türkçeleri kullanılmıştır. Böyle kısaltmalar için de ayrı bir Tercümede Kullanılan Türkçe Terimler ve Kısaltmalar Listesi hazırlanarak tercümeye eklenmiştir. Bu kısaltmalar içinde istatistikle ilgili olanlar da vardır, fakat ülkemizde (hâttâ dünyada) terimler ve kısaltmalar üzerinde birlik henüz sağlanamamış olduğundan okumayı kolaylaştırmak amacıyla böyle bir liste konulmasına ihtiyaç duyulmuştur.

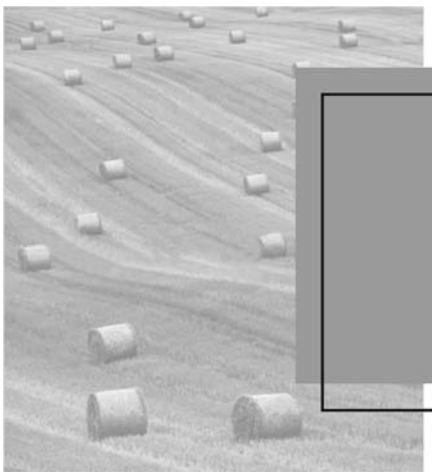
Kitabın aslında, bilgisayar yazılımlarına yapılan bazı ekler ve analiz sonuçlarının çıktıları da vardır. Bunların asılları kitaba aynen konulmuştur. Ancak analiz sonuçlarının çıktılarına ait Türkçeleştirmeler yapılmamıştır.

Diğer önemli bir nokta, kitap içindeki bazı örnekler ile alıştırmalara ait veri kümeleri kullanılması gereklidir. Bu veri kümeleri kitap içinde, eklerde veya başka bir yerde verilmemiş, kitabın Web sayfasına konulmuştur.

Kitapta yer verilen istatistik teknikleri davranış ve sosyal bilimler alanlarında başvurulan yöntemlerdir. Örnekler daha çok psikoloji ve eğitim alanlarından alınmıştır. Bu nedenlerle kitabın psikoloji ve eğitim başta olmak üzere tüm davranış ve sosyal bilimler alanlarındaki öğrenciler ve araştırmacılar için faydalı olabileceği düşünülmektedir.

Bu değerli eseri, kâr amacı gütmeden, dilimize ve ülkemizin bilim insanlarına kazandıran başta Sayın Servet Sarıkaya olmak üzere Pegem Akademi Yayınevi yönetimine, kitabın tercümesini büyük bir dikkatle okuyarak hatalarımızı düzelten ve önerilerde bulunan Doç. Dr. Kasım Kıroğlu'na, Pegem Akademi Yayınevi'nin tercüme kitaplar görevlisi Sayın Özlem Sağlam'a ve emeği geçen tüm Yayınevi çalışanlarına teşekkürü borç biliriz.

Editör
Mayıs 2017



# Tercüme Edenler Listesi

| Bölüm | Bölüm Adı | Tercüme Eden |
|---|---|---|
| **Bölüm 1**<br>**Önsöz**<br>**Yazar Hakkında** | Temel Kavramlar + Önsöz + Yazar Hakkında | Arş. Gör. Dr. Mehmet Bars<br>*Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi* |
| **Bölüm 2** | Verinin Tanımlanması ve Keşfedilmesi | Yrd. Doç. Dr. Derya Çobanoğlu Aktan<br>*Hacettepe Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 3** | Normal Dağılım | Öğr. Gör. Dr. Zafer Çepni<br>*Hacettepe Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 4** | Örneklem Dağılımları ve Hipotez Testi | Öğr. Gör. Dr. Zafer Çepni<br>*Hacettepe Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 5** | Olasılığın Temel Kavramları | Yrd. Doç. Dr. Mustafa İlhan<br>*Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi* |
| **Bölüm 6** | Kategorik Veri ve Ki-Kare | Öğr. Gör. Dr. Gülşen Taşdelen Teker<br>*Sakarya Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 7** | Ortalamalara Uygulanan Hipotez Testleri | Yrd. Doç. Dr. Gülden Kaya<br>*Sakarya Üniversitesi, Eğitim Fakültesi* |

| | | |
|---|---|---|
| **Bölüm 8** | Güç | Doç. Dr. Neşe Güler<br>*Sakarya Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 9** | Korelasyon ve Regresyon | Okt. Dr. Melek Gülşah Şahin<br>*Gazi Üniversitesi, Gazi Eğitim Fakültesi* |
| **Bölüm 10** | Diğer Korelasyon Teknikleri | Yrd. Doç. Dr. Seçil Ömür Sünbül<br>*Mersin Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 11** | Basit Varyans Analizi | Yrd. Doç. Dr. Seçil Ömür Sünbül<br>*Mersin Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 12** | İşlem Ortalamaları Arasında Çoklu Karşılaştırmalar | Yrd. Doç. Dr. Önder Sünbül<br>*Mersin Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 13** | Faktöriyel Varyans Analizi | Yrd. Doç. Dr. Önder Sünbül<br>*Mersin Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 14** | Tekrarlı-Ölçümler Desenleri | Yrd. Doç. Dr. Yeşim İmamoğlu<br>*Boğaziçi Üniversitesi, Eğitim Fakültesi*<br><br>Yrd. Doç. Dr. Zeynep Çiğdem Özcan<br>*Maltepe Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 15** | Çoklu Regresyon | Yrd. Doç. Dr. Mustafa İlhan<br>*Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi* |
| **Bölüm 16** | Genel Lineer Modeller Olarak Varyans ve Kovaryans Analizi | Yrd. Doç. Dr. Mustafa İlhan<br>*Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi* |
| **Bölüm 17** | Meta Analizi ve Tek Durum Desenleri | Doç. Dr. Neşe Güler<br>*Sakarya Üniversitesi, Eğitim Fakültesi* |
| **Bölüm 18** | Yeniden Örnekleme ve Parametrik Olmayan Testler | Doç. Dr. Neşe Güler<br>*Sakarya Üniversitesi, Eğitim Fakültesi* |

# Yazar Hakkında

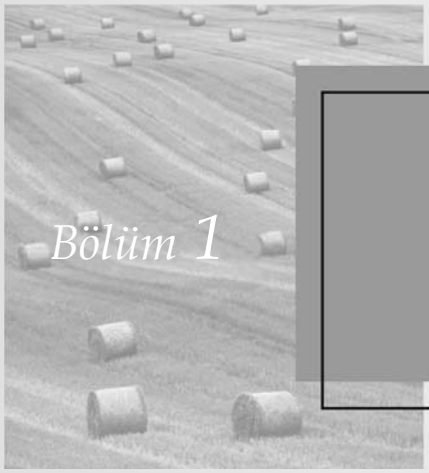


Courtesy David C. Howell

Profesör Howell, Vermont Üniversitesi'nde onursal profesördür. 1967 yılında Tulane Üniversitesi'nden doktorasını aldıktan sonra 2002'deki emekliliğine kadar Vermont Üniversitesinde Psikoloji Bölüm Başkanı olarak görev yapmıştır. Ayrıca İngiltere'deki iki üniversitede birer yıl misafir profesör olarak görev yapmıştır.

Davranışsal Bilimlerde İstatistik Ansiklopedisi'nin Brian Everitt ile birlikte editörüdür. Yakın zamanda, araştırma tasarımı ve istatistik konularındaki çeşitli kitaplarda bölümler yazmıştır.

Profesör Howell Colorado'da yaşamaktadır, başlıca hobileri ise kayak ve yürüyüş yapmaktır.



*Bölüm 1*

# Temel Kavramlar

## Hedefler

Bu kitapta sunulan problem türlerinin ve istatistiksel bir yöntem seçmekle ilgili sorunların incelenmesi.

## Konular


**1.1** Önemli Terimler
**1.2** Betimsel ve Vardamsal İstatistik
**1.3** Ölçmede Ölçekler
**1.4** Bilgisayar Kullanmak
**1.5** Bu Basım Hakkında Bilinmesi Gerekenler

STRES, hepimizin yaşam boyu üstesinden gelmeye zorlandığı bir şeydir. Stres, etrafımızdakilerle ve çevre ile olan günlük etkileşimlerimizde, yaklaşan bir sınav karşısında ve birçok öğrenci için zorunlu bir istatistik dersi almaları gerektiğini fark ettiklerinde ortaya çıkar. Çoğumuz strese tepki vermeyi ve ona uyum sağlamayı öğrensek de bu öğrenme süreci genellikle yavaş ve acı vericidir. Bu oldukça sıkıcı durum, bir istatistik dersi için çok güzel bir giriş olmasa da beraberinde uygulamalı bir araştırma projesinin tanımını getirir ve dolayısıyla bu tanım da, önemli birtakım istatistik kavramlarını açıklar. Yıllar önce, bu projeye benzer bir projede çalıştım; dolayısıyla bu örnek, kuramsal bir proje örneği değildir.

Bir grup eğitimci, stresin nasıl yönetileceğini ve stres yönetiminin özgüven üzerindeki etkisini lise öğrencilerine öğretmek üzere tasarlanan bir ders oluşturdu. Bu grup, onlara, dersin ne kadar iyi bir şekilde gitmekte olduğunu ve özellikle dersi alan öğrencilerin, dersi almamış olanlardan daha fazla özgüveni olup olmadığını söyleyebilecek dışarıdan bir araştırmacıya ihtiyaç duydular. Şu an için, grubun programının bir değerlendirmesini tasarlama görevinin bize verildiğini varsayacağız. Tasarlayacağımız deney tam olmayacak fakat deneyleri tasarlamak ve analiz etmekle ilgili bazı konuları ve size tanıdık gelmesi beklenen bazı istatistik kavramlarını gösterecek.

## 1.1 Önemli Terimler

Stres yönetimi programı lise öğrencileri için tasarlanmış olsa da, bu programı ülkedeki tüm lise öğrencilerinden oluşan evrene uygulamak imkânsızdır. İlk olarak bu türden çok fazla öğrenci vardır. Ayrıca programı, faydalı olup olmadığını anlayana kadar herkese uygulamak anlamsızdır. Tüm lise öğrencileri evrenine uygulamak yerine, bu evrenden bir öğrenciler örneklemi çekeceğiz, programı onlara uygulayacağız. Fakat herhangi bir eski örneklemi değil **tesadüfî (rastgele) bir örneklem** alacağız, bununla beraber gerçek tesadüfî örneklemler normal şartlarda uygulanamaz ve hatta imkânsız olmasa bile normal şartlarda pratik değildir. Tesadüfî bir örneklem almak için, evrendeki her bir elemanın (öğrencinin) eşit seçilme şansı olmasını garantiye alacak belli bir yöntemler bütünü izlenmelidir. (Tesadüfî örneklemi açıklayan en bilinen örnek, isimleri yazıp bir şapkanın içine atmak ve içinden bakmadan isim çekmektir. Neredeyse hiç kimse tam olarak bunu yapmasa da aklımızda olan şeyin iyi bir açıklamasıdır.)

**tesadüfi (rastgele) örneklem**

Öğrencilerden örneklem oluşturulduktan sonra kişilerin yarısı stres yönetimi programının uygulanacağı gruba, diğer yarısı da bu programı almayacak gruba **tesadüfî** olarak atanır.

**tesadüfi atama**

Bu tanım şu ana kadar, ayrıntılara daha fazla girilmesini gerektiren çeşitli kavramları ortaya çıkarmıştır; bunları belirtmek gerekirse evren, örneklem, tesadüfî örneklem ve tesadüfî atamadır. **Evren**, ilgilenilen olayların (öğrencilerin puanları, insanların gelirleri, farelerin koşma hızları vb.) kümesidir. Dolayısıyla eğer Amerika Birleşik Devletleri'ndeki tüm lise öğrencilerinin özgüven puanlarıyla ilgileniyorsak tüm lise öğrencilerinin özgüven puanlarının kümesi bir evren oluşturur. Bu durumda da evren milyonlarca elemandan oluşacaktır. Öte yandan eğer sadece Fairfax, Vermont'daki (4 bin kişiden az insanın yaşadığı bir kasaba) lise son sınıf öğrencilerinin özgüven puanlarıyla ilgileniyorsak evren yaklaşık olarak sadece 100 öğrenciden oluşur.

**evren**

Buradaki önemli nokta evrenin her boyutta olabileceğidir. Evren, nispeten küçük ve kolayca oluşturulabilen sayı kümelerinden, büyük fakat sonlu olan ve tümünü oluşturmanın pratik olmadığı sayı kümelerine kadar değişebilir. Aslında evrenler, öğrencilerin teorik olarak çizebilecekleri tüm olası karikatür çizimleri gibi, toplanması imkânsız olan her boyutta olabilir. Evrenler, nispeten küçük ve kolayca toplanabilen sayı kümelerinden, büyük fakat sonlu olan ve tümünü toplamanın pratik olmadığı sayı kümelerine kadar değişebilir. Aslında evrenler, öğrencilerin teorik olarak çizebilecekleri tüm olası karikatür çizimleri gibi, sonsuz sayıda bir sayı kümesi dahi olabilir. Ne yazık ki ilgilendiğimiz evrenler genellikle çok büyüktür. Bu durumda pratikte varılan sonuç, evrenin tümünün nadiren elde edilebil-

örneklem

diğidir. Onun yerine o evrenden sadece bir gözlemler **örneklem**i alınması ve o örneklemin, evrenin özellikleri hakkında birtakım çıkarımlar yapmak için kullanılması gerekir.

Örneklemin tam anlamıyla tesadüfî olduğu varsayıldığında evrenin sadece belli özelliklerini kestirmekle yetinilmez, aynı zamanda kestirimlerin ne kadar doğru olduğu konusunda da çok iyi bir fikre sahip olunur. Örneklemin tesadüfî olmama derecesine göre, kestirimlerimiz manidar olabilir veya olmayabilir, çünkü örneklem tüm evreni doğru bir şekilde yansıtabilir veya yansıtmayabilir.

Tesadüfî olmanın, dikkate almamız gereken en az iki boyutu vardır. Birinci boyut, örneklemin evreni temsil eder nitelikte olup olmadığıdır, çünkü örneklemin evrenden çıkarımlar yapmak için kullanılması istenir. Bu durum temel olarak, evrenden tesadüfî *örneklem* almayı kapsar ve **dış geçerlik** denilen kavramı ortaya çıkarır. Dış geçerlik, örneklemin evreni yansıtıp yansıtmadığı sorusunu ifade eden bir kavramdır. Nebraska'daki küçük bir kasabadan alınan örneklem, Amerika Birleşik Devletleri nüfusunun İspanyolca konuşan yüzdesinin geçerli bir kestirimini ortaya çıkarmayabilir. Yalnızca Amerika'nın güneybatısından alınan bir örneklem de geçerli bir kestirimi ortaya çıkarmayabilir. Öte yandan Nebraska'daki küçük bir kasabadan alınan örneklem, ansızın ortaya konulan bir uyarıcıya karşı insanların tepki zamanının mantıklı bir kestirimini verebilir. Bu noktada, tesadüfî örneklemeden bahsetmenin problemlerinden birini görüyoruz. Tamamen tesadüfî olmayan bir şekilde kişiler veya katılımcılardan oluşturulan bir örneklem de bizim için tamamen tesadüfî bir örneklemden elde edilen sonuçlar kadar bizi ve başkalarını ikna edici olabilir. Diğer yandan eğer tesadüfî olmayan örneklemimiz, tamamen tesadüfî bir örneklemle elde edebileceklerimizi temsil etmiyorsa, çıkarım yapma becerimiz tehlikeye girer ve sonuçlarımız çok yanıltıcı olabilir.

dış geçerlik

Devam etmeden önce, birçok insanın aklını karıştıran bir noktayı açıklığa kavuşturalım. Bu husus, bir insanın örnekleminin başka bir insanın evreni olabilmesidir. Örneğin, bir öğretme aracı olarak bu kitabın etkinliği ile ilgili bir çalışma yürütseydim, bir sınıfın bir sınavdaki notlarını, bu kitabı kullanan veya kullanma potansiyeli olan tüm öğrencilerin notlarının bir evreninin tesadüfî olmayan bir örneklemi olarak değerlendirebilirdim. Öte yandan bu sınıfın öğretim elemanı büyük bir olasılıkla bu kitapla çok fazla ilgilenmeyecek, onu sadece kendi öğrencileri ilgilendirecektir. Öğretmen, aynı not kümesini bir evren olarak değerlendirecektir. O halde istatistik öğretmek isteyen bir kimse benim evrenimi (kitabımı kullanan herkes), daha büyük bir evrenden (istatistik konusunda herhangi bir kitabı kullanan herkes) alınan ve hiç tesadüfî olmayan bir örneklem olarak değerlendirebilir. Dolayısıyla evren tanımı, çalışmak istediğiniz duruma göre değişir.

Bizim stresle ilgili çalışmamızda, ABD'deki lise öğrencilerinden tamamen tesadüfî bir örneklem almayı ve onlara stres yönetimi programını uygulamayı ciddî olarak değerlendirme olasılığımız çok azdır. Basitçe ifade etmek gerekirse, bu uygulanamaz bir şeydir. O zaman, tesadüfî örnekleme varsayımına dayanan istatistiksel yöntemlerden ve süreçlerden nasıl faydalanacağız? Bunu yapabilmemizin tek yolu sadece, sonuçlarımızın ilgilendiğimiz evreni genel olarak temsil ettiğine inandığımız durumlarda bu yöntemleri ve süreçleri uygulamaya dikkat etmektir. Eğer bu varsayımı yapamazsak çalışmamızı tekrar tasarlamamız gerekir. Buradaki sorun istatistiksel geliştirmeden çok sağduyu sorunudur. Örneklemimizin ABD'deki lise öğrencilerini temsil etmediğini düşündüğümüz oranda sonuçlarımızın yorumunu da sınırlamalıyız. Kestirimlerimizin geçerliği, örneklemin evreni temsil etme derecesi kadardır.

tesadüfi atama

Tesadüfî olma kavramının ikinci boyutu **tesadüfî atama** ile ilgilidir. Tesadüfî seçim, verimizin *kaynağı* ile ilgili ve çalışmamızın sonuçlarını evrenin tümüne genellemek için önemli olsa da kişilerin tedavi gruplarına tesadüfî olarak atanması deneyimizin bütünlüğü için çok önemlidir. Bu **iç geçerlik** kavramıyla ilgilidir. Elde ettiğimiz sonuçların, bu gruplara konulan kişilerden değil, gruplara uygulanan tedavideki farklardan kaynaklandığını sağlamak isteriz.

iç geçerlik

Örneğin, eğer örneklemimizdeki tüm çekingen öğrencileri bir gruba ve tüm iddialı öğrencileri de diğer gruba koyarsak sonuçlarımızın, bu gruplara uyguladığımız tedavilerden çok, öğrencileri gruplara atamamızın bir fonksiyonu olması çok muhtemeldir. Öğrencilerin, girecekleri grupları kendilerinin seçmeleri de benzer riskleri taşır ve bu durum, çalışmanın yapıldığı olayın doğası olmadıkça genellikle olumsuz bir fikirdir. Gerçek uygulamada tesadüfî atama genellikle tesadüfî örneklemeden çok daha önemlidir. Tıptaki çalışmalarda çalışmanın niteliğine göre, gruplara tesadüfî atama son derece karmaşık bir hal alabilir ayrıca işi, sadece deneyin bu boyutunu kontrol etmek olan insanlar da bu çalışmalarda zaman zaman görev alabilmektedirler.

Örneklemdeki insanların (elemanların) seçiminden ve bu insanların tedavi gruplarına atanmasından söz ettikten sonra her bir gruba nasıl tedavi yapıldığını ve sonuçta çıkacak veri kümelerinin özelliklerinin nasıl saptanacağını değerlendirmeye sıra gelir. İnsanların stresle başa çıkabilme becerilerini ve onlara yapılan farklı tedaviler altında yüksek özgüvenlerini koruyabilme becerilerini çalışmak istediğimizden ve strese verilen tepki birçok değişkenin bir fonksiyonu olduğundan çalışmayı planlamanın kritik bir boyutu da çalışılacak değişkenleri seçmektir. Bir **değişken**, farklı değerler alabilen bir nesne veya olayın bir özelliğidir. Örneğin, saç rengi bir değişkendir çünkü bir nesnenin (saçın) bir özelliğidir ve farklı değerler alabilir (kahverengi, sarı, kızıl, gri vb.) Stres yönetimi programı için yapılan tedaviler, öğrencinin özgüveni, sosyal destek, cinsiyet, kişisel kontrol derecesi ve tedavi grubu gibi özelliklerin tümü konuyla ilgili değişkenlerdir.

değişken

İstatistikte değişken kavramı, bağımsız ve bağımlı olma yönlerinden ikiye ayrılır. Örneğimizde grup üyeliği **bağımsız bir değişken**dir, çünkü onu kontrol ediyoruz. Tedavilerin neler olacağına ve her bir tedaviyi kimlerin göreceğine karar veririz. Hangi grubun tedavi göreceğine ve hangisinin tedavi görmeyeceğine biz karar veriyoruz. Eğer erkekleri ve kadınları karşılaştırıyorsak şurası kesindir ki bir insanın cinsiyetini kontrol edemeyiz fakat cinsiyetlerle çalışıp çalışmayacağımıza (oldukça zor bir karar) ve erkekleri kadınlarla karşılaştırmayı isteyip istemediğimize karar veririz. Öte yandan veri kümesi -sonuç olarak çıkan özgüven puanları, kişisel kontrol duygularının puanları vb.- **bağımlı değişken**lerdir. Çalışma temel olarak bağımsız değişkenlerle, sonuçları ise bağımlı değişkenlerle ilgilidir. Bağımsız değişkenler nicel veya nitel ve kesikli veya sürekli olabilir. Bağımlı değişkenler ise genellikle nicel ve süreklidirler, fakat elbette bu durum her zaman geçerli değildir. Bu terimler aşağıda tanımlanmaktadır.[1]

bağımsız değişken

bağımlı değişken

Kesikli değişkenler ile sürekli değişkenler arasında bir ayrım yapmak gerekirse **kesikli değişkenler**, cinsiyet ve lise sınıfı gibi yalnız sınırlı sayıda değerler alabilen değişkenlerdir; **sürekli değişkenler** ise yaş ve özgüven puanı gibi en azından teorik anlamda ölçekteki en düşük ve en yüksek noktalar arasındaki herhangi bir değeri alabilirler.[2] Görüldüğü gibi bu ayrım, veriyi işleme biçimimizde önemli bir rol oynar.

kesikli değişken
sürekli değişken

Kesikli ve sürekli değişkenlerin ayrımıyla yakından ilgili bir diğer kavram nicel ve kategorik veri kümeleri arasındaki ayrımdır. **Nicel veri**, (bazen **ölçme verisi** de denir) derken herhangi bir ölçmenin sonuçları kastedilir. Nicel veriye örnek olarak bir testten alınan puanlar, insanların kiloları, bir özgüven ölçeğindeki puanlar vb. verilebilir. Her durumda, bir özelliği ölçmek için bir araç kullanılır ve ölçme sonucunun, belli bir nesnenin bir özelliğini "ne kadar" temsil ettiği ile ilgilenilir.

nicel veri
ölçme verisi

Öte yandan **kategorik veri (frekans-sıklık verisi** veya **nitel veri kümeleri** olarak da bilinirler), "çalışmamızda 34 kadın ve 26 erkek var" veya "15 kişi 'yüksek kaygıya sahip', 33 kişi 'nötr', 12 kişi de 'düşük' kaygıya sahip olarak sınıflandırıldı" gibi ifadelerle açıklanır. Burada birtakım şeyleri kategorize ediyoruz (sınıflandırıyoruz) ve veri kümelerimiz

kategorik veri (frekans-sıklık verisi)

nitel veri kümeleri

1 Birçok insan bağımlı değişkenin hangisi, bağımsız değişkenin hangisi olduğunu hatırlamak konusunda zorluk yaşar. Bunları, bağımlı ("dependent") ve veri ("data") kelimelerinin her ikisinin de İngilizcede "d" harfi ile başlamasından hatırlayabilirsiniz.

2 Aslında sürekli bir değişken, ölçeğin iki ucu arasındaki herhangi bir değerin (örn. 32,485687...) mümkün olduğu değişkendir. Bununla beraber uygulamada bir değişkeni, birçok farklı değer alabildiği zaman sürekli değişken olarak, sadece birkaç farklı değer alabildiği zaman ise kesikli değişken olarak kabul ederiz.

her bir kategorinin (dolayısıyla kategorik veri) frekanslarından oluşuyor. Bir çalışmada birkaç yüz insan yer alabilir, fakat sonuçlar (veri kümesi) sadece iki veya üç sayıdan oluşabilir ve bunlar da her bir kaygı kategorisine düşen insanların sayılarıdır. Tersine, eğer insanları yüksek, orta ve düşük kaygıya göre sıralamak yerine onların her birine sürekliye yakın bir kaygı ölçeğine dayanan bir puan verilseydi ölçme veri kümesi ile çalışıyor olunacaktı ve veri, o değişken üzerindeki her bir kişinin puanlarından oluşacaktı. Her iki durumda da değişken *kaygı* olarak etiketlenir. Çoğu ayrımda olduğu gibi ölçme verisi ve kategorik veri arasındaki ayrımın da fazlaca üzerine gidilebilir. Bununla birlikte bu ayrım faydalıdır ve bir değişkenin ölçme değişkeni mi yoksa kategorik bir değişken mi olduğu sorusuna verilecek cevap uygulamada hemen her zaman açıktır.

## 1.2 Betimsel ve Vardamsal İstatistik

Strese müdahale programımıza döndüğümüzde, ölçülecek değişkenleri seçtikten ve uygulamayı yaptıktan sonra ham verinin yani puanların elde edilmesi konusuyla karşı karşıya geleceğiz. İstatistik alanının, bu puanlardan veri kümesini elde etme ve onu kullanma ile ilgili iki ana bölümü vardır.

**betimsel istatistik**

Amacımız sadece bir veri kümesini açıklamak olduğunda **betimsel istatistiği** kullanırız. Örneğin; veri kümelerimizle yapmak istediğimiz ilk şeylerden biri onların grafiğini çizmek, ortalamaları ve diğer bazı ölçüleri hesaplamak ve aşırı puanları veya tuhaf bir şekle sahip puan dağılımlarını aramaktır. Bu yöntemlere betimsel istatistik denir çünkü bunlar temel olarak veri kümelerini açıklamayı amaçlarlar. Betimsel istatistik alanı önceleri hiç ilginç olmayan bir alan olduğu gerekçesiyle küçümseniyordu. Bu alanı küçümseyenlerin çoğu da Time dergisi gibi yayınlar için biçimsiz grafikler çizenlerdi. Yirmi beş yıl önce John Tukey, keşfedici istatistik veya **keşfedici veri analizi (KVA)** adını verdiği alanı geliştirdi. Tukey, veri kümelerine yakın bir dikkat göstermenin ve daha teknik yöntemleri uygulamadan önce veriyi incelemenin gerekliliğini gösterdi. Tukey'nin yeniliklerinden bazıları istatistiğin ana görüşleri arasına girmiş durumdadır ve sonraki bölümlerde bunların üzerinde durulacaktır; bazıları ise henüz popüler olmamıştır. Bununla beraber Tukey'nin, veri kümelerini yakından incelenmesi gerektiği konusundaki vurgusu çok etkili olmuştur. Bunun bir nedeni Tukey'ye bir istatistikçi olarak duyulan yüksek güvendir.

**keşfedici veri analizi (KVA)**

**vardamsal istatistik**

Veri kümelerini ayrıntılı bir şekilde tanımladıktan ve sayıların ne demek istediğini yüzeysel olarak anladığımıza emin olduktan sonra **vardamsal istatistik** kavramıyla özellikle ilgileneceğiz. Aslında bu kitabın çoğu vardamsal istatistikle ilgilidir. Stresin özgüven üzerindeki etkisi konusundaki deneyimizi tasarlarken, evrenin tümüyle ilgili ölçmeleri yapmanın mümkün olmadığını kabul etmiştik; bu yüzden o evrenden örneklemler almıştık. Bununla birlikte, temel sorularımız evrenin kendisiyle ilgilidir. Örneğin, gerçekte elimizde olan şey sadece, programa katılmış öğrencilerin bir örnekleminin ortalama puanı olmasına rağmen programımızı almış olabilecek öğrencilerin tüm evren için ortalama bir özgüven puanını sormak isteyebiliriz.

**parametre**
**istatistik**

Ortalama özgüven puanı gibi, evrenin tümüne atıfta bulunan bir ölçüye **parametre** denir. Aynı ölçü eğer bizim topladığımız bir veri örnekleminden hesaplanmışsa ona **istatistik** adı verilir. Parametreler *gerçek olan* ve ilgilenilen oluşumlardır, ilgili istatistikler ise aslında *tahminler*dir. Bu kitapta yaptığımız şeylerin çoğu örneklem istatistiği (veya tahminler de diyebilirsiniz) ile ilgili olsa da ilgilenilen gerçeğin aslında ilgili evren parametresi olduğunu unutmamak gerekir. Örneklemin özellikleri (istatistik) konusunda bildiklerimizden, evrenin özellikleri (parametreler) hakkında bir şeyler vardamak isteriz. Varsayıma dayalı çalışmamızda özellikle, potansiyel olarak programımıza kayıtlı olan öğrencilerin evreninin ortalama özgüven puanının, programa kayıtlı olmayan öğrencilerin evreninin ortalama özgüven puanından daha yüksek veya daha düşük olup olmadığını bilmekle ilgileniriz.Yine vardamsal istatistik alanıyla ilgileniyoruz, çünkü örneklemlerin özelliklerinden evrenlerin özelliklerine varıyoruz.

## 1.3 Ölçmede Ölçekler

Ölçmede ölçekler, bazı yazarların çok önemli, bazılarının ise ilgisiz olduğunu düşündüğü bir konudur. Bu kitabın yazarı ikinci grupta yer alsa da, genel olarak konuya bir yakınlığımızın olması önemlidir. (Bir şeyin, çalışılmaya değer olduğunu düşünmek için onunla aynı görüşü paylaşmanız gerekmez. İncil vaizleri, günah hakkında çok şey bildiklerini iddia ederler buna karşılık ondan uzak durmayı savunduklarını söylemek zordur.) Bir konu olarak istatistik sadece sıradan bir gerçekler bütünü değildir; daha ziyade çok çeşitli yorumlar ve görüşler ile bir araya getirilmiş bir açıklamalar ve düşünceler bütünüdür.

Ölçmede ölçekleri, istatistiksel yöntemlerin seçiminde çok önemli görenlerin başında büyük olasılıkla S. S. Stevens[3] geliyordu. Zumbo ve Zimmerman (2000), ölçeklerden kayda değer ölçüde bahsetmişler ve Stevens'ın sisteminin kendi tarihsel bağlamında görülmesi gerektiğini bize hatırlatmışlardır. 1940'larda ve 1950'lerde Stevens, "sert bilimlerde (hard sciences)" olup bilimsel ölçme konusunda sınırlı bir görüşe sahip olanların aksine psikoloji araştırmalarını savunma girişimlerinde bulunmuştur. Stevens psikolojiyi "saygı duyulan" bir dal haline getirmeye çalışmıştır. Stevens, çok seçkin profesyonel kariyerinin çoğunu psiko-fizik dalı için ölçmede ölçekler geliştirerek geçirmiş ve bilime önemli katkılar yapmıştır. Bununla beraber, o alanın dışında psikolojide Stevens'ın geliştirdiği türdeki ölçekleri geliştirmek için hem az bir çaba sarf edilmiş hem de bu konuya çok fazla gerçek ilgi gösterilmemiştir. Stevens'ın gözünü korkutan eleştiriler büyük ölçüde uçup gitti; bu eleştirilerle birlikte aynı zamanda, ölçmede ölçeklerin, uygun olan istatistik süreçlerini ciddi derecede etkilediği inancının büyük kısmı da uçup gitti. Fakat bu ölçümleri tanımlayan terimler önemli olmaya devam etmektedir.

### Nominal Ölçekler

**nominal ölçekler**

**Nominal ölçekler** bir anlamda gerçekten ölçek değildirler; nominal ölçekler, bir boyut boyunca sıralanan kalemleri ölçeklendirmezler, onları etiketlerler. Cinsiyet ve siyasî parti üyeliği gibi değişkenler nominal değişkenlerdir. Kategorik veri kümeleri genellikle nominal bir ölçekte ölçülürler çünkü gözlemlere sadece kategorik etiketler (örn. erkek veya kadın; Cumhuriyetçi, Demokrat veya Bağımsız) koyarız. Sayısal bir örnek olarak futbolculara verilen sayıların oluşturduğu kümeyi verebiliriz. Genellikle bu sayılar, oyuncuları birbirlerinden ayırmak için kullanılan uygun etiketler olmaktan öte bir anlam taşımazlar. Harfler veya hayvan resimleri de aynı ölçüde kolay bir şekilde kullanılabilir.

### Sıralama Ölçekleri

**sıralama ölçeği**

En basit gerçek ölçek **sıralama ölçeği**dir. Bu ölçek, bir süreklilik boyunca insanların özelliklerini, nesneleri veya olayları sıralar. Böyle bir ölçeğin mükemmel bir örneği Donanma'daki rütbelerdir. Bir deniz yarbayı, saygınlık olarak bir deniz albayından düşük bir seviyededir. Deniz albayı da bir tuğamiralden düşük seviyededir. Fakat bir yarbay ile bir albayın arasındaki saygınlık farkının bir albay ile bir tuğamiral arasındaki saygınlık farkıyla aynı olduğunu düşünmek için bir neden de yoktur. Buna psikolojiden bir örnek de Holmes ve Rahe (1967) tarafından geliştirilen yaşam stresi ölçeğidir. Bu ölçeği kullanarak, bir insanın hayatında son altı ayda meydana gelen değişiklik sayısını (evlilik, taşınma, yeni iş vb.) belirtebiliriz. 20 puana sahip birinin, 15 puana sahip birinden daha fazla stres yaşadığı; 15 puana sahip kişinin de 10 puana sahip birinden daha fazla stres yaşadığı varsayılır. Dolayısıyla insanlar, yaşamlarında yakın zamanda olan değişikliklere göre stres konusunda

3 Bu görüşe temel oluşturan matematik konularını incelemek isteyenler için, Stevens'ın Deneysel Psikolojinin El Kitabı (1951) adlı eserindeki Bölüm 1 mükemmel bir kaynaktır.

sıralanırlar. Bu sıralama ölçeğine bir örnektir, çünkü ölçekteki noktalar arasındaki farkların anlamı yoktur. 10 ile 15 puan arasındaki stres farkının 15 ile 20 arasındaki stres farkına eşit olduğunu varsayamayız.

## Eşit Arılıklı Ölçekler

**eşit aralıklı ölçek**

**Eşlit aralıklı ölçekle** elimizde, ölçek noktaları arasındaki farkların ölçek boyunca eşit olduğundan bahsedebiliriz. Bunun bilinen bir örneği Fahrenheit sıcaklık ölçeğidir. Bu ölçekte 10 puanlık bir fark, ölçek boyunca herhangi bir noktada aynı anlamı ifade eder. Dolayısıyla 10° F ve 20° F sıcaklıkları arasındaki fark ile 80º F ve 90º F sıcaklıkları arasındaki fark aynıdır. Bu ölçeğin aynı zamanda önceki iki ölçeğin özelliklerine de sahip olduğu unutulmamalıdır. Bununla birlikte eşit aralıklı ölçekte sahip olmadığımız şey, oranlar hakkında anlamlı bir şekilde konuşamadığımızdır. Dolayısıyla örneğin 40° F sıcaklığın, 80º F derece sıcaklığın yarısı olduğunu veya 20º F sıcaklığın iki katı olduğunu söyleyemeyiz. Bu amaç için oran ölçeklerini kullanmalıyız. (Bu konuda şunu söylemek gerekir: bir aralıklı ölçekten diğerine tamamen geçerli dönüşümler yaptığımız zaman bile –örneğin Fahrenheit sıcaklık ölçeğinden Celsius ölçeğine dönüşüm– aynı oranları elde etmeyiz. Dolayısıyla, sıcaklıklar birbirleriyle karşılaştırılabilir olsalar bile Fahrenheit ölçeğindeki 40º ve 80º dereceler arasındaki oran, Celsius ölçeğindeki 4,4º ve 26,7º dereceler arasındaki orandan farklıdır. Bu durum, aralık ölçekleriyle çalışırken karşımıza çıkan oranların keyfî doğasını göstermektedir.)

## Oran Ölcekleri

**oran ölçeği**

**Oran ölçeği**nin, gerçek bir *sıfır noktası* vardır. Sıfır noktasının, 0º F ve hatta 0º C gibi keyfi bir sıfır noktası değil, gerçek bir sıfır noktası olması gerekir. (Gerçek bir sıfır noktası, ölçülen şeyin yokluğuna karşılık gelen noktadır. 0º F ve 0º C, sıcaklığın veya moleküler hareketin yokluğunu temsil etmediği için gerçek sıfır noktaları değildirler.) Oran ölçeklerinin örnekleri uzunluk, hacim, zaman vb. yaygın fiziksel ölçeklerdir. Bu ölçeklerle sadece daha önceki ölçeklerin özelliklerine sahip olunmaz, aynı zamanda oranlar hakkında konuşulabilir. 10 saniyenin, 5 saniyenin iki katı uzun zaman süresi olduğunu veya 100 pound ağırlığın 300 pound ağırlığın üçte biri ağırlık olduğunu fiziksel anlamda söyleyebiliriz vb.

Çalışmamızda kullandığımız ölçeğin türünün açıkça belli olacağını düşünebiliriz. Ne yazık ki özellikle davranış bilimlerinde topladığımız ölçümler için bu durum nadiren geçerlidir. Kullandığınız ölçeğin türü genellikle onu kullanma amacımıza bağlıdır. Aynı vücut yapısına sahip yetişkinler arasında ağırlık, ağırlık olarak oran ölçeğindedir, fakat en iyi ihtimalle fiziksel formun veya sağlığın bir sıralama ölçeğidir.

Nasıl kullanıldığına bağlı olan ve bir ölçeğe sahip ölçüm şekline örnek olarak bir evin sıcaklığını düşünün. Fahrenheit sıcaklığının genellikle aralık ölçeğinde olduğunu belirtiriz. Az önce bunu, aralıklı ölçeğin bir örneği olarak açıkladık ve şüphe yoktur ki bir fizikçi için 62° F ve 64° F arasındaki fark, 92° F ve 94° F arasındaki fark ile tamamen aynıdır. Bununla birlikte sıcaklığı, bir moleküler etkinlik endeksinden ziyade bir konfor endeksi olarak ölçüyorsak aynı sayılar artık bir eşit aralıklı ölçek oluşturmaz. 62° F sıcaklıkta olan bir odada oturan bir insan için sıcaklığın 64° F dereceye çıkması belirgin biçimde fark edilebilir (ve hoş karşılanan) bir durum olur. Aynı şey, 92° F ve 94° F olan oda sıcaklıkları arasındaki fark için söylenemez. Bu durum, *ölçtüğümüz özelliğin temeldeki değişken (örn. konfor) olduğu, sayıların kendileri olmadığı ve bunun ölçeği tanımlarken önemli olduğu gerçeğine işaret eder.* Bir *konfor* ölçeği olarak Fahrenheit dereceleri bir aralıklı ölçek oluşturmaz. Hatta Fahrenheit dereceleri sıralama ölçeği bile oluşturmaz çünkü konfor belli bir noktaya kadar sıcaklıkla birlikte artar, daha sonra da düşmeye başlar.

## Ölçmede Ölçeklerin Rolü

Yazarların, ölçeklere atfedilen önem konusunda farklı fikirlere sahip olduğunu daha önce belirtmiştim. Bazı yazarlar problemi tamamen görmezden gelmiştir, diğerleri ise kitapların tümünü farklı ölçeklerin etrafında düzenlemişlerdir. Bu konudaki mantıklı görüş (diğer bir deyişle benim görüşüm) ana meselenin, nesnelerden veya olaylardan topladığımız sayıları zihinlerimizde, kendilerine atfedilen nesne ve olaylara göre ayırmanın kesinlikle gerekli olduğudur. Böyle bir sav, oda sıcaklığı örneği için öne sürülmüştü. Burada ölçek (aralıklı ölçek veya sıralama ölçeği), sıcaklığın fiziksel bir özelliğini mi yoksa sıcaklığın insanlar üzerindeki etkisini mi (bir başka deyişle konforu) ölçmekle ilgilendiğimize göre değişiyordu. 2° F derecelik bir fark fiziksel olarak ölçeğin herhangi bir yerinde aynıdır; fakat zaten sıcak olan bir odadaki 2° F derecelik bir fark, nispeten daha serin bir odadaki 2° F derecelik bir fark kadar çok hissedilmeyebilir. Başka bir deyişle fiziksel birimlerin aralık ölçeğine sahibiz, fakat konfor için elimizde sıralama ölçeğinden fazlası bulunmuyor (yine belli bir noktaya kadar).

İstatistiksel testler, sayıların atıfta bulunduğu nesneleri veya olayları dikkate almadan söz konusu sayıları kullandıkları için işin temelindeki ölçeğin türünden bağımsız olarak standart matematiksel işlemlerin herhangi birini (toplama, çarpma vb.) uygulayabiliriz. Bu konuda mükemmel, eğlenceli ve çok tavsiye edilen bir yazı, Lord (1953) tarafından yazılmış "Futboldaki Sayıların İstatistiksel Olarak Ele Alınması" başlıklı yazıdır. Yazıda Lord, bu sayıların istediğiniz şekilde ele alınabileceğini çünkü "sayıların nereden geldiklerini hatırlamadıklarını" (s. 751) ifade eder.

Problem, herhangi bir çeşit istatistiksel işlemin sonuçlarını yorumlamak gerektiğinde ortaya çıkar. O noktada, istatistiksel sonuçların söz konusu nesnelerle veya olaylarla herhangi bir anlamlı yolla ilişkili olup olmadığını sormamız gerekir. Bu noktada artık istatistiksel bir sorunla değil, yöntembilimsel bir sorunla uğraşacağız. Hiçbir *istatistiksel* yöntem, bir kaygı anketinde bir grubun diğer bir gruptan daha yüksek puanlar elde etmesi gerçeğinin, temeldeki kaygı seviyeleri konusunda gruplar arasındaki farklar hakkında herhangi bir şey gösterip göstermediği hususunda bir şey söyleyemez. Ayrıca anketin, kaygı puanlarının (50 puanı, 25 puanının iki katı büyüklüktedir) bir oran ölçeğini sağlamasından memnun olmak da, puanlardaki artışlarla doğru orantılı bir şekilde artmayabilecek olan kaygıyı ölçmeye başlayacağımız gerçeğini gözden kaçırmamız anlamına gelir. İstatistiksel testlerimiz sadece elde ettiğimiz sayıları kapsayabilir ve ölçtüğümüzü düşündüğümüz nesneler veya olaylar hakkındaki ifadelerin geçerliliği temel olarak bu nesneler veya olaylar hakkındaki bilgimize dayanır; ölçüm yapma ölçeğine dayanmaz. Ölçümlerimizin, ölçmek istediğimiz şeyle mümkün olduğunca yakından ilişkili olmasını garanti etmek için elimizden geleni yapacağız. Fakat işin sonunda sonuçlarımız sadece elde ettiğimiz sayılar ve bu sayılar ve onların altında yatan nesneler veya olaylar arasındaki ilişkiye olan inancımız olacaktır.[4]

Yukarıda tartışılan konudan çıkan -ve bu kitapta kabul edilen- açık sonuç, ölçmede başvurduğumuz ölçeğin, istatistiksel teknik konusundaki seçimimizde çok önemli olmadığıdır. Açıkçası, bu istatistiksel manipülasyonların sonuçlarını yorumlamak için belli bir miktar sağduyu gereklidir. Sadece bir aptal, bir kişi tarafından mükemmel olarak, başka bir kişi tarafından ise çok kötü olarak değerlendirilen bir tablonun vasat olarak sınıflandırılması gerektiği sonucuna varır.

# 1.4 Bilgisayar Kullanmak

Otuz beş yıl önce bu kitabın ilk basımının yazıldığı sırada istatistiksel analizlerin çoğu masaüstü veya el hesap makineleriyle yapılıyordu; kitaplar da buna göre yazılıyordu. Şimdi yöntemler değişti ve hesaplamaların çoğu artık bilgisayarlarla yapılıyor.

---

4 Cohen'ın (1965) da belirttiği gibi "Bir keresinde Thurstone, bizim psikolojide insanları gölgelerine göre ölçtüğümüzü söylemiştir. Aslında klinik psikolojide insanları genellikle bir dans salonunda dans ederken ve dönen bir eski moda çok yüzlü aynanın yansımaları tarafından aydınlatılmış haldeyken ölçeriz" (s. 102).

Bu kitap, bilgisayarları yazıya dâhil ederek bilgisayarların artan kullanımını vurgulamaktadır. Kitap ilerledikçe ve hesaplamalar daha zahmetli hale geldikçe bilgisayarın olaya dâhil olma düzeyi kayda değer biçimde artar.

Basit yöntemlerde, kavramların tanımlanmasında hesaplamaya ilişkin formüller önemlidir. Örneğin, standart sapma için veya *t* testi için olan formül, standart sapmanın veya *t* testinin gerçekten ne olduğunu tanımlar ve o kavramları anlamlı hale getirir. Bu durumlarda, bilgisayar çözümlerinin örnekleri de verilse elle hesaplama vurgulanır. Kitabın ilerleyen bölümlerinde, örneğin, çoklu regresyon veya kayıp veri ile uğraşma tekniklerinde formüller daha az bilgilendirici hale gelir. Beş tahmin değişkenine sahip regresyon katsayılarını elde etmek için gereken formülün, sonuçta çıkan istatistikler konusundaki anlayışınıza bir şeyler eklemesini beklemek makul bir beklenti değildir. Aslında bu durum sadece konuyu daha da karmaşıklaştıracaktır. Böyle durumlarda hemen hemen sadece bilgisayar çözümlerine güveneceğiz.

Peki, hangi istatistik paketini kullanacağız? Günümüzde, istatistiksel analizler yürüten normal araştırmacı veya öğrenciler için birçok istatistik yazılım paketi mevcuttur. Bu kitapla birlikte ihtiyaç duyacağınız hemen her analizi yapabilecek en önemli büyük istatistik paketleri Minitab®, SAS®, SPSS™, Stata® ve S-PLUS paketleridir. Çok güvenilir ve kullanımı nispeten kolay olan paketler vardır ve bunlardan biri veya birkaç tanesi genellikle her yüksekokulun veya üniversitenin bilgisayar merkezinde bulunur. (Fakat hangi paketin en iyi olduğuna dair uzmanlar arasında tartışmalar vardır.) Bu kitapta SPSS çıktısının birçok örneği görülecektir. Önceleri çok çeşitli paketler kullandım, fakat eleştirmenler bir pakete yoğunlaşmamı önerdiler ve SPSS davranış bilimlerinde en yaygın kullanılan paket olma özelliğini taşır. *http://core.ecu.edu/psyc/wuenschk/sas/sas-programs.htm* adresinde SAS kodunun mükemmel program dosyalarını bulabilirsiniz ve bunların çoğu bu kitabın örnekleri ile bağlantılıdır.

İhtiyaç duyacağınız hemen her şeyi yapacak ucuz bir paket indirmek isterseniz Open Stat paketini tavsiye ederim. Bunun için, *http://statpages.org/miller/openstat/* adresine gitmeniz yeterli olacaktır. Veri dosyalarını okumayı biraz fazla detaycı bir yaklaşım olarak görüyorum fakat program iyi çalışıyor. Programı açıp örnekler girebilirsiniz veya benim örneklerimi değiştirdiğiniz zaman neler olabileceğini görebilirsiniz.

R denilen bir programlama ortamından da bahsetmem gerekir. Bu ortam, ücretsiz olarak sunulan kamuya açık bir yazılımdır. Bu, sevdiğim bir programdır (veya en son kullandığım zamandan beri onu ne kadar unuttuğuma bağlı olarak nefret etmeyi sevdiğim bir programdır). R, çok esnek bir programdır ve birçok metin, analizlerinin bazıları için R kodunu vermeye başlamıştır. Ayrıca programın, insanların bu program için yazdığı ve kamuya açık hale getirdiği özel amaçlara yönelik yüzlerce işlevi vardır. Şaşırtıcı biçimde R, birçok farklı disiplinde popüler olmaya başlamıştır ve bu durumun da devam edeceğini beklemekteyim. R'yi, *http://www.r-project.org/* adresinden indirebilirsiniz. Bu kitap boyunca arada bir, kendi yazdığım ve önemli hesaplamaları yapmak için R kodunu içeren Web sayfalarına atıfta bulunacağım ve bu hesaplamalar kolaylıkla R'ye de yapıştırılabilir. R'yi kullanmaya başlamanızı sağlayacak bir Web sayfası *www.uvm.edu/~dhowell/methods8/R-Programs/Using-R.html* adresidir. Temel içeriği daha sonra göreceğimiz materyalden oluşsa da, başlamak için belki daha da iyi bir kaynak Amholt (2007) tarafından yazılan yazıdır.

Excel gibi yaygın bir şekilde kullanılan tablolama programlarından da bahsetmem gerekir. Bu programlar birçok istatistiksel hesapları yapabilirler ve oldukça güzel grafikler çizebilirler. Bu programlar elle hesaplama yapmanın da mükemmel bir yoludur. Ayrıca bu programlar kullanıcıyı, hesaplamalara mantıksal olarak başlamaya zorlarken diğer yandan daha sonra yapacağınız incelemeler için gereken tüm ara adımları da elde tutmanızı sağlar. Geçmişte bu tür programlar sonuçlarının doğruluğu konusunda eleştirilmişti. Söz konusu programlar için son dönemlerde yazılan ek programlar, bu programların doğruluk derecesini büyük ölçüde arttırmıştır, bu yüzden bu programların değerini küçümsememek gerekir. Excel gibi programların aynı zamanda, çoğu kişinin kişisel bilgisayarında bir veya birden fazlasının kurulmuş olma avantajları vardır.

## 1.5 Bu Basım Hakkında Bilinmesi Gerekenler

Bazı yönlerden bu basım, geçmiş basımlardan sonra bir kırılmayı temsil etmektedir. Yıllar içinde her basım belli ölçüde yeni materyal içeriyordu ve metin, teori ve pratikteki değişiklikleri yansıtmak için değiştiriliyordu. Bu basımda, öğrencilerin anlaması gereken belli sayıdaki ek konudan bahsetmeyi düşündüm ve bunu da, bazı eski materyalleri kaldırmadan yapamazdım. Buna karşılık internet sayesinde materyal silinmeyecek, sadece yer değiştirecek. Örneğin, çoklu karşılaştırma testini veya log-doğrusal modelleri içeren bölümün eski ve tam bir versiyonunu görmek isterseniz benim web sitemden eski versiyonu indirebilirsiniz. Aynı şey, silinen diğer temel konular için de geçerlidir.

Bu kitabın ilk basımı 1973-1974 yıllarında yazılmıştı; o sıralarda ben İngiltere'deki Durham Üniversitesi'nde uzun dönemli akademik izindeydim. O dönemde, psikoloji dalındaki istatistik derslerinin çoğu deneysel psikologları hedefliyordu ve bu çalışmaların büyük bir yüzdesi de varyans analizi ile ilişkiliydi. B. J. Winer tarafından yazılan *Deneysel Tasarımda İstatistik İlkeleri* (1962, 1971) adlı eser, psikolojik istatistik alanının büyük bir kısmına hakim oldu ve deneylerin tasarlanma, yürütülme ve analiz edilme şekillerini etkiledi. Benim kitabımın ilk basımı büyük ölçüde bu trendi takip etti ve bunu da hakkıyla yaptığını düşünüyorum. Veri kümelerinin istatistiksel analizi o zamandan beri oldukça fazla değişikliğe tanık oldu. Varyans analizi hâlâ hakim teknik olsa ve çoklu regresyon onu yakından izlese de, birçok başka konu da listeye girmiştir. Örneğin, etki büyüklüğü ölçüleri şu anda, o zaman olduğundan çok daha önemli durumdadır. Meta çözümlemesi önceleri adeta hiçbir şey iken, bir alanı özetlemenin önemli bir yolu haline gelmiştir. Kayıp veriye yapılacak işlemler de, lineer karma modeller ve veri değerlendirme yöntemleri sayesinde önemli ölçüde değişmiştir. Yüksek kalitede araştırma yapmak isteyen öğrencilere bu konuları açıklamadan onları çalışmaya başlatamayız.

Fakat eğer metin yönetilebilir bir büyüklükte kalacaksa, bu konuları ve benzer konuları kitaba dahil etmek için bazı kısımların çıkarılması gerekiyor. Geçmişte örneğin çoklu karşılaştırma yöntemleri, birbiriyle rekabet halinde olan birçok yaklaşımın kayda değer dikkat gösterdiği bir konu olmuştu. Fakat eğer insanların literatürde gerçekten ne yaptıklarına bakarsanız, birbiriyle rekabet halindeki yaklaşımların çoğu uygulayıcılardan çok teorisyenlerin ilgisini çekmektedir. Benim düşünceme göre kayıp veriye yapılacak işlemler hakkında yazmak, ortalamalar arasında karşılaştırma yapmak konusundaki tüm alternatifleri listelemekten çok daha faydalıdır. Bu durum özellikle, eski materyali okumak isteyen herkes için onu kolayca bulunabilir hale getirmek kolay olduğundan doğrudur.

O halde hangi değişiklikleri bekleyebilirsiniz? Bu yazıdan başlayarak dahil edilecek olan en önemli konular, kayıp veriye daha kapsamlı işlemler yapılması, tekrarlayan ölçümleri ele almak için lineer karma modellerin kullanılması, güven aralıklarının üzerinde daha çok durma ve ayrıca klinik müdahalelerde çok önemli olan meta analizi ve tek konuda tasarıma giriş konularıdır. Bu materyalin tümünü istediğim derinlikte işleyemem fakat en azından sizin olaya girmenizi sağlayabilirim ve sizi ilgili kaynaklara yönlendirebilirim. Çoğu durum için benim kendi oluşturduğum ek materyaller mevcut ve bu materyallere internetten ulaşılabilir. Bu materyallere *http://www.uvm.edu/~dhowell/methods8* veya *http://www.uvm.edu/~dhowell/StatPages*/adresinden ulaşılabilir. Birinci sitede tüm veri dosyalarına, tek sayılı alıştırmalar için detaylı cevaplara, hatalar ve düzeltmelere ve SPSS konusunda benim kendi kılavuzlarıma erişim sağlayabilirsiniz. "Bu kitabın web sitesi" dediğim zaman bu siteyi kastediyorum. Yapmanız gereken tek şey oradaki linkleri izlemektir.

İstatistiksel yöntemler alanında yapılmakta olan belki de en büyük değişiklik "yeniden örnekleme" başlığı altındadır. 100 yıldan beri istatistik problemlerini, test istatistiğini hesaplayarak ve daha sonra da "eğer örneklem aldığım evrenin belli özellikleri varsa o zaman istatistiğimi uygun tablo değeri ile tam anlamıyla karşılaştırabilirim" diyerek çözdük. Fakat eğer evren bu özelliklere sahip değilse ne olur? Neyse; evrenin o özelliklere sahip olmasını umalım.

Standart bir dizüstü bilgisayar bile çok yüksek bir hesaplama hızına sahip olduğu için evren konusundaki bu varsayımlardan bazılarını eleyebiliriz. Herhangi bir evrenden örneklem aldığımızda istatistiğin nasıl görüneceğini düzenleyebiliriz. Bu yeniden örnekleme yöntemleri çok güçlü yöntemlerdir ve önemli bir rol oynamaya başlamıştır. Geleneksel yaklaşımlar yakın gelecekte durumlarını koruyacaktır, bunun tek nedeni, bu yaklaşımların insanların kullanmayı öğrendiği gibi kalmış olmasıdır. Fakat bu durum değişmektedir. Programınızın ekranlarından birinde "benzet" ("simulate") veya "önyükle" ("bootstrap") gibi bir şey söyleyen bir seçme kutusu gördüğünüzde bu, geleneksel yöntemlerin tabutuna çakılan bir başka çivi anlamını taşımaktadır.

Bu değişime ayak uydurmak için eldeki kitap, bilinen yaklaşımlarla beraber yeniden örnekleme (resampling) yöntemlerini de sunacaktır. Bu yöntemler yaşamınızı daha *zor hale* getirmeyecek çünkü anlaşılmaları kolay yöntemlerdir. Sizden istenilen sadece işin temelindeki kavramları öğrenmenizdir. Örneklemeyi yapabilmeniz için kısa programlar da vereceğim. İsterseniz onları kullanmayabilirsiniz. Yeni yöntemler, eski yöntemlerin hangi konularla ilgili olduğunu daha iyi anlamanıza yardımcı olacaktır.

## Anahtar Terimler

Tesadüfî örneklem (1.1)
Tesadüfen atama (1.1)
Evren (1.1)
Örneklem (1.1)
Dış geçerlik (1.1)
Tesadüfî atama (1.1)
İç geçerlilik (1.1)
Değişken (1.1)
Bağımsız değişken (1.1)
Bağımlı değişken (1.1)
Kesikli değişkenler (1.1)
Sürekli değişkenler (1.1)
Nicel veri (1.1)
Ölçme verisi (1.1)
Kategorik veri (1.1)
Frekans verisi (1.1)
Nitel veri (1.1)
Betimsel istatistik (1.2)
Keşfedici Veri Analizi (KVA) (1.2)
Vardamsal istatistik (1.2)
Parametre (1.2)
İstatistik (1.2)
Nominal ölçek (1.3)
Sıralama ölçeği (1.3)
Aralık ölçeği (1.3)
Oran ölçeği (1.3)

## Alıştırmalar

1.1 Hangi koşullar altında yüksekokulunuzun veya ünivesitenizin tüm öğrencileri bir evren olarak değerlendirilir?

1.2 Hangi koşullar altında yüksekokulunuzun veya üniversitenizin tüm öğrencileri bir örneklem olarak değerlendirilir?

1.3 Eğer yüksekokulunuzun veya üniversitenizin öğrencileri, Alıştırma 1.2'deki gibi bir örneklem olarak değerlendirilseydi, bu örneklem tesadüfi bir örneklem mi olurdu, tesadüfî olmayan bir örneklem mi olurdu? Neden?

1.4 Yerel bir telefon rehberinden isimler seçmek neden o şehirde oturanlardan tesadüfî bir örneklem oluşturmaz? O durumda kim az temsil edilirdi, kim fazla temsil edilirdi?

1.5 Bağımsız değişkenlere ve bağımlı değişkenlere ikişer örnek veriniz.

1.6 Bağımsız ve bağımlı değişkenlerin kullanıldığı bir deneyi açıklayan bir cümle yazınız.

1.7 Sürekli değişkenlere üç örnek veriniz.

1.8 Kesikli değişkenlere üç örnek veriniz.

1.9 Bir evrenin ortalama puanını kestireceğiniz bir çalışma örneği yazınız.

1.10 Bir evrenin ortalamasının gerçek sayısal değeriyle ilgilenmediğimiz, fakat bir evrenin ortalamasının diğer bir evrenin ortalamasından daha büyük olup olmadığını öğrenmek istediğimiz bir örnek veriniz.

1.11 Kategorik veriye üç örnek veriniz.

1.12 Ölçüm verisine üç örnek veriniz.

1.13 Çalıştığımız şeyin bir ölçme değişkeni veya kategorik bir değişken olabileceği bir örnek veriniz.

1.14 Her bir tür ölçme ölçeğine birer örnek veriniz.

1.15 Bazı amaçlar için oran ölçeği kullanılarak ölçülebilecek, başka amaçlar için ise aralıklı ölçek veya sıralama ölçeği kullanılarak ölçülebilecek bir değişken örneği yazınız.

1.16 Fareleri, onlara yiyecek verdiğimiz zaman düz yoldan oluşan bir labirentte koşmaları için eğittik. 12. denemede bir fare labirentin yarısında yattı ve uyudu. Bu durum, "hız" bir öğrenme endeksi olarak kullanıldığı zaman ölçmenin ölçeği hakkında ne söyler?

1.17 Alıştırma 1.16, motivasyon endeksi olarak kullanılan hız hakkında ne söyler?

1.18 Asıl ilgilendiğimiz şeyin, değişkenler arasındaki ilişkiye bakmak olduğu çalışmalara iki örnek veriniz

1.19 Asıl ilgilendiğimiz şeyin, gruplar arasındaki ilişkiye bakmak olduğu çalışmalara iki örnek veriniz.

1.20 İç geçerliliği açık bir şekilde açıklayan yazılar için bir Google araması yapınız.

1.21 Bağımsız değişken olarak tanımladığımız şeyin eş anlamlılarını bulmak için bir Google taraması yapınız. (Sunduğum şeyleri tam olarak anlamadığınız zaman Google'ın arkadaşınız olduğunu unutmayınız.)

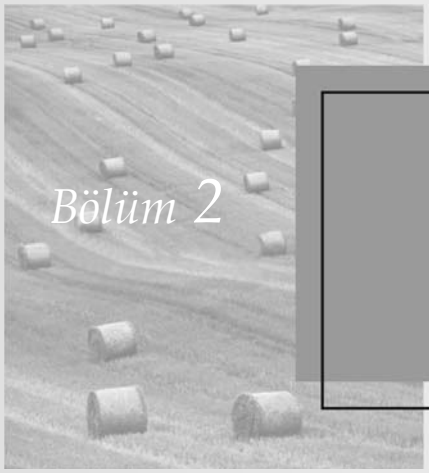
## Tartışma Soruları

1.22 21 Temmuz 1995 tarihli *Chicago Tribune* gazetesi, Beth Peres isimli bir dördüncü sınıf öğrencisinin bir çalışmasını haber yaptı. Peres, daha fazla harçlık için düzenlediği kampanyasına destek için kanıt toplarken, sınıf arkadaşlarının harçlık olarak ne aldıkları konusunda anket uyguladı. Ankete cevap veren 11 kız öğrencinin ortalama olarak haftalık 2,63$ aldığını, buna karşılık ankete cevap veren 7 erkek öğrencinin haftalık olarak ortalama 3,18$, bir başka deyişle kızlardan %21 fazla aldığını keşfedince Peres şaşırdı. Aynı zamanda erkek öğrenciler harçlıklarını kazanmak için kızların yaptığından daha az ev işi yapmak zorundaydı. Hikâye, kayda değer bir ulusal şöhret kazandı ve yetişkin kadınların yetişkin erkeklere karşı olan gelir eşitsizliğinin yaşamın çok erken bir döneminde başlayıp başlamadığının sorgulanmasına neden oldu.

a. Bu çalışmadaki bağımlı ve bağımsız değişkenler nelerdir ve nasıl ölçülürler?
b. Burada nasıl bir örneklemle uğraşıyoruz?
c. Örneklemin özellikleri Beth'in bulduğu sonuçları nasıl etkileyebilir?
d. Beth, tesadüfî örneklemeyi nasıl ele alabilir? Beth, tesadüfî atamayı nasıl ele alabilir?
e. Eğer bu çalışmada tesadüfî atama mümkün değilse bu durumun çalışmanın geçerliliği konusunda birtakım olumsuz çıkarımları olur mu?
f. Erkeklerin ve kızların gelirleri arasındaki herhangi gerçek evren farklarından ayrı olarak bu çalışmanın sonucunu etkileyebilecek değişkenlerden bazıları nelerdir?
g. Bu örneğin betimsel ve vardamsal istatistiksel özelliklerini açık bir şekilde birbirinden ayırınız.

1.23 *Journal of Public Health* dergisi, sigara içme ve sağlık arasındaki ilişkiye ait veri kümesini yayımladı (bkz. Landwehr & Watkins [1987]). Dergide, 21 gelişmiş batı ülkesindeki yetişkin başına düşen sigara tüketimi ve her bir ülke için koroner kalp hastalığı oranı rapor edildi. Veri kümesi açıkça gösterdi ki koroner kalp hastalığı oranı en çok, en yüksek sigara tüketim oranına sahip ülkelerdedir.

a. Niçin bu çalışmadaki örnekleme Batılı ve gelişmiş ülkelerle kısıtlanmış olabilir?
b. "Ölçmenin ölçekleri" olarak etiketlediğimiz şey açısından bu iki değişkenin özelliklerini nasıl saptarsınız?
c. Eğer amacınız sigara içmenin sağlık üzerindeki etkilerini çalışmaksa bu veri genel soruyla nasıl ilişkilidir?
d. Böyle bir çalışmada başka hangi değişkenlerin değerlendirilmesi gerekebilir?
e. Sigara şirketlerinin Asya'da çok büyük reklamlar yaptıkları rapor edildi. Şu anda Çinli kadınların sadece %7'si sigara içmektedir (buna karşılık Çinli erkeklerin ise %61'i sigara içmektedir). Bir sağlık psikologu, Çinli kadınların sigara içmesi olayındaki olası değişikliklerin sağlık etkilerini çalışmaya nasıl başlayabilir?



*Bölüm 2*

# Verinin Tanımlanması ve Keşfedilmesi

## Hedefler

Verinin grafik gösterimi, merkeze eğilme ve dağılma ölçüleri yardımıyla özetlenerek nasıl daha kolay yorumlanabilir hale getirilebileceğinin gösterilmesi.

## Konular


**2.1** Verinin Grafiğinin Çizilmesi
**2.2** Histogramlar
**2.3** Veri Kümesine Ait Grafiğin Sürekli Hale Getirilmesi
**2.4** Gövde-ve-Yaprak Gösterimi
**2.5** Dağılımların Tanımlanması
**2.6** Sembol ile Gösterim (Notation)
**2.7** Merkeze Eğilme Ölçüleri
**2.8** Değişme (Yayılma) Ölçüleri
**2.9** Kutu Grafikleri: Dağılımın ve Uç Değerlerin Grafik Gösterimleri
**2.10** Değişme Ölçülerinin SPSS Yardımıyla Elde Edilmesi
**2.11** Yüzde Birlikler, Çeyrekler ve Onda Birlikler
**2.12** Doğrusal Dönüşümlerin Veri Üzerindeki Etkisi

HAM VERİ KÜMESİ, tek başına, seçim gününden önce e-postanıza gelen gereksiz mesajlar kadar bilgilendirici ve ilgi çekicidir. Veriyi ister düzgün bir şekilde veri toplama formunda satırlar halinde düzenleyin, isterseniz çöp sepetine atacağınız tarihi geçmiş bir duyuru kâğıdının arkasına karalayarak yazın, onlar hâlâ sadece bir araya getirilmiş sayılardan ibarettir. Yorumlanabilir olmaları için öncelikle belli bir mantık çerçevesinde düzenlenmeleri gerekir. Aşağıdaki gerçek deney bu adımların bazılarını açıklamaktadır.

İnsanlar kısa süreli belleklerinde depolanan bilgiyi nasıl işlerler? Eğer size biraz önce ekranda gösterilen beş basamaklı sayının "6" rakamını içerip içermediği sorulsa, kısa süreli belleğinizdeki ekranı sırasıyla tarayarak "hayır birinci basamakta değildi, hayır ikinci basamakta değildi ve benzeri" biçimde sıralı işlem mi kullanırsınız? Yoksa belleğinizdeki bütün basamakları aynı anda "6" rakamı ile karşılaştırmak için paralel işlem mi yaparsınız? İkinci yaklaşım daha hızlı ve daha etkili olacaktır. Ancak insanlar işlerini her zaman en hızlı ve en etkili biçimde yapmazlar. Peki ya siz bu işlemleri nasıl yaptığınızı düşünüyorsunuz? Siz belleğinizin arka bölgelerini araştırıp içeriye giren kişiyi nasıl Jennifer olarak belirliyorsunuz? Onu belleğinizdeki aynı yaştaki bütün kadınlarla mı karşılaştırdınız, yoksa karşılaştırmayı paralel bir şekilde mi yaptınız? (İkinci örnek kısa süreli bellekten çok uzun süreli belleği kullanır, fakat sorular benzerdir.)

Saul Sternberg 1966 yılında, insanların kısa süreli belleklerinden bilgiyi nasıl hatırladıklarını araştıran basit, ünlü ve önemli bir çalışma gerçekleştirmiştir. Bu çalışma halen, araştırma literatüründe geniş olarak atıf almaktadır. Bu çalışmada deneklere, kısa süreliğine önlerindeki ekranda bir, üç ve beş basamaklı karşılaştırma sayıları gösterilmiş. Bu gösterimden kısa süre sonra ekranda deneklere tek bir test rakamı gösterilip, deneklerden test rakamını karşılaştırma setinde gördülerse bir butona (pozitif butonu), eğer görmedilerse diğer butona (negatif butonu) basmaları istenmiştir. Bunun için örnek olarak aşağıdaki iki uyaran gösterilebilir:

| | | | | | |
|---|---|---|---|---|---|
| Karşılaştırma | 2 | 7 | 4 | 8 | 1 |
| Test | | | 5 | | |

(İki uyaran kümesinin aynı anda değil sırasıyla gösterildiğini unutmayalım, yani gösterimde buradaki satırlardan sadece biri görünür durumdaydı.) 5 rakamı karşılaştırma kümesinin bir parçası değildir ve denek negatif butonuna basarak cevap vermelidir. Sternberg deneklerin cevaplamak için harcadıkları süreyi saniyenin 100'de biri duyarlılığında ölçtü. Bu işlemler tesadüfî olarak düzenlenmiş denemelerde defalarca tekrarlandı. Çünkü Sternberg insanların bilgiyi nasıl işlediklerini, cevap verme süresinin karşılaştırma kümesindeki basamak sayısına ve test rakamının karşılaştırma kümesinde bulunup bulunmadığına bağlı olarak nasıl değiştiğini araştırmak istedi. (Eğer karşılaştırmayı sıralı olarak yaparsanız karar vermek için gerekli süre karşılaştırma kümesindeki basamak sayısı arttıkça artar. Karşılaştırmayı paralel biçimde yaparsanız, karşılaştırma kümesindeki basamak sayısının etkisinin olmaması gerekir.)

Her ne kadar Sternberg'in amacı farklı koşullardaki veri kümelerini karşılaştırmak olsa bile, uyarıcı koşullarına bakılmaksızın biz cevap sürelerinin bütününü ele alarak veri kümesi hakkında hemen bir izlenim elde edebiliriz. Tablo 2.1'deki veri kümesi Sternberg'in deneyine benzer biçimde, fakat sadece bir denekten -benden- toplandı. Hiç bir biçimde değiştirilmesine izin verilmedi, yani burada sunulan veri kümesi sadece doğru denemelerden gelmektedir.

## 2.1 Verinin Grafiğinin Çizilmesi

Görebileceğiniz gibi, bir bakışta yorumlayabilmek için Tablo 2.1'de çok fazla miktarda rakam vardır. Veriyi düzenlemenin en basit yöntemlerinden biri onların grafiğini çizmektir.

**Tablo 2.1** Sayı belirleme deneyinden elde edilen cevaplama süresi veri kümesi

| Karşılaştırma Uyarıcısı* | Cevaplama süresi, saniyenin 100'de biri olarak | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 40 | 41 | 47 | 38 | 40 | 37 | 38 | 47 | 45 | 61 | 54 | 67 | 49 | 43 | 52 | 39 | 46 |
| 1Y | 47 | 45 | 43 | 39 | 49 | 50 | 44 | 53 | 46 | 64 | 51 | 40 | 41 | 44 | 48 | 50 | 42 |
| | 90 | 51 | 55 | 60 | 47 | 45 | 41 | 42 | 72 | 36 | 43 | 94 | 45 | 51 | 46 | 52 | |
| | 52 | 45 | 74 | 56 | 53 | 59 | 43 | 46 | 51 | 40 | 48 | 47 | 57 | 54 | 44 | 56 | 47 |
| 1N | 62 | 44 | 53 | 48 | 50 | 58 | 52 | 57 | 66 | 49 | 59 | 56 | 71 | 76 | 54 | 71 | 104 |
| | 44 | 67 | 45 | 79 | 46 | 57 | 58 | 47 | 73 | 67 | 46 | 57 | 52 | 61 | 72 | 104 | |
| | 73 | 83 | 55 | 59 | 51 | 65 | 61 | 64 | 63 | 86 | 42 | 65 | 62 | 62 | 51 | 62 | 72 |
| 3Y | 55 | 58 | 46 | 67 | 56 | 52 | 46 | 62 | 51 | 51 | 61 | 60 | 75 | 53 | 59 | 56 | 50 |
| | 43 | 58 | 67 | 52 | 56 | 80 | 53 | 72 | 62 | 59 | 47 | 62 | 53 | 52 | 46 | 60 | |
| | 73 | 47 | 63 | 63 | 56 | 66 | 72 | 58 | 60 | 69 | 74 | 51 | 49 | 69 | 51 | 60 | 52 |
| 3N | 72 | 58 | 74 | 59 | 63 | 60 | 66 | 59 | 61 | 50 | 67 | 63 | 61 | 80 | 63 | 60 | 64 |
| | 64 | 57 | 59 | 58 | 59 | 60 | 62 | 63 | 67 | 78 | 61 | 52 | 51 | 56 | 95 | 54 | |
| | 39 | 65 | 53 | 46 | 78 | 60 | 71 | 58 | 87 | 77 | 62 | 94 | 81 | 46 | 49 | 62 | 55 |
| 5Y | 59 | 88 | 56 | 77 | 67 | 79 | 54 | 83 | 75 | 67 | 60 | 65 | 62 | 62 | 62 | 60 | 58 |
| | 67 | 48 | 51 | 67 | 98 | 64 | 57 | 67 | 55 | 55 | 66 | 60 | 57 | 54 | 78 | 69 | |
| | 66 | 53 | 61 | 74 | 76 | 69 | 82 | 56 | 66 | 63 | 69 | 76 | 71 | 65 | 67 | 67 | 55 |
| 5N | 65 | 58 | 64 | 65 | 81 | 69 | 69 | 63 | 68 | 70 | 80 | 68 | 63 | 74 | 61 | 85 | 125 |
| | 59 | 61 | 74 | 76 | 62 | 83 | 58 | 72 | 65 | 61 | 95 | 58 | 64 | 66 | 66 | 72 | |

*Y = Evet, test uyarana dahildi; N= Hayır, dahil değildi; 1, 3 ve 5 karşılaştırma uyarıcısının basamak sayısını gösterir

© Cengage Learning 2013

Veri kümesinin grafikle gösterilmesinde kullanılan birkaç yaygın yöntem vardır. Bunlardan bazıları frekans dağılımları, histogramlar ve gövde ve yaprak gösterimleridir. Bunlar bu bölümde sırasıyla açıklanacaktır. (Ben grafiklerin mümkün olduğunca basit olması ve gereksiz detaylarla zihin karıştırıcı mesajlar vermemesi gerektiğine inanırım. Ancak grafiklerin başka hiçbir yöntemle görülemeyen önemli bilgileri nasıl ortaya koyduğunu gösteren bir örnek görmek isterseniz http://blog.ted.com/2007/06/hans_roslings_j_1.php adresini ziyaret ediniz, çok etkileyicidir.)

## Frekans Dağılımları

frekans dağılımları

Veri kümemizi belli bir mantık çerçevesinde düzenlemenin ilk adımı olarak onların **frekans dağılımları** oluşturulabilir. Bizim örneğimiz için, her bir cevap süresinin kaç kez tekrarlandığını sayabiliriz. Örneğin, denek 5 kez 50/100 saniyede ve 12 kez 51/100 saniyede cevap vermiştir. Bir defasında telaşlanmış ve cevaplaması 1,25 saniye (125/100) sürmüştür. Veri kümesindeki her bir cevaplama süresinin hangi sıklıkla meydana geldiğini raporlayan frekans dağılımı Tablo 2.2'de gösterilmiştir.

Tablo 2.2'de gösterilen dağılımdan cevaplama sürelerinin en düşük 36/100 saniye ile en yüksek 125/100 saniye gibi geniş bir dağılım gösterdiğini görebiliriz. Verinin çoğu 40/100 ile 90/100 saniye arasındadır ve 60/100 etrafında toplanma eğilimi göstermektedir. Bu eğilimi Tablo 2.1'deki düzenlenmemiş veri kümesinde görmek mümkün değildi.

**Tablo 2.2** Cevaplama sürelerinin frekans dağılımı

| Cevaplama Süresi, Saniyenin 100'de Biri | Frekans | Cevaplama Süresi, Saniyenin 100'de Biri | Frekans |
|---|---|---|---|
| 36 | 1 | 71 | 4 |
| 37 | 1 | 72 | 8 |
| 38 | 2 | 73 | 3 |
| 39 | 3 | 74 | 6 |
| 40 | 4 | 75 | 2 |
| 41 | 3 | 76 | 4 |
| 42 | 3 | 77 | 2 |
| 43 | 5 | 78 | 3 |
| 44 | 5 | 79 | 2 |
| 45 | 6 | 80 | 3 |
| 46 | 11 | 81 | 2 |
| 47 | 9 | 82 | 1 |
| 48 | 4 | 83 | 3 |
| 49 | 5 | 84 | 0 |
| 50 | 5 | 85 | 1 |
| 51 | 12 | 86 | 1 |
| 52 | 10 | 87 | 1 |
| 53 | 8 | 88 | 1 |
| 54 | 6 | 89 | 0 |
| 55 | 7 | 90 | 1 |
| 56 | 10 | 91 | 0 |
| 57 | 7 | 92 | 0 |
| 58 | 12 | 93 | 0 |
| 59 | 11 | 94 | 2 |
| 60 | 12 | 95 | 2 |
| 61 | 11 | 96 | 0 |
| 62 | 14 | 97 | 0 |
| 63 | 10 | 98 | 1 |
| 64 | 7 | 99 | 0 |
| 65 | 8 | . . . | . . . |
| 66 | 8 | . . . | . . . |
| 67 | 14 | 104 | 2 |
| 68 | 2 | . . . | . . . |
| 69 | 7 | 125 | 1 |
| 70 | 1 | | |

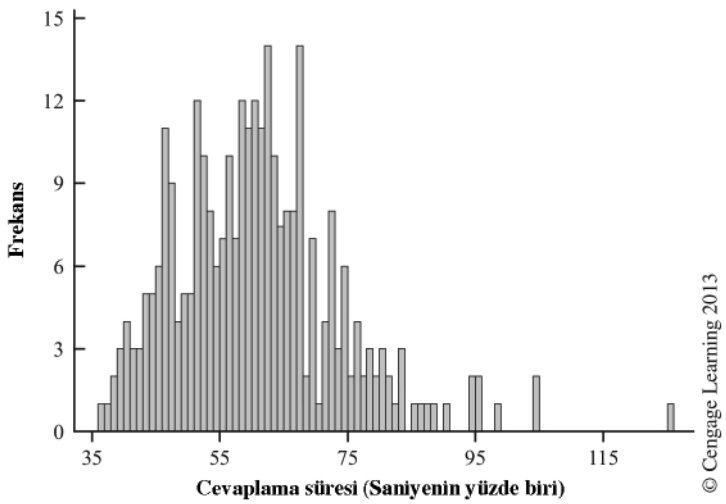
© Cengage Learning 2013

## 2.2 Histogramlar

Tablo 2.1'deki verinin dağılımından Şekil 2.1'deki grafiği kolaylıkla çizebiliriz. Fakat bunun gibi birçok farklı değerler alan bir değişken ile çalıştığımızda her bir değer genellikle düşük frekansta görülür ve ardışık aralıkların frekansları arasında önemli değişiklikler bulunur. Örneğin, on dört tane 67, fakat sadece iki tane 68 var. Bu gibi durumlarda,

histogram

**histogram**da[1] ardışık değerleri gruplamak daha mantıklıdır. Bunu yapmaktaki amacımız anlamlı olması olası olmayan bazı tesadüfî "paraziti" örtmek ve veri kümesindeki önemli eğilimleri korumaktır. Örneğin, biz veriyi 5/100 saniyelik bloklar halinde gruplayıp, 35 ile 39, 40 ile 44 ve benzer şekilde frekansları birleştirebiliriz. Böyle bir dağılımın örneği Tablo 2.3'te görülmektedir.




**Şekil 2.1** Cevaplama süresi frekans grafiği

gerçek alt sınır
gerçek üst sınır

Tablo 2.3'te, aralıkların alt ve üst sınırları, tablonun okunması daha kolay olsun diye, tam sayı olarak belirlenmiştir. Ancak şunun farkına varılmalıdır ki aralıkların gerçek sınırları (**gerçek alt sınır** ve **gerçek üst sınır** olarak bilinir) bir aralığın en üstü ile diğerinin en altının tam ortasındaki ondalık değerdir. Aralığın gerçek alt sınırı aralığın içinde yer alan en küçük değerdir. Benzer biçimde gerçek üst sınır aralıktaki en büyük değerdir. Örneğin cevaplama sürelerini saniyenin yüzde biri yerine binde biri olarak ölçseydik, aralıklar 35-39 yerine 34,5 ile 39,5 arasındaki tüm değerleri içerecekti, çünkü bu noktalar arasına düşen bütün değerler bu aralık arasındaki değerlere yuvarlanacaklardır.

**Tablo 2.3** Gruplandırılmış frekans dağılımı

| Aralık | Orta Nokta | Frekans | Yığmalı Frekans | Aralık | Orta Nokta | Frekans | Yığmalı Frekans |
|---|---|---|---|---|---|---|---|
| 35–39 | 37 | 7 | 7 | 85–89 | 87 | 4 | 291 |
| 40–44 | 42 | 20 | 27 | 90–94 | 92 | 3 | 294 |
| 45–49 | 47 | 35 | 62 | 95–99 | 97 | 3 | 297 |
| 50–54 | 52 | 41 | 103 | 100–104 | 102 | 2 | 299 |
| 55–59 | 57 | 47 | 150 | 105–109 | 107 | 0 | 299 |
| 60–64 | 62 | 54 | 204 | 110–114 | 112 | 0 | 299 |
| 65–69 | 67 | 39 | 243 | 115–119 | 117 | 0 | 299 |
| 70–74 | 72 | 22 | 265 | 120–124 | 122 | 0 | 299 |
| 75–79 | 77 | 13 | 278 | 125–129 | 127 | 1 | 300 |
| 80–84 | 82 | 9 | 287 | | | | |

© Cengage Learning 2013

1 İnsanlar "histogram"lardan bahsederken ona farklı anlamlar yükleyebilirler. Bazıları kategorilerin birleştirilip birleştirilmemesine bakmaksızın verinin dağılımı için kullanırlar. Diğerleri ise ardışık değerlerin birleştirildiği durumları histogram olarak adlandırırlar. Ben "histogram" ve "frekans dağılımı" kelimelerini birbirinin yerine kullanıyorum.

(İnsanlar genellikle bir birey tam olarak 39,500000000 puan alırsa, ne yapacakları konusunda endişelenirler ve bu yüzden iki aralığın kırılma noktasının ortasını kullanırlar. Bu konuda endişelenmeyin. İlk olarak, bu genellikle meydana gelmez. İkinci olarak her zaman yazı tura atabilirsiniz. Üçüncü olarak endişelenilecek daha önemli şeyler var. Sadece böyle bir durumda ne yapacağınız konusunda ihtiyari bir kural oluşturun ve ona uyun. Bu, insanlara istatistik çalışmalarının kafa karıştırıcı, sıkıcı ve her ikisi birden olduğunu düşündüren önemsiz konulardan biridir. Ben çift sayıya yuvarlamayı tercih ederim. Yani 25,5000'i 26'ya ve 44,5000'i 44'e yuvarlarım.)

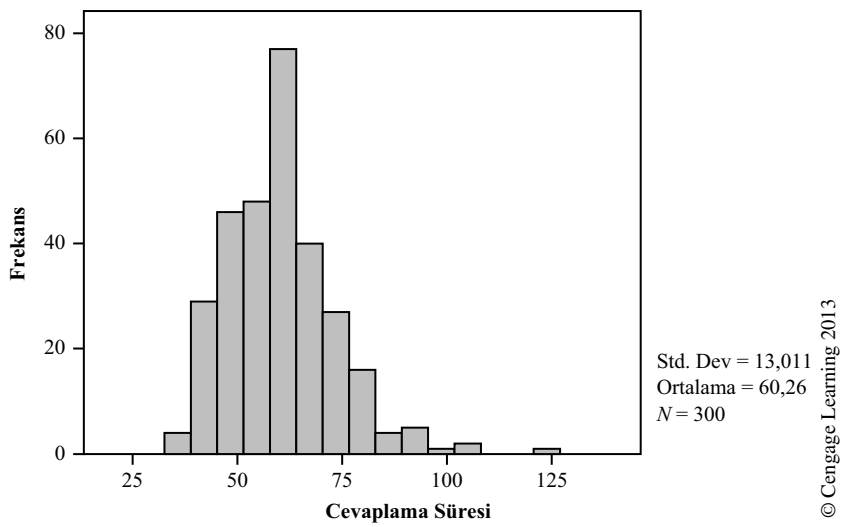
**orta noktalar**

Tablo 2.3'te listelenen **orta noktalar** alt ve üst sınırların ortalamalarıdır ve kolaylık için sunulmuşlardır. Biz veri kümesinin grafiğini çizdiğimizde genellikle onları aralıkların orta noktalarına denk geliyorlarmış gibi çizeriz.

Tablo 2.3 aynı zamanda her bir aralıktaki frekansları da gösterir. Örneğin, 35/100 ve 39/100 saniye aralığında yedi veri bulunmaktadır. Tablo 2.3'teki dağılımın grafiği Şekil 2.2'deki histogram ile gösterilmektedir.

İnsanlar genellikle veriyi gruplarken kullanılacak en uygun aralık sayısını sorarlar. Keen (2010), en uygun aralık sayısını belirlemede kullanılacak altı olası kuralı vermektedir, fakat bu kurallar öncelikle genel durumları ele alan bilgisayar programı yazarları için oluşturulmuştur. Eğer siz belli bir veri kümesiyle ilgiliyseniz, resmi kuralları bir kenara bırakıp anlamlı sonuçlar elde edene kadar ayarları değiştirebilirsiniz. Her ne kadar kullanılacak aralık sorusuna doğru bir cevap olmasa da, 10 veya 12 arası aralık genellikle mantıklıdır. Bu örnekte ben 19 aralık kullandım çünkü sayılar doğal olarak bu biçimde ayrılıyor ve ben birçok gözlem yaptım. Genel ve pratik olarak, aralığı 10 tanımlanmış aralığa bölmektense, sayı sistemindeki doğal aralıkları (örn. 0-9,10-19,...,100-119,120-139) kullanmak en iyisidir. Ancak eğer diğer sınırlar veriyi daha yorumlanır hale getiriyorlarsa o zaman o sınırları kullanmak daha yararlaı olur. Unutmayın ki siz veriyi daha anlamlı hale getirmeye çalışıyorsunuz; sizin probleminizi hiç görmemiş birinin ortaya koyduğu katı kurallara uymaya çalışmayın.

Şekil 2.2'deki cevaplama sürelerine dikkat ederseniz, verinin genellikle saniyenin yüzde 50-70 arasında yığıldığını, dağılımın düzgün bir şekilde yükseldiğini ve alçaldığını ve dağılımın sağa doğru giderek daraldığını görebilirsiniz. Bu sağa doğru frekansların azalması beklenen bir durumdur (sağa doğru çarpık olarak adlandırılır), çünkü kişinin cevap-




**Şekil 2.2** Cevaplama sürelerinin gruplanmış frekans histogramı

lama hızının bir sınırı vardır, fakat ne kadar yavaş cevaplayacağının bir sınırı yoktur. Dikkat edilirse uç değer (veri kümesinden kopmuş değerler) 125/100'dür. Bu değer uç değer olarak adlandırılır çünkü verinin geri kalanından oldukça ayrıktır. **Uç değerler** sıklıkla veri kaydındaki hataları temsil ederler, fakat bu durumda deneğin hangi butona basacağına karar verememesinden kaynaklanmıştır.

uç değerler

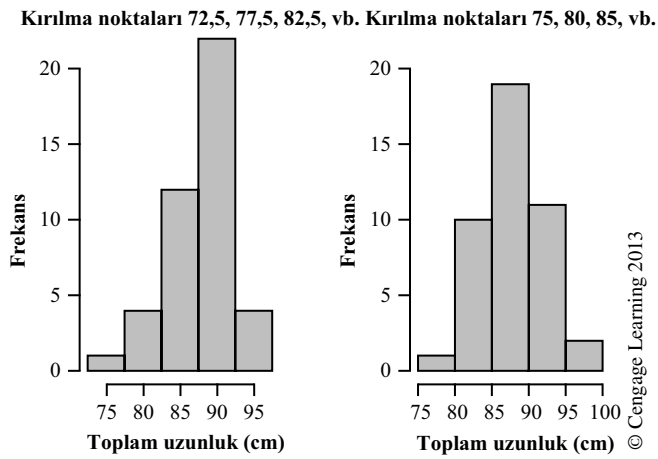
## 2.3 Veri Kümesine Ait Grafiğin Sürekli Hale Getirilmesi

Histogramlar, örneğin Şekil 2.1 ve 2.2'de gösterilenler gibi, genellikle veriyi anlamlı bir biçimde göstermek için kullanılırlar, fakat histogramların bazı problemleri vardır. Çok sayıda insanın belirttiği gibi, histogramlar yaygın kullanımlarına rağmen veriyi net bir biçimde betimleme işinde genellikle başarısız olurlar. Bu özellikle, aralıkların konumunun veya genişliğindeki küçük değişikliklerin dağılımın şeklinde önemli farklar oluşturduğu küçük örneklemlerde doğrudur. Wilkinson (1994) bu ve benzeri sorunlar hakkında mükemmel bir makale yazmıştır. Maindonald ve Braun (2007) Şekil 2.3'te görülen keseli farelerin uzunluklarının grafiğini örnek olarak vermektedir. Birinci grafik veri kümesini 72,5, 77,5, 82,5, ...vb. kırılma noktalarındaki kutulara böler. İkinci grafik 70, 75, 80 kırılma noktalarını kullanır. Keseli fareleri önemseyenler (evet böyle insanlar var), kullanılan kırılma noktasına bağlı olarak bu iki grafikten farklı sonuçlara ulaşabilirler. Sağdaki histogramdaki veri oldukça simetriktir, fakat soldaki histogramda sola doğru fark edilir bir kuyruk mevcuttur.

Şekil 2.2 cevaplama süreleri için oldukça uygun bir gösterimdir, fakat biz genellikle veriye sürekli hale getirilmiş eğriler uydurarak daha iyi sonuçlar alabiliriz. Birçok eğri uydurma yaklaşımı arasından burada iki tanesini anlatacağım. Bunlardan birisi, histogramın üzerine normal dağılımın çizilmesi (bir sonraki bölümde detaylı olarak açıklanacak) ve çekirdek yoğunluk grafikleri olarak bilinen diğer kullanımlarıdır.

### Normal Dağılıma Uyum

Her ne kadar henüz Bölüm 3'ü okumamış olsanız da, normal dağılım eğrisine aşina olabilirsiniz. Bu eğri, genellikle çan eğrisi olarak adlandırılır ve dağılımın orta noktası etrafında simetriktir ve her iki ucuna doğru azalır. Normal dağılımın belirli bir tanımı vardır, fakat biz bunu bir sonraki bölüme erteleyeceğiz. Şimdilik bizim için verimizin normal dağılıma




**Şekil 2.3** Aynı kuyruklu sıçan uzunlukları veri kümesine ait iki histogram

sahip olduğunu varsayacağımızı söylemekle yetineceğiz ve histogramın üzerine normal dağılım eğrisi çizmek bu varsayımın[2] ne kadar mantıklı olduğu hakkında bize bir nebze fikir verecektir.

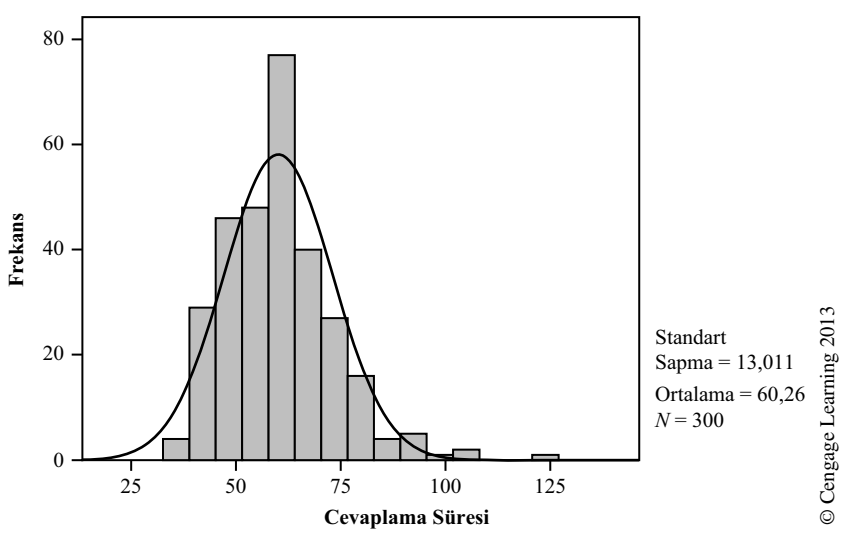
Şekil 2.4 SPSS kullanılarak çizildi ve gördüğünüz gibi her ne kadar veri kümesinin kabaca normal dağılım gösterdiği söylense de, gerçek dağılımın sol tarafı biraz kesilmiş ve en sağ tarafta beklenenden daha fazla gözlem var gibi. Normal dağılımın veriye uyumu çok kötü olmamakla birlikte bundan daha iyisini yapabiliriz. Bu yaklaşıma bir alternatif çekirdek yoğunluk grafiklerini oluşturmaktır.

## Çekirdek Yoğunluk Grafikleri

Şekil 2.4'te kuramsal bir dağılımı, veri kümesinin histogramı üzerine yerleştirdik. Bu dağılım, veri kümesinin sadece ortalama ve standart sapması kullanılarak oluşturuldu, ayrıca eğriyi verinin gerçek şekline uydurmak için hiçbir çaba sarf edilmedi. Biraz daha açık ifade edecek olursak, veri kümesinin sadece ortalama ve standart sapmasını (bu bölümün ileriki kısımlarında açıklanacak) hesaplayarak normal dağılımı histogramın üzerine çizebiliriz. Değişik veri noktalarının ve onların dağılımlarının, normal dağılımın histogram üzerine çizilmesinde hiçbir rolü yoktur. **Çekirdek yoğunluk grafikleri** ise bunun tam tersini yapar. Bu grafikler veriye düzgün bir biçimde uyan eğrileri çizerken aynı zamanda gözlemlerde eğriyi çok fazla bozabilecek tesadüfî parazitlenmelere de izin verilmemesi gerçeğini göz önünde tutar. Çekirdek yoğunluk grafikleri gözlemlerin ortalama ve standart sapmasını göz önüne almaz.

**çekirdek yoğunluk grafikleri**

Çekirdek yoğunluk grafiklerinin arkasında yatan temel fikir, her bir gözlemin bir miktar farklı olabileceğidir. Örneğin, cevaplama süresinin ölçüldüğü bir denemede, puanın 82 yerine 79 olması mantık çerçevesi dahilindedir. Benzer şekilde 86 yerine 73 de olabilirdi ama 20 veya 100 olması ihtimal dahilinde değildir. Başka bir deyişle, elde edilen her bir veri civarında olası alternatif dağılımlar vardır ve bu elde edilen tüm değerler için geçerlidir. Biz bu gerçeği veriye oldukça iyi uyan bir eğriyi (grafiği) çizmek için kullanacağız.




**Şekil 2.4** Üzerine normal dağılım çizilmiş cevaplama süresi histogramı

2 Bir sonraki bölümde göreceğiniz gibi bu, bir dağılımın normal olup olmadığını belirlemenin en iyi yolu değildir. Ancak yaygın olarak kullanılan bir yoldur.

Çekirdek kestirimleri Everitt ve Hothorn'dan (2006) alınan örnek ile grafik yardımıyla gösterilebilir. Bu araştırmacılar (*X*) bağımlı değişkeni için aşağıdaki basit veri kümesini kullandılar.

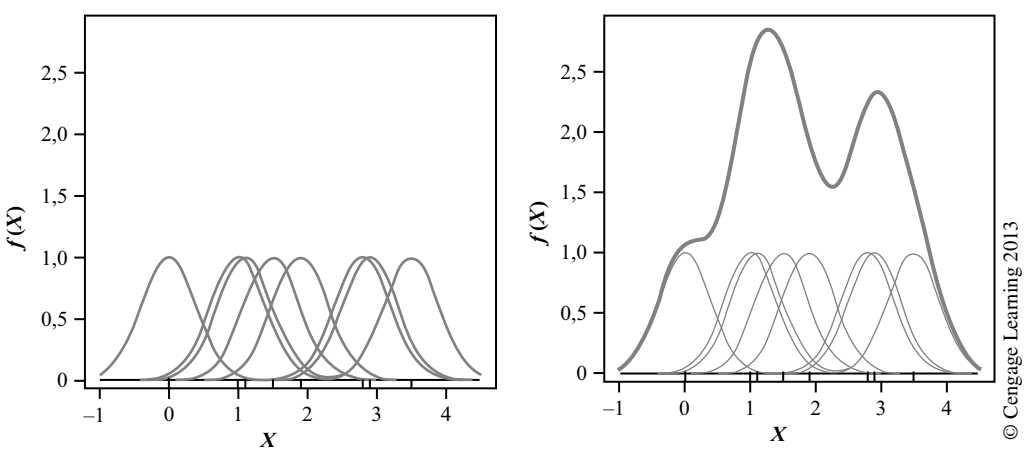
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *X* | 0,0 | 1,0 | 1,1 | 1,5 | 1,9 | 2,8 | 2,9 | 3,5 |

Eğer *X* ekseni boyunca bu noktaların grafiğini çizerseniz ve üzerlerine de bu gerçek değerler yerine elde edilebilecek alternatif değerlerin dağılımlarını çizerseniz Şekil 2.5'te solda görülen dağılım elde edilir. Everitt ve Hothorn'un bu küçük dağılımlara verdiği teknik terim "kasisler"dir. Dikkat edilecek olursa bu kasisler normal dağılımlardır, fakat normal dağılımın uygun olmadığı düşünülürse başka bir dağılım şekli de belirlenebilir.

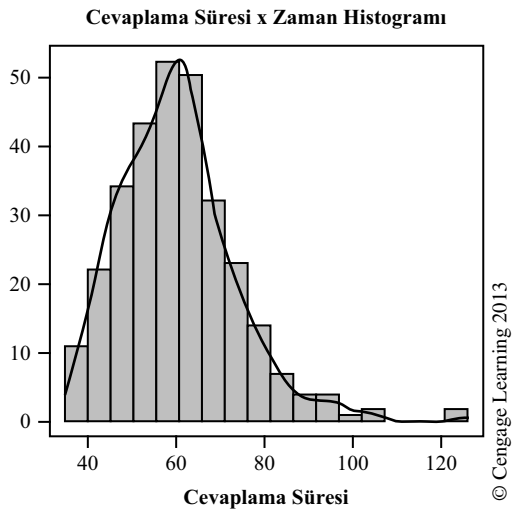
Şimdi kelimenin tam anlamıyla biz bu kasisleri dikey olarak toplayacağız. Örneğin, biz bu kasislerin her birini merkezlerindeki puanlar ile adlandıralım. *X* ekseninde 3,8 değerinin üzerinde biraz kasis_2,8, biraz kasis_2,9 ve büyük miktarda kasis_3,5 var. Bu üç kasisin *X* = 3,8'deki yüksekliklerini toplayarak eğrinin o noktadaki çekirdek yoğunluğuna ulaşılabilir. Aynı işlem her bir *X* değeri için yapılabilir. Bu yapıldığında Şekil 2.5'te sağdaki grafik bulunur. Kasislerin üzerindeki eğri büğrü grafik, başladığımız veri kümesi altında yatan dağılım hakkındaki en iyi tahminimizi gösterir.

Şimdi tekrar cevaplama sürelerine geri dönebilir ve histogramın üzerine çekirdek yoğunluk grafiğini çizebiliriz. (Sayıları çok fazla olduğu ve karmaşa oluşturdukları için kasisler çizilmemiştir.) Sonuçta ortaya çıkan grafik Şekil 2.6'da görülmektedir. Görüldüğü gibi, bu grafik veri kümesini temsil etme işini normal dağılımdan daha iyi yapmaktadır. Özellikle dağılımın kuyruk kısmına oldukça iyi uyum göstermiştir.

Çekirdek yoğunluk grafiklerini SPSS'te syntax kullanılarak çizebilirsiniz; SAS ve S-Plus kullanarak da oluşturabilirsiniz. Bu programların internette örneklerini bulmak oldukça kolaydır. Psikoloji, nöroloji ve sağlık bilimleri gibi alanlara doğru genişledikçe çekirdek yoğunluk grafikleri gibi teknikler daha da yaygınlaşmaktadır. Bu grafiklerin arkasında yatan teknik kavramlar vardır, örneğin, onları oluşturmak için kullanılan kasislerin şekli ve bant genişliği; şimdi siz bu grafikleri anlamak için gerekli temel bilgiye sahipsiniz. Çekirdek yoğunluk grafiklerinin SPSS ve R'de nasıl hesaplanacağını gösteren web sayfasının adresi http://www.uvm.edu/~dhowell/methods8/Supplements/R-Programs/kernel-density/kernel-density.html ve R programı kitabın web sayfasında http://www.uvm.edu/~dhowell/methods8/Supplements/R-Programs/R-Programms.html adresinde mevcuttur.




**Şekil 2.5** X için çekirdek yoğunluk fonksiyonunun gösterilmesi




**Şekil 2.6** Cevaplama süresi veri kümesine ait çekirdek yoğunluk grafiği

## 2.4 Gövde-ve-Yaprak Gösterimleri

Her ne kadar histogramlar, frekans dağılımları ve çekirdek yoğunluk grafikleri yaygın olarak kullanılsalar da her birinin olumsuz yönleri de vardır. Örneğin, histogramlar gözlemleri gruplanmış aralıklarda ortaya koydukları için her bir aralıktaki bireysel puanların gerçek değerleri görülmez. Frekans dağılımları ise her bir gerçek veriyi korurken verinin yeterince özetlenmediği durumlarda kullanılması zordur. Her iki eleştiriden de kaçınmamızı sağlayan alternatif yaklaşım **gövde-ve-yaprak gösterimi**dir.

gövde ve yaprak gösterimi

açımlayıcı veri analizi

**Açımlayıcı veri analizi** olarak bilinen genel veri analizi yaklaşımının bir kısmında John Tukey (1977) veri kümesinin anlamlı görseller olarak gösterilmesini sağlayan çeşitli metotlar geliştirmiştir. Çoğu temel istatistik programlarında da mevcut olan gövde-ve-yaprak gösterimi bu metotların en basitlerinden biridir. Burada örnek olarak cevaplama süresi verisini gözlemlerin sayısı oldukça fazla olduğu için kullanamayız. Bunun yerine 100 öğrencinin haftalık elektronik oyun oynamak için harcadıkları sürenin kaydedildiği kurgusal (farazi) veri setini kullanacağız. Ham verinin bir bölümü Şekil 2.7'de gösterilmiştir. Şeklin sol tarafı verinin bir kısmını göstermektedir (oyun oynamak için haftada 40 ile 80 dakika arasında harcayan öğrencilerin verisi) ve sağ tarafında bütün sonuçların gövde-ve-yaprak gösterimi bulunmaktadır.

önder basamaklar

en önemli basamaklar

gövde

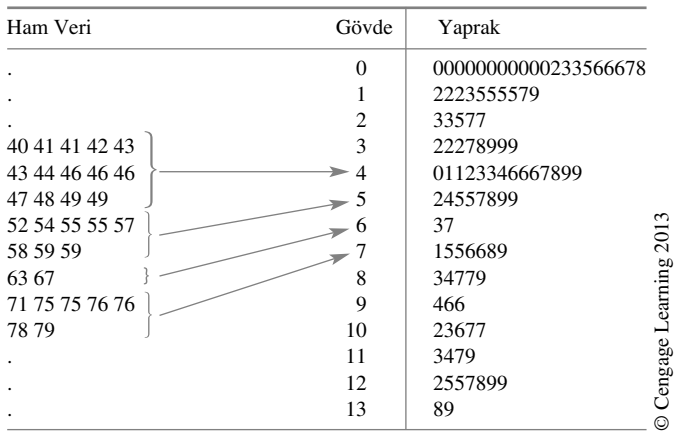
izleyici

daha az önemli basamak

yapraklar

Şekil 2.7'deki ham veriden, puanların arasında birkaç 40'lı, birçok 50'li, iki 60'lı ve birkaç 70'li değerler olduğu görülebilir. Biz burada 4, 5, 6 ve 7'yi bu puanlar için **önder basamaklar** (bazen **en önemli basamaklar**) olarak adlandırırız. Bu önder basamaklar **gövde**dir, gösterimimizin dikey ekseninde yer alırlar. On dört tane 40'lı sayı kümesinin içinde bir 40, iki 41, bir 42, iki 43, bir 44, 45 yok, üç 46, bir 47, bir 48, ve iki 49 vardır. Sayılardaki 0, 1, 2, 3 ve benzeri birler basamağı **izleyici** (veya **daha az önemli**) **basamak** olarak nitelenir. Bunlar **yapraklardır** - gösterimimizin[3] yatay elemanlarıdırlar.

3 Onlar basamağının gövdeyi ve birler basamağının yaprakları gösterdiği her zaman doğru değildir. Örneğin, eğer veri 100 ile1000 aralığında değişiyorsa yüzler basamağı gövde, onlar basamağı yaprakları gösterir ve birler basamağını ihmal ederiz.




**Şekil 2.7** Elektronik oyun veri kümesinin gövde-ve-yaprak gösterimi

Şekil 2.7'nin sağ tarafında 4'ün (gövdenin) hizasında bir 0, iki 1, bir 2, iki 3, bir 4, üç 6, bir 7, bir 8 ve iki 9 vardır. Bu yaprak değerleri ham verideki birler basamağına karşılık gelir. Benzer biçimde 5'in (gövdenin) hizasındaki yaprakların 50'li sayıların birler basamaklarına karşılık geldiği görülebilir. Gövde-ve-yaprak gösteriminden, gösterime giren ham verinin tamamını tekrar oluşturabilirsiniz. Örneğin, 11 öğrencinin elektronik oyun oynamak için 0 dakika harcadığını, bir öğrencinin iki dakika, iki öğrencinin ise üç dakika harcadığını görebiliriz. Ayrıca gösterimin şekli çarpık histograma benzemektedir, bu da bize gövde-ve-yaprak gösteriminin, grafik metotların getirilerini de sunduğunu gösterir.

Bu basit gövde-ve-yaprak gösterimin görünen bir kusuru, bazı veri kümeleri için gruplamanın çok büyük olmasıdır. Aslında tam bu nedenden dolayı tanıtıcı örnekte ben kurgusal (farazi) veri kullandım. Cevaplama süresi verisini kullanmayı denediğimde 50 gövdesine karşılık gelen 88 yaprak olduğunu gördüm. Neyse ki Tukey bizden önce bu yollardan geçip, bu problem için çözüm üretmiş.

Eğer problem 50 ile 59 arasındaki her şeyi bir arada göstermekten kaynaklanıyorsa, belki de aralıkları daha küçük aralıklara ayırmalıyız. Örneğin, 50-54 ve 55-59 ve benzeri aralıkları kullanmayı deneyebiliriz. Fakat o zaman gövde olarak sadece 5'i kullanamayız, çünkü o zaman iki aralığı birbirinden ayırt edemeyiz. Tukey 50-54 için "5*" ve 55-59'u göstermek için"5." alternatif yöntem olarak önerdi. Fakat burada bu öneri de bizim sorunumuzu halen çözmüyor, çünkü kategoriler halen çok büyük. Bu yüzden Tukey aralıkları daha da küçülterek 50-51 için "5*", 52-53 için "5t", 54-55 için "5f", 56-57 için "5s", 58-59 için "5." kullanmayı önerdi. (5'in yanındakiler aralıklardaki sayıların İngilizce baş harfleridir: iki ve üç her ikisi de t harfi ile başlar.) Eğer bu taslağı yanıtlama süresi verisine uygularsak Şekil 2.8'deki sonuçları elde ederiz. Kullanılacak gövde sayısına karar verme problemi, histogramlardaki aralık sayısını belirleme problemine benzer. Veriyi anlamlı bir şekilde gösteren durumları seçmek uygun olur. Buradaki tek kısıtlama, gövdelerin aynı genişlikte olmasıdır. Bir gövde 50-54 arasında iken diğeri 60-69 olamaz.

Dikkat edilirse Şekil 2.8'de diğerlerinde yaptığım gibi uç değerleri listelemedim. Gövde yerine Yüksek kelimesini kullandım ve gerçek değerleri gösterdim. Bu hem yerden kazanmak hem de uç değerlerin varlığının altının çizilmesi için yapılmıştır.

| Ham Veriler | Gövde | Yaprak |
|---|---|---|
| 36 37 38 38 39 39 39 40 | 3s | 67 |
| 40 40 40 41 41 41 42 42 | 3. | 88999 |
| 42 43 43 43 43 43 44 44 | 4* | 0000111 |
| 44 44 44 45 45 45 45 45 | 4t | 22233333 |
| 45 46 46 46 46 46 46 46 | 4f | 44444555555 |
| 46 46 46 46 47 47 47 47 | 4s | 66666666666777777777 |
| 47 47 47 47 47 48 48 48 | 4. | 888899999 |
| 48 49 49 49 49 49 50 50 | 5* | 00000111111111111 |
| 50 50 50 51 51 51 51 51 | 5t | 222222222233333333 |
| 51 51 51 51 51 51 51 52 | 5f | 4444445555555 |
| 52 52 52 52 52 52 52 52 | 5s | 66666666667777777 |
| 52 53 53 53 53 53 53 53 | 5. | 88888888888999999999 |
| 53 54 54 54 54 54 54 55 | 6* | 00000000000011111111111 |
| 55 55 55 55 55 55 | 6t | 22222222222223333333333 |
| … | 6f | 444444455555555 |
| | 6s | 6666666677777777777777 |
| | 6. | 889999999 |
| | 7* | 01111 |
| | 7t | 22222222333 |
| | 7f | 44444455 |
| | 7s | 666677 |
| | 7. | 88899 |
| | 8* | 00011 |
| | 8t | 2333 |
| | 8f | 5 |
| | 8s | 67 |
| | 8. | 8 |
| | 9* | 0 |
| | 9t | |
| | 9f | 4455 |
| | 9s | |
| | 93 | 8 |
| | High | 104; 10; 125 |

© Cengage Learning 2013

**Şekil 2.8** Cevaplama süresi verisi için gövde-ve-yaprak gösterimi

Gövde-ve-yaprak gösterimler özellikle iki farklı dağılımın karşılaştırılması için kullanışlıdırlar. Bu tür bir karşılaştırma iki dağılımın gövdenin zıt taraflarında gösterilmesiyle başarılır. Şekil 2.9 benim öğrettiğim deneysel yöntemler dersinde bayanların ve bayların aldıkları puanların gerçek dağılımını göstermektedir. Bunlar gerçek veridir. Dikkat ederseniz gövde kısmında 6* (60-64 için) ve 6. (65-69 için) kullanılmıştır. Ek olarak çizelgenin altındaki ham puanlara nasıl dönüştürüleceğiyle ilgili kodu da gözden kaçmamalıdır. Bu kod 14*11'in 4.1'i veya 410 değil 41'i gösterdiğini belirtmektedir. Son olarak şeklin erkekler ile kadınlar arasındaki performans farkını güzel bir biçimde ortaya koyduğuna da bakabiliriz.

| Erkek | Gövde | Kadın |
|---|---|---|
| | 3* | |
| 6 | 3. | |
| | 4* | 1 |
| | 4. | |
| | 5* | |
| | 5. | |
| 2 | 6* | 03 |
| 6. | 6. | 568 |
| 32200 | 7* | 0144 |
| 88888766666655 | 7. | 555556666788899 |
| 4432221000 | 8* | 0000011112222334444 |
| 7666666555 | 8. | 5566666666666677888888899 |
| 422 | 9* | 000000000133 |
| | 9. | 56 |
| Kod \|4*\|1 = 41 | | |

© Cengage Learning 2013

**Şekil 2.9** Deneysel metotlar dersindeki puanların yüzde olarak cinsiyete göre ayrı ayrı grafiği

## 2.5 Dağılımların Tanımlanması

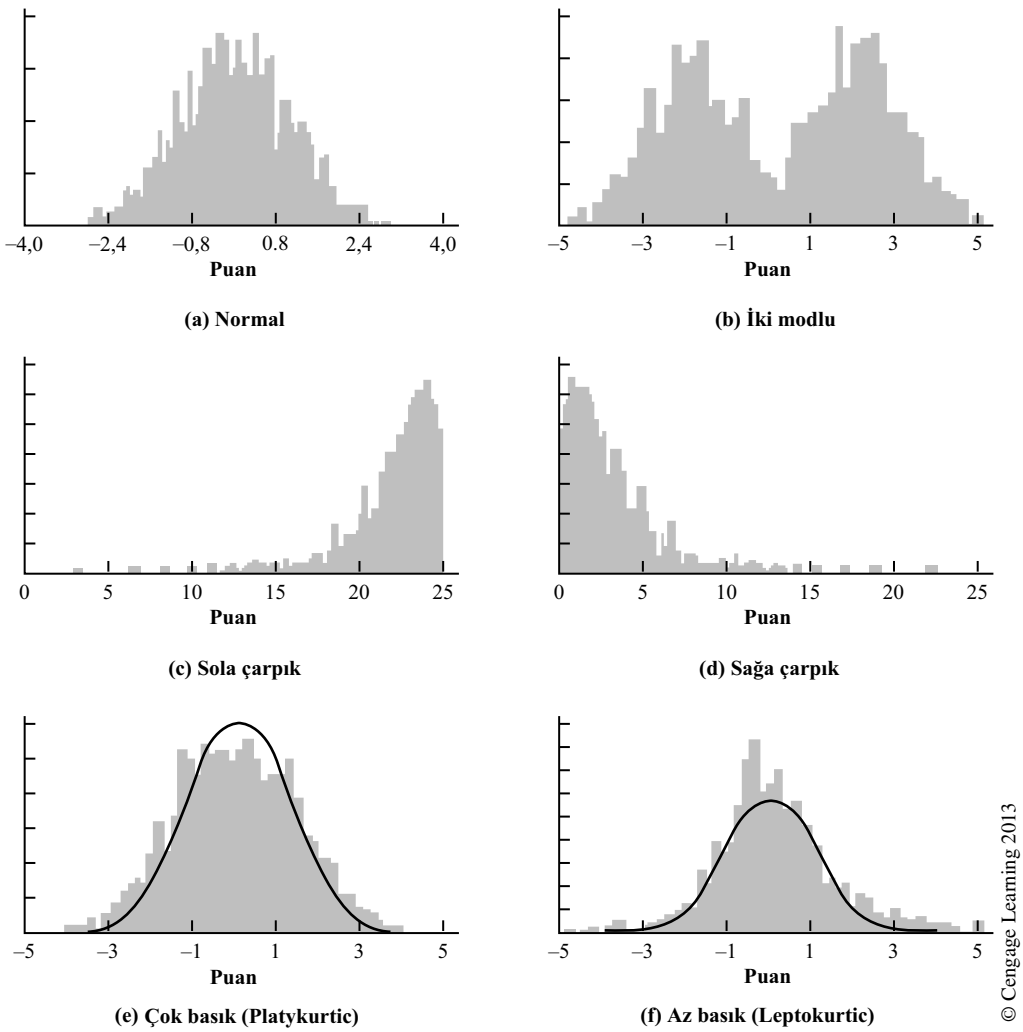
Şekil 2.1 ve 2.2'de gösterilen puan dağılımları her ne kadar tam olarak **simetrik** olmasalar da maksimum değere yükselen ve sonra düzgün bir biçimde azalan aşağı yukarı düzgün biçimli dağılımlardı. Ancak, bütün dağılımların hepsi merkezde tepe değere ulaşıp sonra da kenarlarda düzgün bir biçimde azalmadıkları için (Şekil 2.6'daki gövde-ve-yaprak gösterimine bakınız) farklı dağılımları betimlemede kullanılan terimleri anlamak önemlidir. Şekil 2.10(a) ve (b)'de gösterilen dağılımları ele alalım. Bunlar, belli şekillere sahip evrenlerden gelen bilgisayar tarafından üretilmiş veri kümelerinin grafikleridir. Bunlar ve Şekil 2.10'daki diğer dört grafik 1000 gözlem örneklemine dayanır ve küçük düzensizlikler tesadüfî değişkenliklerdir. Şekil 2.10(a) ve (b)'deki her iki dağılım da simetriktir çünkü merkezin her iki tarafı da aynı şekildedir.

Şekil 2.10(a)'da gösterilen dağılım daha sonra normal dağılım olarak adlandıracağımız dağılımdan elde edilmiştir. Şekil 2.10(b)'de gösterilen dağılım iki modlu olarak adlandırılır çünkü iki tepesi vardır. **İki modlu** (tepedeğerli) terimi tepeleri aynı boyda olsa da olmasa da iki baskın tepesi olan dağılımlar için kullanılır. Eğer dağılımın tek bir tepesi mevcutsa ona **tek modlu** denir. Dağılımdaki başlıca tepelerin sayısını nitelendirmek **modluluktur** (modality).

iki modlu
tek modlu
modluluk

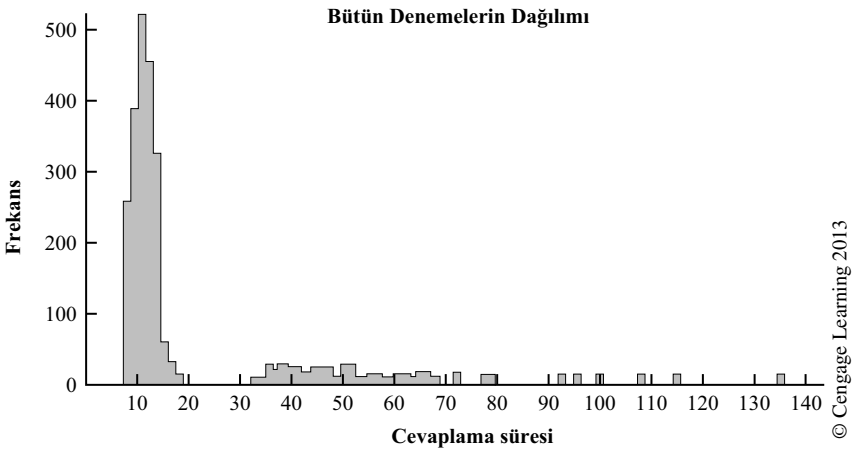
Şimdi de Şekil 2.10(c) ve (d)'yi ele alalım. Bu iki dağılımın simetrik olmadıkları açıkça görülmektedir. Şekil 2.10(c)'deki dağılımın kuyruğu sola doğru iken, Şekil 2.10(d)'nin kuyruğu sağa doğrudur. (c)'dekine **sola çarpık**, (d)'dekine **sağa çarpık** dağılım deriz. (Sağa çarpık olanın kuyruğu sağa, sola çarpık olanın kuyruğu sola doğrudur.) Asimetrikliğin veya **çarpıklığın** derecesini gösteren istatistiksel ölçüler mevcuttur, fakat davranış bilimlerinde yaygın olarak kullanılmazlar.

sola çarpık
sağa çarpık
çarpıklık




**Şekil 2.10** Frekans dağılımlarının şekilleri: (a) normal, (b) iki modlu, (c) sola çarpık, (d) sağa çarpık, (e) çok basık, (f) az basık (sivri)

Gerçek hayattan ilginç bir sağa çarpık, hafif iki modlu dağılım örneği Şekil 2.11'de görülmektedir. Bu veri kümesi Bradley (1963) tarafından deneklerden hafif bir ışık gördüklerinde butona basmaları istenerek toplanmıştır. Verinin çoğu saniyenin yüzde biri cinsinden yaklaşık olarak 7 ile17 arasında düzgün bir dağılım göstermektedir, ama sağa doğru 30 ile 70 arasında ufak ancak fark edilir veri noktaları vardır. Bu ikinci tabakadaki veri kümesi birinci denemede butona basmayı başaramayan deneklere aittir. Çalışmaya bu verinin dahil edilmesi dağılımın şeklini önemli ölçüde etkilemektedir. Böyle bir veri kümesine sahip bir araştırmacının belli bir maksimum değerden sonraki değerleri ayrı olarak değerlendirmesi -bu değerlerin cevaplama süresinin hızından çok psiko-motor davranışın doğruluğunu gösterdiği gerekçesiyle ayrı olarak değerlendirmesi- önerilir. Her ne kadar dağılımın daha iyi görünmesini sağlasak bile, bu kaçırılan cevapların analizlere dahil edilip edilmemesini sorgulamamız gerekir.




**Şekil 2.11** Bradley'nin cevaplama süresi verisine ait frekans dağılımı

Şekil 2.11'de gösterilen Bradley'nin verisiyle benim ürettiğim Şekil 2.1 ve 2.2'de görülen veri kümeleri arasındaki farkları göz önünde bulundurmak önemlidir. Her iki dağılım da sağa çarpık olmasına rağmen, benim veri kümelerim genellikle daha uzun yanıtlama süresi içermelerine rağmen ikinci bir veri tabakası söz konusu değildir. Aradaki fark, benim hangi butona basacağıma karar vermem iken, Bradley'nin çalışmasındaki deneklerin ışığı gördükleri anda tek bir butona basmalarıdır. Karar vermek zaman alır. Ayrıca kullandığım program yanlışlara değil, sadece doğru yanıtlara verilen süreleri kaydediyordu. Yapılan yanlış yanıtları düzeltme şansı yoktu. Bunları belirtmemin sebebi, veri kümelerinin toplanma biçiminden gelen gözle görülür farklara işaret etmektir.

basıklık

Bir dağılımın inceleyeceğimiz son özelliği **basıklık**tır. Basıklığın matematiksel bir tanımı vardır, ama temel olarak basıklık bir dağılımda merkezdeki puanların, alt ve üst uçlardaki (kuyruklar) ve omuzlardaki (kuyruklar ile merkez arası) puanlara göre ne kadar toplandığını gösterir. Şekil 2.10(e) ve (f)'de karşılaştırmayı daha iyi yapabilmek için veri kümelerinin histogramlarının üzerine normal dağılım eğrileri de çizilmiştir. Normal dağılım (Bölüm 3'te açıklanacaktır) **normal basık**tır **(mesokurtic)**. Normal dağılımın kuyruk kısımları ne çok ince ne de çok kalındır ve merkezde de ne çok az ne de çok fazla puan toplanmıştır. Eğer normal dağılım ile başlar, kuyruktaki ve merkezdeki puanları omuzlara doğru kaydırırsanız grafiğiniz düzleşir ve **çok basık (platykurtic)** diye adlandırılır. Şekil 2.10(e)'de bu açıkça görülmektedir, dağılımın merkez kısmı oldukça düzdür. Öte yandan, eğer omuzlardaki puanları merkeze ve kuyruklara doğru kaydırırsanız, dağılımın kuyrukları kalınlaşır ve sivrileşir. Bu tür grafikler **az basık (leptokurtic) (sivri)** olarak adlandırılır ve bir örneği Şekil 2.10 (f)'dir. Dikkat edilirse bu dağılımda merkezde ve kuyruklarda çok fazla puan olduğu gözlenir.[4]

normal basık

çok basık

az basık

Dağılımın şekli hakkında tam fikir sahibi olabilmek için, özellikle basıklık için, örnekleme ait veri kümesinin oldukça büyük olmasına dikkat etmek önemlidir. Otuz civarındaki örneklem büyüklükleri ile ancak veri kümesinin dağılımının kuyruklarında toplanıp toplanmadığını gözleyebiliriz veya dağılımın belli bir yönde büyük ölçüde çarpık olup olmadığını saptayabiliriz.

4 East Carolina Üniversitesi'nden Karl Wuensch'e basıklık ve çarpıklığı anlamamda yardımcı olan önerileri için teşekkür etmek isterim. Her ne kadar basıklık konusundaki ifadelerimden memnun olacağından emin olmasam da onun fikirleri burada yansıtılmıştır. Birkaç kişi hariç, istatistikçiler de dahil olmak üzere, gerekli hesaplamaları yapmadan, hiç kimse dağılıma bakarak bir dağılımın az mı yoksa çok mu basık olduğunu ayırt edemez.

Tartışmanın buraya kadar olan kısmında sayılardan hiç bahsedilmedi. Veri kümesinin nasıl organize edilebileceği veya dağılım olarak nasıl sunulabileceğini gördük ve dağılımları karakterize edecek simetriklik, basıklık, çarpıklık ve mod gibi çeşitli özelliklerinden bahsettik. Her ne kadar bu gibi bilgiler belli durumlarda faydalı olsalar bile, diğerlerinde yeterli değillerdir. Halen daha ortalama cevaplama süresini veya cevaplama sürelerinin bir denemeden diğerine göre ne kadar benzediğini veya farklı olduğunu bilmiyoruz. Bunun için, verimizi gerekli bilgileri içeren ölçüler kümesine indirgemeliyiz. Bunlar dağılıma ait, konumu veya merkeze eğilimi, dağılımı veya değişimi belirten ölçülerdir. Bu konular bölüm 2.8 ve 2.9'da ele alınacaktır. Fakat bu bölümlere geçmeden önce bu konularda kullanılacak sembol sistemi tanıtılacaktır.

## 2.6 Sembol ile Gösterim (Notation)

İstatistiksel teknikleri içeren herhangi bir konunun açıklanmasında, matematiksel işlemleri belirten bir sembol sistemi kullanılır. İstatistik konularının açıklanmasında ise standart bir sembol ile gösterim sisteminin olmadığını öğrendiğinizde şaşırabiliriz. Her ne kadar bu amaçla girişimlerde bulunulduysa da, herhangi iki istatistik ders kitabında aynı kavramlar için farklı sembollerin kullanıldığını görebiliriz.

Yaygın olarak kullanılan sembol sistemlerinin hem çok basitleri hem de çok karmaşıkları vardır. Karmaşık semboller duyarlılık (precision) içerirken kolayca anlaşılabilirliklerini yitirirler, öte yandan basit semboller ise kolayca anlaşılabilirler ama duyarlılıklarını kaybederler. Genellikle duyarlılık kaybı kazanılacak anlaşılabilirlik miktarıyla karşılaştırıldığında az miktarlarda olduğu için bu kitapta oldukça basit bir sembol sistemi kullanılacaktır.

### Değişkenlerin Sembolleri

Genellikle bir kural olarak değişkenler $X$ veya $Y$ gibi büyük harfle gösterilir. Harf ve alt indis ise o değişkene ait tek bir değeri temsil eder. Örneğin, [45, 42, 35, 23, 52] sayıları üçüncü sınıf öğrencilerinin nefeslerini kaç saniye tutabildiklerini göstersin. Bu puanların kümesi $X$ ile temsil edilir. Kümenin birinci elamanı (45) $X_1$ ile, ikincisi (42) $X_2$ ve diğerleri de benzeri biçimlerde gösterilirler. Hangisi olduğunu belirtmeden herhangi bir puandan bahsetmek istediğimizde $X_i$'yi kullanırız. Burada $i$ bu örnek için 1 ile 5 arasında değerler alabilir. Pratikte alt indis genellikle dikkat dağıtıcı olduğu için, bir karışıklık oluşmadığı sürece kullanılmaz.

### Toplama Sembolü

**sigma ( ∑ )**

İstatistikte en yaygın kullanılan sembollerden birisi toplama işlemini gösteren büyük Yunan harfi **sigma**'dır **(∑)**. Anlamı, "önündeki her şeyi topla"dır. Bu yüzden $\sum X_i$'nin anlamı "bütün $X_i$'leri topla"dır. Tam doğru olarak söylemek gerekirse, $X$'in N tane değerinin hepsini toplamak $\sum_{i=1}^{N} X_i$ ile gösterilir. Bu da i = 1'den i = $N$'e kadar bütün $X_i$ değerlerini toplamak anlamına gelir. Pratikte ise nadiren tam olarak kesinlikle ne yapılması gerektiğini belirtiriz ve çoğu durumda alt indisler kullanmayıp $X_i$'lerin toplamını belirtmek için $\sum X$ yazarız.

$\sum X$'nin birkaç değişik durumunun tam olarak anlaşılması önemlidir. Bunlardan biri $\sum X^2$'dir. Bu gösterim $X$ değerlerinin karelerinin toplamı olarak okunur (örnek, $45^2+42^2+35^2+23^2+52^2 = 8247$). Dikkat edilecek olursa $\sum X^2$ ile $(\sum X)^2$ birbirinden oldukça

**Tablo 2.4** Toplama sembolünü içeren işlemlerin gösterilmesi

| | Kaygı Puanı ($X$) | Kaçırılan Testler ($Y$) | $X^2$ | $Y^2$ | $X-Y$ | $XY$ |
|---|---|---|---|---|---|---|
| | 10 | 3 | 100 | 9 | 7 | 30 |
| | 15 | 4 | 225 | 16 | 11 | 60 |
| | 12 | 1 | 144 | 1 | 11 | 12 |
| | 9 | 1 | 81 | 1 | 8 | 9 |
| | 10 | 3 | 100 | 9 | 7 | 30 |
| Toplam | 56 | 12 | 650 | 36 | 44 | 141 |

$$\sum X = (10 + 15 + 12 + 9 + 10) = 56$$
$$\sum Y = (3 + 4 + 1 + 1 + 3) = 12$$
$$\sum X^2 = (10^2 + 15^2 + 12^2 + 9^2 + 10^2) = 650$$
$$\sum Y^2 = (3^2 + 4^2 + 1^2 + 1^2 + 3^2) = 36$$
$$\sum (XY) = (7 + 11 + 11 + 8 + 7) = 44$$
$$\sum XY = (10)(3) + (15)(4) + (12)(1) + (9)(1) + (10)(3) = 141$$
$$\sum (X)^2 = 56^2 = 3136$$
$$\sum (Y)^2 = 12^2 = 144$$
$$(\sum (X - Y))^2 = 44^2 = 1936$$
$$(\sum X)(\sum Y) = (56)(12) = 672$$

© Cengage Learning 2013

farklıdır. $(\sum X)^2$ ifadesi $X$'lerin toplamını bulup sonra karesini almayı ifade eder. Bu da $(\sum X)^2 = (45+42+35+23+52)^2 = (197)^2 = 38809$. Bütün durumlarda uygulanan genel kural parantezler içindeki işlemleri parantezlerin dışındakilerden önce uygulamaktır. Bu yüzden $(\sum X)^2$ için önce X'leri toplar, sonra sonucun karesini alırız. Buna karşın $\sum X^2$ için önce $X$'lerin karelerini alır sonra toplarız.

İki değişken ($X$ ve $Y$) ile çalışırken kullanılan diğer yaygın bir gösterim $\sum XY$'dir. Bu da $X$ ve $Y$'nin karşılık gelen değerlerini çarpıp sonra bu çarpım değerlerinin toplanması anlamına gelir. Bu ve diğer gösterimlerin kullanımı aşağıdaki örnekte gösterilecektir.

Beş öğrencinin kaygı düzeylerini ($X$) puanları ile ve okula gelmedikleri için kaçırdıkları sınavların sayısını da ($Y$) ile kaydettiğimiz bir deney düşünelim. Bu deney için veri ve basit toplama işlemleri Tablo 2.4'de gösterilmiştir. Yapılan işlemlerin bazıları önceden tartışıldı, diğerleri de sonraki bölümlerde tartışılacaktır.

## Çift Alt İndisler

$X$'in tam olarak hangi değerini kastettiğimizi kesin bir biçimde belirtmek için genellikle iki veya daha fazla indis yaygın olarak kullanılır. Örneğin, elimizde Tablo 2.5'teki veri kümesinin olduğunu varsayalım. Eğer $i$'yinci satır ve $j$'yinci sütundaki veriyi belirtmek istersek bunu $X_{ij}$ ile gösteririz. Yani ikinci günün üçüncü denemesindeki puan $X_{2,3} = 13$'tür. Bazı gösterim sistemleri bunun için $\sum_{i=1}^{2} \sum_{j=1}^{5} X_{ij}$ kullanırlar, bu $i$'nin 1 ile 2 ve $j$'nin 1 ile 5 arasındaki değerlerini aldığı $X_{ij}$'lerin toplamıdır. Bazı istatistik kitapları bu gösterim sistemini kullandıkları için bunların da bilinmesi yararlı olur. Ancak bu kitapta daha basit ama daha az kesin olan yerlerde mümkün oldukça $\sum X$, gerek oldukça da $\sum X_{ij}$ kullanılacak ama $\sum\sum X_{ij}$ asla kullanılmayacaktır.

**Tablo 2.5** Sembolleri açıklayan hipotetik veri

| | | Deneme | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | Toplam |
| Gün | 1 | 8 | 7 | 6 | 9 | 12 | 42 |
| | 2 | 10 | 11 | 13 | 15 | 14 | 63 |
| | Toplam | 18 | 18 | 19 | 24 | 26 | 105 |

© Cengage Learning 2

En temel istatistiksel işlemleri öğrenmek için bile bu gösterim sistemini tam olarak anlamanız gerekmektedir. Tablo 2.4'ü yapılan işlemleri tam olarak anlayana kadar iyice inceleyiniz.

## 2.7 Merkeze Eğilme Ölçüleri

Veri kümesinden bazı sonuçlar çıkarmaya başlayabilmek için kümenin nasıl görselleştirilebileceğini inceledik. Veri kümesinin grafiklerinin çizilmesi dağılımın genel şekli ve içerdiği sayıların büyüklükleri hakkında genel görsel bir fikir verir. Bu kısımda dağılımın "merkezini" temsil etmede kullanılan birkaç istatistiği inceleyeceğiz. Bu istatistiklere **merkeze eğilme ölçüleri** adı verilir. Bir sonraki kısımda ise bir adım ileri giderek bu gözlemlerin merkezin etrafında nasıl dağıldıklarını (yayıldıklarını) inceleyeceğiz, ancak bu kısımda öncelikle dağılımın merkezini belirlemeyi ele almalıyız.

merkeze eğilme ölçüleri

konum ölçüleri

Merkeze eğilme ölçüleri ve bazen de **"konum ölçüleri"** terimleri dağılımın merkezinin yerini belirtir. Bu ölçüler, veriyi ne ölçüde kullandıkları, özellikle uç değerleri konusunda birbirlerinden farklılaşırlar, fakat hepsi bize dağılımın merkezinin nerede olduğu hakkında bir şeyler söylemeye çalışırlar. Üç temel merkeze eğilme ölçüsünden, tepedeğer (mod), birkaç veriye dayanır; ortanca (medyan), verinin çoğu kısmını göz ardı eder; ortalama (mean), bütün veri kümesinin tamamından hesaplanır. Bu ölçüleri en az kullanılandan (genellikle en az yararlı), tepedeğerden, başlayarak sırasıyla ele alacağız.

### Tepedeğer (Mod)

tepedeğer

**Tepedeğer (mod)** basit olarak en sık gözlenen puan olarak tanımlanabilir. Yani bir dağılımda en tepe noktaya karşılık gelen $X$ değeri tepedeğerdir. Eğer iki tane birbirini takip eden değer en yüksek ve eşit frekansta iseler genellikle bu iki değerin ortalaması alınır ve bu değere tepedeğer (mod) denir. Öte yandan eğer, iki tane birbirini izlemeyen değer eşit (veya birbirine yakın) frekansta ise bu, dağılımın çift tepedeğerli (çift modlu) olduğunu belirtir ve her iki değeri de tepedeğer olarak rapor ederiz. Örneğin, elektronik oyun oynamak için harcanan süre hemen hemen çift tepedeğerlidir (modludur). Birinci tepe 0-9 dakika arasında ikinci tepe 40-49 dakika arasındadır (Şekil 2.7'ye bakınız).

### Ortanca (Medyan)

Ortanca (Medyan)

**Ortanca (Medyan)**, veri kümesindeki sayılar sıraya konulduğunda, tam ortadaki (altında ve üstündeki sayıların %50'sinin bulunduğu) değerdir. Bu tanıma göre ortanca (medyan) 50'nci yüzdeliktir[5]. Örneğin, (5, 8, 3, 7, 15) sayılarını ele alalım. Eğer sayılar sıralanırsa

5 Belli bir yüzdelik, puanların belirtilen yüzdenin altında veya üstünde o noktadaki değeriyle tanımlanır.

(3, 5, 7, 8, 15) ortadaki puan olan 7 ortancadır. Ancak elimizde çift sayıda puan varsa, örneğin (5, 11, 3, 7, 15, 14) bunları sıraya koyduğumuzda (3, 5, 7, 11, 14, 15) elde ederiz, bu sayıların %50'nciye denk geleni (orta noktası) yoktur. Aslında bu nokta 7 ile 11'in arasındadır. Böyle durumlarda genellikle ortadaki iki puanın (7 ve 11) ortalaması ortanca (9) olarak alınır.[6]

ortancanın konumu

Ortancanın konumu terimine ihtiyacımız var. *N* tane puanın bulunduğu bir veri kümesinde **ortancanın konumu** aşağıdaki gibi tanımlanır:

$$\text{Ortancanın Konumu} = \frac{N+1}{2}$$

Beş sayı için ortancanın konumu = (5+1)/2 = 3, bu da bize sıralanmış veri kümesinde üçüncü sayının ortanca olduğunu gösterir. On iki tane sayı için ortancanın konumu = (12+1)/2 = 6,5; ortanca altıncı ve yedinci sayının ortalamasıdır.

Tablo 2.2'de görülen cevaplama süresi veri kümesi için ortancanın konumu = (300+1) / 2 = 150,5'tir. Veri sıralandığında 150'nci veri 59 ve 151'nci veri 60'tır. Yani ortanca (59 + 60) / 2 = 59,5'tir. Elektronik oyun için 100 veri vardı, yani ortancanın konumu 50,5'tir. Şekil 2.7'den 50'nci verinin 44 ve 51'nci verinin 46 olduğunu görebiliriz. Buradan ortancayı 44 ve 46'nın ortalamasını alarak 45 buluruz.

## Ortalama

ortalama

En yaygın kullanılan ve biraz daha açıklamaya ihtiyaç duyulan merkeze eğilme ölçüsü **ortalama**dır. Ortalama ($\overline{X}$) puanların toplamının puanların sayısına bölümüdür ve genellikle $\overline{X}$ ("*X* çizgi" olarak okunur) ile gösterilir.[7] Toplam sembolü kullanılarak ortalamanın tanımı aşağıdaki gibi yazılır:

$$\overline{X} = \frac{\sum X}{N}$$

Burada $\sum X$ bütün *X* değerlerinin toplamı ve *N* de *X*'lerin sayısıdır. Örnek olarak 3, 5, 12 ve 5'in ortalaması şu şekildedir:

$$\frac{3+5+12+5}{4} = \frac{25}{4} = 6,25$$

Cevaplama süresi veri kümesi için Tablo 2.2'deki gözlemlerden elde edilen sayıların toplamı 18078'dir. Bu sayıyı *N* = 300'e böldüğümüzde 18078/300 = 60,26 buluruz. Dikkat edilirse bu değer ortanca değer olan 59,5'e yakındır. Ortalama ve ortancanın yakın olduğu durumlarda dağılım yaklaşık olarak simetriktir (bakınız simetriklik tanımı). Ayrıca bu değer, tepedeğer aralığıyla (60-64) da uyumludur.

---

6 Ortancanın tanımı istatistikçilerin tartışmayı sevdikleri diğer konulardan biridir. Burada verilen tanım en çok tercih edilendir. Aynı zamanda ortancanın 50. yüzdelik olduğu tanımıyla paraleldir. Öte yandan ortancanın sıralanmış veri kümelerinde en ortada olanı olduğunu söylemekten mutlu olanlar da vardır (*N* tek olduğunda) veya ortadaki iki sayının ortalaması (N çift ise). Bu tür tartışmaları okumak tartışılacak önemli bir konu yokken fakülte toplantısına gitmeye benzer. Konu ne kadar önemsiz ise o kadar fazla söylenecek şey vardır.

7 Amerikan Psikolji Birliği $\overline{X}$ yerine M'yi tercih ediyor, ama ben uzun yıllardır $\overline{X}$ kullandım ve bundan sonra da böyle yapmaya devam edeceğim. İstatistik camiası da genel olarak benimle aynı görüşü paylaşıyor, bu yüzden kitapta $\overline{X}$ kullanacağız.

## Tepedeğer (Mod), Ortanca (Medyan) ve Ortalamanın Avantajları ve Dezavantajları

Ortalama ve ortanca sadece dağılım simetrik olduğunda eşit ve sadece dağılım simetrik ve tek tepedeğerli olduğunda üç ölçü de eşit olacaktır. Ele alacağımız hemen hemen tüm durumlar dahil diğer bütün durumlarda belli bir merkeze eğilme ölçüsünün seçilmesi gerekir. Merkeze eğilme ölçüsü seçmek için iyi bir kural yoktur, fakat üçü arasından mantıklı bir seçim yapmak mümkündür.

### Tepedeğer (Mod)

Mod en sık gözlenen puandır. Bu tanıma göre mod, gerçekte gözlenmiş bir puandır, buna karşın ortalama ve bazen de ortanca veri arasında gözlenmeyebilir. Ayrıca tepedeğerin en belirgin avantajı en çok sayıda insanı temsil etmesidir. Küçük bir dükkanı işleten bir kişinin tepedeğer üzerinde yoğunlaşması işe yarayacaktır. Eğer müşterilerinin %80'i çok büyük ekonomik aile boyu deterjan ve %20'si de küçük tek kişilik olanı istiyorsa, diğer merkeze eğilme ölçülerine yönelerek sadece normal boyu stoklamak bu dükkan sahibi için akıllıca olmayacaktır.

### Ortanca (Medyan)

**Ortancanın (Medyanın)** tepedeğer ile paylaştığı başlıca avantajı uç noktalardan etkilenmemesidir. Örneğin, hem (5, 8, 9, 15, 16) hem de (0, 8, 9, 15, 206) sayılarının ortancası 9'dur. Uç noktaların nadiren ortaya çıktığı, fakat bir önem arz etmediği bazı çalışmalarda ortancanın bu özelliği kullanışlı bulunur. Örneğin, bir fare kısa bir koşu yolunu ortalama bir ile iki saniye arasında koşabilir. Arada bir bu aynı fare açıklanamayan bir şekilde yarı yolda durur yere yatar, burnunu fotosellere sokar, kendini kaşır. Bu gibi durumlarda koşu yolunun diğer ucuna farenin 30 saniyede veya 10 dakikada ulaştığının pratikte hiçbir anlamı yoktur. Hatta deneyi yapan kişi fareyi yerinden kaldırmak için onu kalemle dürtüklemek zorunda kalabilir. Eğer bir fareyi günde üç kez koşturup ve (1,2; 1,3 ve 20 saniye) sürelerini elde ettiğimizde, farenin görev hakkında ne bildiği konusunda (1,2; 1,3 ve 136,4 saniye) sürelerini elde ettiğimiz durum aynı anlama gelir. Her iki durumda da ortanca 1,3'tür. Ancak her iki durumun günlük ortalamaları birbirinden oldukça farklıdır (7,5 ve 46,3 saniye). Bu problem sıklıkla deneyi yapanların günlük ortalama yerine ortancayı tercih etmelerine sebep olur.

### Ortalama (Mean)

Üç temel merkeze eğilme ölçüsü arasından büyük farkla en yaygın olarak kullanılan **ortalama**dır. Çoğu insan için istatistik, ortalamanın çalışılmasıyla hemen hemen eş anlama gelir desek durumu çok da abartmış sayılmayız.

Önceden de gördüğümüz gibi ortalamanın belli dezavantajları vardır: Uç değerlerden etkilenir, değeri veri kümesindeki sayılar arasında olmayabilir ve yorumlanması eldeki verinin en azından eşit aralıklı ölçekte olmasını gerektirir. Eğer ortalama bütün bu belirtilen dezavantajlara sahipse, siz kibarca bazı diğer istatistik kavramlar gibi, ortalamanın da yıllardır hiç adı duyulmayan kavramlar arasına karışmasını önerme eğilimi gösterebilirsiniz. Ancak ortalama daha güçlü materyalden oluşmuştur.

Ortalamanın, dezavantajlarına göre daha ağır basan birkaç önemli avantajı vardır: Tarihsel açıdan baktığımızda, muhtemelen bunların en önemlisi ortalamanın cebirsel olarak hesaplanabilmesidir. Başka bir deyişle, ortalamayı bir eşitlikte kullanabilir ve cebirin normal kurallarını kullanarak onu hesaplayabiliriz. Özellikle ortalamayı tanımlayan eşitliği yazabilmek mümkündür. Öte yandan tepedeğer veya ortancanın standart bir eşitliği yazılamaz, standart cebir işlemleri kullanarak bu ölçüleri hesaplamanın bir yolu yoktur. Deza-

vantajlarına rağmen, bu sebeplerden dolayı ortalama yaygın bir şekilde uygulanmaktadır. Ortalamanın ikinci önemli avantajı ise, evren ortalamasının kestirimi olarak kullanılmasıyla ilgili birkaç istenen özelliğinin bulunmasıdır. Özellikle, eğer belli bir evrenden birçok örneklem seçersek, elde edilen örneklem ortalamaları örneklem ortancaları ve tepedeğerlerine göre evrenin merkeze eğilme kestirimi olarak daha kararlıdır (az değişim gösterir). Örneklem ortalamasının, evren ortalamasının daha iyi bir kestirici olması tepedeğer veya ortancaya göre daha yaygın olarak kullanılmasının başlıca sebebidir.

### Kırpılmış Ortalamalar

**Kırpılmış Ortalama**

**Kırpılmış ortalama**lar dağılımın her iki ucundaki belli bir yüzdede verinin atılmasıyla hesaplanan ortalamalardır. Örneğin, eğer elimizde 100 gözlem varsa ve biz %10 kırpılmış ortalama hesaplamak istiyorsak, en yüksek 10 ve en düşük 10 puanı çıkarıp, kalanlarla ortalamayı hesaplarız. Bu fikir yeniden moda olmaya başlayan eski bir fikirdir ve en güçlü savunucusu da Rand Wilcox'dır (Wilcox, 2003, 2005).

Örneklemi kırpmanın birçok sebebi vardır. Bölüm 1'de söz edildiği ve bu kitap boyunca tekrar tekrar ifade edileceği gibi örneklem ortalamasını hesaplamanın temel amacı elimizdeki örneklemden evren ortalamasını kestirmektir. Eğer iyi bir kestirim istiyorsanız, bir örneklemden diğerine çok az değişmesini istersiniz. (İlerideki bölümlerde sözü edilecek terimi kullanırsak, standart hatası küçük bir kestirim isteriz). Yayılımı çok fazla olan bir örneklemimiz mevcutsa, yani çok yüksek ve çok düşük puanlarımız varsa, bizim örneklemimizin ortalaması evren ortalamasının iyi bir kestirimi olmayacaktır. Uç değerleri örneklemden kırptığımızda evren ortalaması kestirimimiz daha kararlı olacaktır. Bu genellikle bir avantaj olsa bile, aynı zamanda yararlı bilgiyi atmadığımızdan da emin olmamız gerekir.

Örneklemi kırpmamızın diğer bir sebebi, çarpıklık problemini kontrol etmektir. Eğer oldukça çarpık bir dağılımımız varsa bu uç değerler ortalamayı kendilerine doğru çekerler ve daha kötü bir evren ortalamasına yol açarlar. Kırpmanın bir sebebi de bu uç değerlerin etkisini ortadan kaldırmaktır. Ancak Şekil 2.10'da gösterilen, Bradley'nin (1963) cevaplama verisini ele alırsak, uzun cevaplama sürelerinde muhtemelen cevaplayıcıların butonu ıskalamalarından kaynaklandığını söyleyebiliriz, bu yüzden bunları tam olarak cevaplama süresi ile ilişkilendirmeyebiliriz ve verinin içinden çıkarabiliriz; fakat aynı sayıda gözlemi ölçümlerin diğer ucundan da gerçekten çıkarmak istiyor muyuz?

Kırpma ile ilgili problemler üzerine Wilcox birçok çalışma yaptı ve ben onun lâyıkıyla hak edilmiş ününe saygı duyuyorum. Ayrıca günümüz literatüründe de tartışıldıkları için öğrencilerin kırpılmış ortalamaları öğrenmeleri gerektiğini düşünüyorum. Fakat kullanımlarını savunma konusunda Wilcox kadar ileri gideceğimi düşünmüyorum. Ancak benim bu düşüncelerim insanları bu konuyu ciddi bir şekilde ele almaktan alıkoymamalıdır ve ben Wilcox'ın kitabını öneririm (Wilcox, 2003).

## 2.8 Değişme (Yayılma) Ölçüleri

Bir önceki kesimde dağılımın merkezine ilişkin birkaç ölçüyü ele aldık. Ancak dağılım için ortalama bir değer (bu tepedeğer, ortanca veya ortalama olabilir) hikâyenin tamamını vermeyebilir. Her bir gözlemin ortalamadan ne ölçüde saptığını veya ne kadar eşit olduğu

veya nerelerde toplandığını gösterecek bazı ek ölçü veya ölçülere ihtiyaç duyarız. Ortalama bir değer, puanların çoğunun genel konumunu yansıtabilir veya puanlar geniş bir değer aralığında yayılabilirler ve bu ortalama değer tüm gözlemlerin iyi bir temsilcisi olmayabilir. Öğrencilerin hepsinin yaklaşık olarak aynı notu aldıkları ve bazılarında öğrencilerin en iyi ile en kötü arasında değişen puanlar aldığı sınavlar görülmüştür. Bu iki sınav durumu arasındaki farklara karşı gelen ölçüler, tepedeğer, ortanca veya diğer noktalar etrafında değişim, **değişkenlik** veya **yayılma**dan bahsettiğimizde aklımızdan geçenlerdir. Bu bölümde genel olarak özellikle ortalama civarındaki değişmeden söz edilecektir.

değişkenlik
yayılma

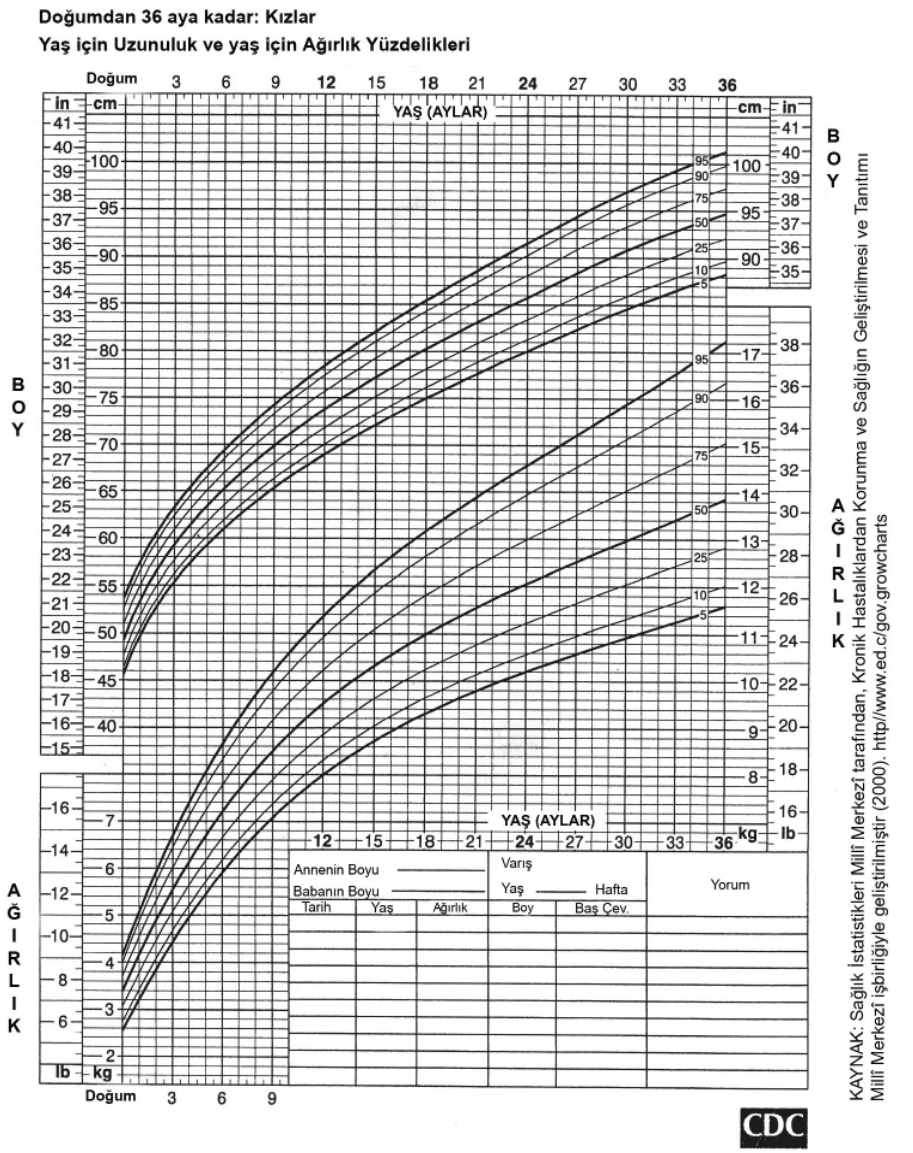
Değişkenliği örneklemek için kullanılacak örnek Weaver (1999) tarafından tavsiye edilmiştir ve bu örnek eminim hepinizin aşina olduğu standart yeni doğan gelişim grafiğidir. Grafik Şekil 2.12'de görülmektedir. Alt kısmında ise doğum ile 36 ay arası kızların normal ağırlıkları görülebilir. 50 ile etiketlenen koyu çizgi her bir yaşta ortalama ağırlığı temsil eder. Her iki taraftaki çizgiler ise dağılımın orta yarısının içinde bulunmasını beklediğimiz sınırları gösterir; her iki doğrultuda ilerledikçe görülecek bir sonraki iki çizgi ise sırasıyla çocukların orta %80'lik ve orta %90'lık sınırlarını gösterir. Bu şekilden yola çıkarak çocukların yaşı arttıkça yayılımın arttığı kolayca görülebilir. Çoğu yeni doğanın ağırlığı ortalama civarındaki bir pound içinde iken üç yaşındakilerin ağırlıkları ortalama etrafında her iki yönde beş pound olarak yayılmaktadır. Açıkça görüldüğü gibi, ortalama da artmaktadır ancak biz burada yayılma ile ilgileniyoruz.

İkinci örneğimiz olarak Langlois ve Roggman (1990) tarafından yüzlerin algılanan çekiciliği üzerine toplanmış ilginç bazı veri kümelerini dikkate alacağız. Bir dakikalığına çekici olarak düşündüğünüz bazı yüzleri düşünün. Bu yüzler olağandışı özelliklere mi sahipler yoksa olağan özelliklere mi? Langlois ve Roggman, yüzleri nelerin çekici yaptığını araştırmakla ilgileniyorlardı. Bu amaçla, öğrencilere bilgisayar tarafından üretilmiş yüz resimleri gösterdiler. Bu resimlerin bazıları dört farklı insanın resimlerinin ortalaması bir bileşke haline getirilerek oluşturulmuşlardı. Bu resimler Set 4 olarak etiketlenmiştir. Diğer resimler (Set 32) 32 farklı insanın çekilmiş fotoğraflarının ortalamasıyla oluşturulmuşlardır. Öngörüleceği gibi, dört kişinin ortalamasını aldığınızda bileşkede bireysellik için biraz yer vardır. Örneğin, bazı bileşkeler sivri bir yüze sahipken diğerleri yuvarlak bir yüze sahiptir. Ancak, 32 kişinin ortalaması çok "ortalama" sonuçlar verir. Burunlar ne çok uzun ne çok kısa; kulaklar başa ne çok yakın ne çok uzaktır. Öğrencilerden, ortaya çıkan resimleri incelemeleri ve her birine çekicilik bakımından 5 üzerinden puan vermeleri istenmiştir. Araştırmacılar öncelikle Set 4'teki yüzlerin ortalama puanlarının ve Set 32'deki yüzlerin ortalama puanlarından daha az olup olmadığı ile ilgileniyorlardı. Ayırt edici özelliklere sahip yüzlerin sıradan yüzlere göre daha az çekici bulunduğu sonucuna ulaştılar. Ancak biz bu bölümde daha çok yüzlere verilen puanların benzerlik derecesi ile ilgileniyoruz. Öngörümüz, 32 yüzün bileşkesinin dört yüze göre daha homojen olacağı ve bu nedenle puanların birbirine daha yakın olacağı yönündedir.

Veri kümesi Tablo 2.6'da görülmektedir.[8] Tablodan, Langlois ve Roggman'ın Set 32 yüzlerinin Set 4 yüzlerinden daha çekici olarak puanladığı görülebilir. (Ortalamalar sırasıyla 3,6 ve 2,64). Fakat 32 yüzün bileşkesinin dört yüzün bileşkesine göre daha homojen olduğu görülür. Şekil 2.13 bu veri kümelerinin standart histogramlarını göstermektedir.

---

8 Bu veri kümesi Langlois ve Roggman'ın topladıkları gerçek veri değil, ama orijinal veri ile aynı ortalama ve standart sapmaya sahip olacak biçimde üretilmiş veridir. Langlois ve Roggman her bir sette altı bileşke fotoğraf kullanmışlardır. Ben bu bölümdeki amacıma daha iyi hizmet etmesi için her bir sette 20 fotoğraf kullandım. Bu veriden çıkaracağınız sonuçlar, orijinal veriden çıkaracağınız sonuçlarla aynıdır.




**Şekil 2.12** Yeni doğan ağırlıklarının yaşın fonksiyonu olarak grafiği

Her ne kadar Şekil 2.13'ten dört fotoğrafın bileşkesindeki değişkenliğin 32 fotoğrafınkinden daha büyük olduğunu görsek de değişkenlikteki farkı yansıtacak belli bir ölçüye ihtiyacımız var. Bunun için birçok ölçü kullanılabilir, en basitinden başlayarak sırasıyla bunlar açıklanacaktır.

**Tablo 2.6** Bileşke yüzlerin çekicilik puanları

| Set 4 | | Set 32 | |
|---|---|---|---|
| Resim | 4 yüzün bileşkesi | Resim | 4 yüzün bileşkesi |
| 1 | 1,20 | 21 | 3,13 |
| 2 | 1,82 | 22 | 3,17 |
| 3 | 1,93 | 23 | 3,19 |
| 4 | 2,04 | 24 | 3,19 |
| 5 | 2,30 | 25 | 3,20 |
| 6 | 2,33 | 26 | 3,20 |
| 7 | 2,34 | 27 | 3,22 |
| 8 | 2,47 | 28 | 3,23 |
| 9 | 2,51 | 29 | 3,25 |
| 10 | 2,55 | 30 | 3,26 |
| 11 | 2,64 | 31 | 3,27 |
| 12 | 2,76 | 32 | 3,29 |
| 13 | 2,77 | 33 | 3,29 |
| 14 | 2,90 | 34 | 3,30 |
| 15 | 2,91 | 35 | 3,31 |
| 16 | 3,20 | 36 | 3,31 |
| 17 | 3,22 | 37 | 3,34 |
| 18 | 3,39 | 38 | 3,34 |
| 19 | 3,59 | 39 | 3,36 |
| 20 | 4,02 | 40 | 3,38 |
| Ortalama: 2.64 | | Ortalama = 3.26 | |

© Cengage Learning 2013

## Açıklık (Range)

**açıklık (ranj)**

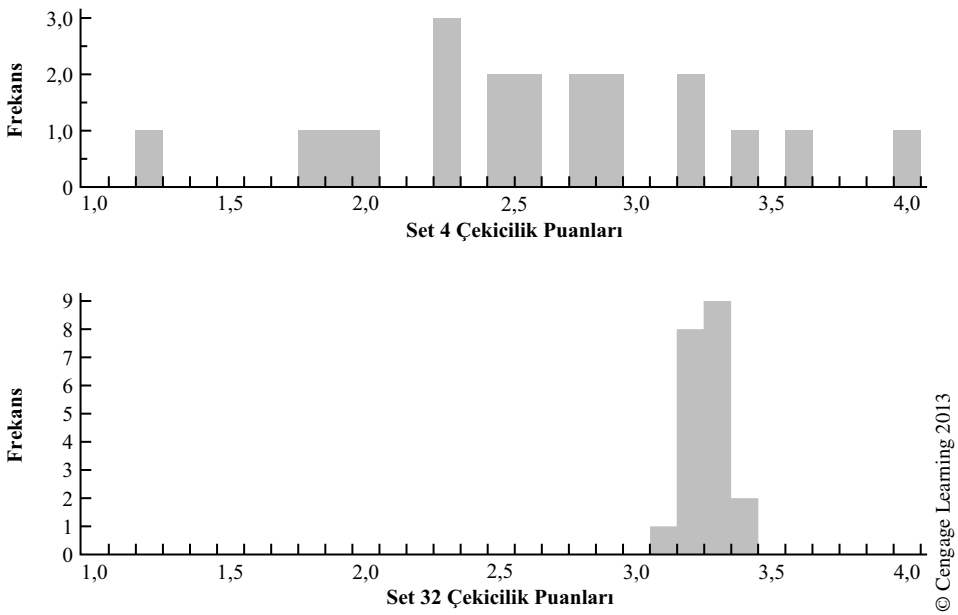
**Açıklık (ranj)**, veri kümesindeki en yüksek puan ile en düşük puan arasındaki uzaklığın ölçüsüdür. Bizim veri kümemizde, Set 4 için (4,02 – 1,20) = 2,82 birim; Set 32 için (3,38 – 3,13) = 0,25 birimdir. Ranj çok yaygın kullanılan bir ölçüdür. Örneğin, günlük hayatımızda "kırmızı biberin fiyatı 0,99 $ ile 3,99 $ arasında değişiyorsa kırmızı biber fiyatının açıklığı 3 $'dır. Ancak açıklık tamamen uç değerlere bağlı olduğundan, çok alışılmadık uç değerler varsa açıklık bundan çok etkilenir, bu nedenle açıklık değişkenlik hakkında yanıltıcı bir resim ortaya koyabilir.

## Çeyrekler Arası Açıklık (Interquartile Range) ve Diğer Açıklık İstatistikleri

**çeyrekler arası açıklık**

**birinci çeyrek**
**üçüncü çeyrek**
**ikinci çeyrek**

Çeyrekler arası açıklık, açıklığın uç değerlere bağlı sorunlarının üstesinden gelme çabalarının sonucudur. **Çeyrekler arası açıklık**, dağılımın en üstteki %25 ile en alttaki %25'i çıkarıldıktan sonra kalan kısmının açıklığıdır. Dağılımın en alttaki %25'ini kesen noktaya **birinci çeyrek** denir ve genellikle $Q1$ ile gösterilir. Benzer biçimde dağılımın en üstteki %25'ini kesen noktaya da **üçüncü çeyrek** adı verilir ve $Q3$ ile gösterilir. (Ortanca **ikinci çeyrek**tir, $Q2$). Birinci çeyrek ve üçüncü çeyrek arasındaki fark ($Q_3 - Q_1$) **çeyrekler arası açıklık**tır. Elimizdeki çekicilik veri kümesinin en düşük beş puanını ve en yüksek beş puanını çıkararak bu kümenin çeyrekler arası açıklığını hesaplayabiliriz. Bu durumda




**Şekil 2.13** Çekicilik puanları dağılımları

çeyrekler arası açıklık Set 4 için 0,58 ve Set 32 için 0,11'dir. Çeyrekler arası açıklığın kutu grafiği olarak bilinen kullanışlı bir grafik yöntemde önemli bir yeri vardır. Bu yöntem Bölüm 2.10'da tartışılacaktır.

Çeyrekler arası açıklık, açıklığın yaşadığı sıkıntıların tersi sıkıntılara sahiptir. Bunlardan başlıcası çeyrekler arası açıklığın çok fazla veriyi göz ardı etmesidir. Eğer bir fotoğraf kümesinin diğerine göre daha değişken olup olmadığını bilmek istiyorsak, en uç değerleri ve ortalamadan en çok uzakta ve ortalamadan çok farklı olan değerleri göz ardı etmek pek mantıklı gelmeyebilir.

Çeyrekler arası açıklığı hesaplarken dağılımın en alt ve en üst %25'ini eleme konusunda kutsal bir neden yoktur. Aslında istediğimiz herhangi bir yüzdeyi, mantığını kendimize ve diğer kişilere açıklayabildiğimiz sürece eleyebiliriz. Gerçekte elemek istediğimiz, çalışmak istediğimiz değişkenlik değil, hata sonucu veya olağan dışı ortaya çıkması olası puanlardır.

**kırpılmış örneklem**

**Winsorize örneklem**

Önceki bölümlerde kırpılmış ortalamaları hesaplamak için kırpılmış örneklemlerden bahsettik. Kırpma işlemi büyük uç değerlerin bulunduğu dağılımlarda ve çarpık dağılımlarda değerli bir yaklaşım olabilir. Fakat değişkenliği kestirmek için **kırpılmış örneklem** kullandığımızda **Winsorize örneklem**e dayalı değişkenliği kullanırız. Örneğin %10'luk Winsorize örneklemi, puanların en düşük %10'unu çıkarıp, yerlerine geriye kalan en küçük puanın kopyalarını koyarak ve sonra puanların en yüksek %10'unu çıkarıp yerlerine geri kalan en büyük puanın kopyalarını koyarak oluştururuz. Daha sonra düzenlenmiş veri kümelerinde değişme ölçüsünü hesaplarız.

## Ortalama Sapma

Eğer puanların ortalama etrafında nasıl dağıldıklarını (örneğin, ortalamadan sapmasını) ölçmek istiyorsak ilk başta bütün sapmaları $(X_i - \bar{X}_i)$ elde edip sonra da ortalamalarını almak yapılacak en mantıklı şey olarak görülebilir. Mantık çerçevesinde puanlar daha geniş dağıl-

dıkça, sapmaların daha büyük olacağını ve bu yüzden sapmaların ortalamasının daha büyük olacağını düşünebilirsiniz. Ancak, hisleriniz sizi burada yanlış sonuca sevk etti. Eğer ortalamadan sapmaları hesaplarsanız, bazı puanlar ortalamanın üstünde olup sapma pozitif olurken bazı puanlar da ortalamanın altında ve sapma negatif olacaktır. Sonunda pozitif ve negatif sapmalar birbirini götürecek ve sapmaların toplamı her zaman sıfır olacaktır. Bu fikir bizi pek bir yere götürmeyecektir.

## Ortalama Mutlak Sapma

ortalama mutlak sapma

Ortalama sapmalardan kullanışlı bir şeyler elde etme zorluğu hakkında düşünürken, problemi sapma değerlerinin mutlak değerini alarak çözmeyi önerebilirsiniz. (Bir sayının mutlak değeri o sayının eksi işareti çıkarılmış sayı değeridir. Mutlak değer sayı etrafındaki dik çizgiler ile gösterilir, örneğin $|-3| = 3$.) Mutlak değer kullanma önerisi mantıklıdır çünkü ortalamanın altında veya üstünde olmasına bakmaksızın puanların ortalamadan ne kadar saptıklarını bilmek istiyoruz. Burada önerilen ölçü oldukça meşrudur ve adı da **ortalama mutlak sapma**dır. Bütün basitliğine ve sezgisel cazibesine karşın, ortalama mutlak sapma istatistiksel metotlar arasında önemli bir rol oynamadı. Bunun yerine daha kullanışlı olan varyans ve standart sapma gibi ölçüler kullanılır.

## Varyans

standart sapma

örneklem varyansı ($s^2$)

evren varyansı

Şimdiye kadar ele aldığımız ölçüler psikologlardan daha çok istatistikçilerin ilgisini çekmektedir. Bu konuları da bilmeniz önemlidir, ancak varyans ve onu izleyen **standart sapma** dikkatinizi daha çok çekmelidir. Bu bölümde ele alacağımız **örneklem varyansı ($s^2$)**, sapmaların ortalamasının sıfır olması problemine başka bir açıdan çözüm getirir. (Örneklem varyansı yerine **evren varyansı**ndan bahsederken $\sigma^2$ (sigma kare) sembolü kullanılır.) Varyansta negatif sayıların karesinin pozitif olmasının avantajından faydalanılır. Yani, mutlak sapmaların yerine sapmaların karelerinin toplamı alınır. Ortalama almak istediğimiz için daha sonra toplamı puanların sayısı olan $N$'in bir fonksiyonuna böleriz. Her ne kadar $N$'e bölmeyi bekleseniz bile, $(N - 1)$'e böleriz. Örneklem varyansında bölen olarak $(N - 1)$'i kullanırız, çünkü birazdan göreceğiniz gibi bu sayede evren varyansının daha iyi bir kestirimini örneklem varyansı olarak elde ederiz. (Evren varyansı, her bir değer için sapmaların karelerinin toplamının $N$'e bölünmesiyle hesaplanır. Ancak evren varyansı nadiren hesaplanır, nerdeyse her zaman örneklem varyansından kestirilir.)

Eğer $s^2$'nin temsil ettiği değişkeni tam (kesin) olarak nitelemek önemli ise, temsil edilen değişken için alt indis kullanılabilir. Yani Set 4'teki veriyi $X$ ile gösterirsek, varyansını $s_X^2$ olarak yazabiliriz. $s_{\text{Set4}}^2$ olarak yazmayı ise uzun alt indisler hoş olmadığı için tercih etmeyiz. Genellikle değişkenleri $X$ veya $Y$ gibi basit harflerle gösteririz.

Örneğimizdeki Set 4 ve Set 32'nin örneklem varyanslarını aşağıdaki gibi hesaplayabiliriz:[9]

Set 4 ($X$)

$$s_x^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$
$$= \frac{(1,20 - 2,64)^2 + (1,82 - 2,64)^2 + ... + (4,02 - 2,64)^2}{20 - 1}$$
$$= \frac{8,1569}{19} = 0,4293$$

Set 32 ($Y$)

$$s_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N - 1}$$
$$= \frac{(3,13 - 3,26)^2 + (3,17 - 3,26)^2 + ... + (3,38 - 3,26)^2}{20 - 1}$$
$$= \frac{0,0903}{19} = 0,0048$$

Yapılan bu hesaplamalardan varyanslardaki farkların dağılımlardaki farklılıkları yansıttığını görebiliriz.

Her ne kadar varyans son derece önemli bir kavram ve en yaygın kullanılan istatistiklerden biri olsa da, doğrudan sezgisel bir yorumu yoktur. Çünkü elde edilen sonuçlar, sapmaların karelerinin alınmasıyla hesaplanır ve kare birimindedir. Set 4'ün ortalaması 2,64 iken varyans kare biriminde 0,4293'tür. Ancak kare birimini veriyle karşılaştırdığımızda çok az bir anlam ifade eder. Neyse ki bu problemin çözümü basittir: Varyansın kare kökü alınır.

## Standart Sapma

Standart sapma ($S$ veya $\sigma$) varyansın pozitif karekökü olarak tanımlanır ve örneklem için $S$ sembolü ile (gerektiğinde değişkeni belirten alt indis kullanılır) veya nadiren $SS$ ile gösterilir ($\sigma$ sembolü evren standart sapması için kullanılır).[10] Örneklemin standart sapması aşağıdaki formülle tanımlanır:

$$s_X^2 = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}$$

Örneğimiz için,

$$s_X = \sqrt{s_X^2} = \sqrt{0,4293} = 0,6552$$
$$s_Y = \sqrt{s_Y^2} = \sqrt{0,0048} = 0,0689$$

Kolaylık sağlaması için bu sayıları sırasıyla 0,66 ve 0,07'ye yuvarlayacağım.

---

9 Kitap içinde buradaki ve diğer hesaplamalarda, sizin hesaplamalarınız benimkilerden biraz farklı çıkabilir. Bunlar büyük ihtimalle yuvarlamalardan kaynaklanır. Bu yüzden benzer sonuçlar elde etmeniz yeterlidir.

10 Amerika Psikoloji Birliği "SD"yi tercih eder ama geri kalan herkes "$s$"yi tercih etmektedir.

Eğer standart sapmanın formülüne bakarsanız, mutlak sapmaların ortalamasında olduğu gibi, bu ölçünün de temel olarak her bir puanın ortalamadan farklarının ortalaması olduğunu görürsünüz. Diyelim ki bu sapmaların kareleri alındı, toplandı ve bölündü, ancak özünde bunlar hâlâ sapmalardır. Her ne kadar $N$ yerine $(N-1)$'e bölünmüş olsalar bile hâlâ bu sapmaların ortalamasına çok benzer bir ölçüye sahibiz. Bu nedenle çok fazla çarpıtmadan Set 4 için çekicilik ölçümlerinin ortalamadan ortalama olarak 0,66 birim saptığını ve Set 32 için ortalamadan sapmaların ortalama olarak sadece 0,07 birim olduğunu söyleyebiliriz. Standart sapmaya bu şekilde sapmaların ortalaması olarak düşünmek, kavrama ciddi haksızlıklar etmeden ona anlam vermemizi sağlar.

Bu sonuçlar çekicilik hakkında bize iki ilginç şey söyler. İlki, eğer bu deneyde katılımcı olsaydınız, birçok yüzün ortalaması alınarak oluşturulan Set 32'deki bileşkelerin birbirine çok benzer olduğu gerçeği, sizin Set 32'ye verdiğiniz puanların çok fazla değişkenlik göstermemesi olarak kendini gösterecekti. İkincisi ise, bu puanların Set 4'ünkinden daha yüksek bir ortalamaya sahip olmasının, daha fazla yüzün ortalamasının alınmasının daha çekici yüzler ortaya çıkarır gibi göründüğünü göstermektedir. Bu sizin günlük deneyimlerinizi doğruluyor mu? Mesela ben farklı özellikleri olan yüzlerin daha çekici olarak puanlanacaklarını düşünürdüm, ama demek ki yanılmışım. Geriye bakıp çekici olarak sınıflandırdığımız yüzleri şöyle bir düşünelim. Bunlar gerçekten farklı özellikleri olan yüzler mi? Eğer öyle ise, bu sonuçları açıklamak için ek hipoteziniz var mı?

Standart sapmaya, aynı zamanda puanların ne kadarının ortalamanın bir standart sapma altı veya üstü aralığında kaldığını betimlemek için de bakabiliriz. Simetrik ve tepe şeklindeki birçok dağılımda, gözlemlerin yaklaşık olarak üçte ikisinin ortalamanın bir standart sapma (altında veya üstünde) aralığında bulunduğunu söyleyebiliriz. (3. Bölüm'de anlatılacak olan normal dağılım için bu oran tam olarak üçte ikidir.) Her ne kadar istisnalar olsa da, özellikle çok çarpık dağılımlar hariç, bu kural oldukça kullanışlıdır. Eğer size ilkokul öğretmenlerinin beklenen maaşlarının ortalamasının 39.259 ve standart sapmasının 4000 dolar olduğunu söylersem, sizin öğretmenlerin üçte ikisinin maaşlarının 35.000 ile 43.000 dolar arasında olduğu çıkarımınız pek de yanlış olmayacaktır. Ayrıca çoğu (%95) ortalamanın iki standart sapma aralığında olacaktır. (Eğer maaşlarla gerçekten ilgileniyorsanız, medyan ortalamaya göre daha iyi bir ölçü olacaktır.)

## Varyans ve Standart Sapma İçin Hesaplama Formülleri

Her ne kadar daha önce verilen varyans ve standart sapma formülleri tam olarak doğru olsalar bile mâkul sayıda veri için bile hesaplamaları oldukça meşakkatlidir. Ayrıca kesirli sayıların karelerinin alınmasını içerdikleri için yuvarlama hatalarına açıktır. Tanım formülleri olarak mükemmeldirler, ama hesaplamalar için biz daha pratik formülleri ele alacağız. Bu yeni formüller cebirsel olarak daha öncekiler ile eşdeğerdirler, yani aynı sonuçları daha az çabayla elde etmemizi sağlarlar.

Örneklem varyansının tanımsal formülü aşağıdaki gibidir:

$$s_X^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

Daha pratik hesaplama formülü ise

$$s_X^2 = \frac{\sum X^2 - \frac{\left(\sum X^2\right)}{N}}{N - 1}$$

Benzer biçimde standart sapma:

$$s_X = \sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}}$$

$$= \sqrt{\frac{\sum X^2 - \frac{\left(\sum X^2\right)}{N}}{N - 1}}$$

Son zamanlarda görüşlerine saygı duyduğum bazı insanlar, artık bu hesaplamaların nadiren elle yapıldıkları için bu kitaptan çıkarılmasını önerdiler. Her ne kadar bu doğru olsa ve ben kendi derslerimin çoğunda bir çok formüle yer vermesem de çoğu insan hâlâ daha bu hesaplamaları yapabilmenin önemli olduğuna inanıyor. Daha da önemlisi bu formülleri farklı biçimlerde ileriki bölümlerde görecek olmamızın, bu formülleri tanımanın yeni konuları anlamada yardımcı olacak olmasıdır. Ancak bu eleştirilere daha karışık formüller için katılıyorum ve bu yüzden kitabın son birkaç basımında daha tanımsal formüllere ağırlık verdim.

Set 4 için örneklem varyansı, hesaplama formülü yardımıyla aşağıdaki gibi hesaplanır:

$$s_X^2 = \frac{\sum X^2 - \frac{\left(\sum X^2\right)}{N}}{N - 1}$$

$$= \frac{1,20^2 + 1,82^2 + ... + 4,02^2 - \frac{52,89^2}{20}}{19}$$

$$= \frac{148,0241 - \frac{52,89^2}{20}}{19} = 0,4293$$

Dikkat edilirse burada elde edilen değer, tanımdaki formülü kullanarak elde edilenle tamamen aynıdır. Ayrıca önceden de belirttiğim gibi, dikkat edilirse $\sum X^2 = 148,0241$ değeri $\left(\sum X\right)^2 = 52,89^2 = 279,35$ değerinden oldukça farklıdır. Set 32'nin varyans hesabı okuyucuya bırakılmıştır.

Biraz önce yapılan hesaplamalardaki matematik düzeyi ile kitabın diğer herhangi bir bölümünde gerekecek matematik düzeyinin hemen hemen aynı olacağını bilerek rahatlayabilirsiniz, çünkü çoğu uygulamalı istatistiği anlamak için ileri düzeyde matematik gerekmez.

## Uç Değerlerin Varyans ve Standart Sapmaya Etkisi

Varyans ve standart sapma uç değerlere oldukça duyarlıdır. Farklı bir biçimde ifade edecek olursak, uç değerler varyansın belirlenmesinde orantısız bir rol oynarlar. Ortalaması 5 ve kabaca 0 ile 10 aralığında değişen bir veri kümesini ele alalım. Varyansın tanımından, 5 puanının (ortalamanın) varyansa hiçbir katkısı olmadığı görülebilir, çünkü sapma puanı 0'dır. $(X - \overline{X})^2 = (6 - 5)^2 = 1$ olduğundan 6 puanının $s^2$'ye katkısı *1/(N – 1)*'dir. Fakat 10

puanının $s^2$'ye katkısı $25/(N-1)$ birimdir. Bu sebeple her ne kadar 6 ve 10 ortalamadan 1 ve 5 puan sapsalar da, onların varyansa etkisi 1 ve 25'tir. İşte bu, büyük sapmaların orantısız olarak temsil edilmesi diye söz edilen konudur. Uç noktalarda, güvenilir olmayan ölçme aracı kullanıldığında bunun akılda tutulması gerekir. Bu uç değerler, verimizin yorumlanmasında çok büyük etkiye sahip olabilirler. Bu bizim çarpık dağılmış veri kümelerinden hoşlanmamamızın başlıca nedenidir.

## Değişim Katsayısı

İstatistikte en yaygın olarak yaptığımız şeylerden biri iki veya daha çok grubun veya değişkenin ortalamalarının karşılaştırılmasıdır. Bu grupların veya değişkenlerin değişimlerini karşılaştırmak da kayda değer ve yerinde etkinliklerdendir. Örneğin, uzun-süreli belleği ölçmek için birbiriyle yarışan iki testimiz olsun.

Bu testlerden birisi ortalaması genellikle 15 ve standart sapması 3,5; ikincisi de oldukça farklı olarak ortalaması 75 ve standart sapması 10,5 olan veri kümesi üretsin. Diğer bütün faktörler aynı olduğunda, bu testlerden hangisi uzun süreli belleği daha iyi ölçer? İkinci testin standart sapması daha büyük olduğundan insanlar arasındaki farklılıkları gösteren yeterli değişimi ölçebilmek açısından daha iyi olduğunu savunmak gibi bir argümanı ortaya koymaya meyilli olabiliriz. Ancak her iki testin ortalamalarının birbirinden oldukça farklı olduklarını aklımızda tutmalı ve bu durumu göz önünde bulundurmalıyız.

Eğer standart sapmanın ortalamadan farklar olduğunu bir anlığına göz önünde bulundurursak herhangi bir puanın büyük ortalamadan sapmasının, küçüğüne göre daha büyük olduğunu düşünmek mantıklı gibi görünür. Örneğin öğretimin etkililiğini, ortalaması 3 olan 7'lik bir ölçekte ölçersek 4'ten büyük bir sapma elde etmemiz mümkün olmaz. Öte yandan ortalaması 30 olan 70'lik bir ölçekte 10 veya 20 gibi sapmalar çoğunlukta olacaktır. Bu iki ölçümdeki değişimi karşılaştırırken ikinci durumdaki büyük sapmalar için belli bir ayar yapmamız gerekir. Başka bir deyişle standart sapmalara bakarken ortalamaların büyüklüklerini de göz önünde bulundurmamız gerekir.

**değişim katsayısı (DK)**

Ortalamaları oldukça farklı olan standart sapmaları karşılaştırmanın en basit yolu, standart sapmaları ortalamalarıyla ölçeklendirmektir. İşte bu bizim **değişim katsayısı (DK)** ile yaptığımız işlemdir.[11] Bu katsayı basit bir biçimde standart sapmanın ortalamaya bölümü olarak tanımlanır:

$$\text{DK} = \frac{\text{Standart sapma}}{\text{Ortalama}} = \frac{s_X}{\bar{X}} \times 100$$

(100 ile çarpılmasının nedeni sonucun yüzde olarak ifade edilmesidir.) Bellek örneğimize dönecek olursak, ilk test için DK = (3,5/15) x 100 = 23,3'tür. Burada standart sapma yaklaşık olarak ortalamanın %23,3'üdür. İkinci test için DK = (10,5/75)x 100 = 14. Bu durumda değişim katsayısı yaklaşık olarak birincinin yarısıdır. Eğer birinci durumdaki büyük değişim katsayısının itinasız ölçüm sonucu olmadığına inanırsak, birinci testi ikincisine tercih ederiz.

İkinci örnek olarak Katz, Lautenschlager, Blackburn ve Harris (1990) çalışmasında SAT'da öğrencilerden çoktan seçmeli maddelerden oluşan bir grup soruyu cevaplamalarını istediler.[12] Bir grup ilgili metni okuyup soruları cevapladı. Diğer grup sorularla ilgili metni okumadan sorulara cevap verdiler - Moğol tarihi dersini almadan o dersin çoktan seçmeli sorularını cevaplar gibi.

---

[11] Andrew Giplin'e (kişisel iletişim, 1990) değişim katsayısının yararını hatırlattığı için teşekkür ederim. Bu anlamlı istatistik çoğunlukla gözden kaçmaktadır.

[12] Basit haliyle SAT olarak bilinen test, son zamanlarda SAT-1 olmuştur.

Veri kümesi aşağıdaki gibidir:

| | Metni okudu | Metni okumadı |
|---|---|---|
| Ortalama | 69,6 | 46,6 |
| $s_X$ | 10,6 | 6,8 |
| DK | 15,2 | 14,6 |

İki standart sapmanın oranı 10,6/6,8 = 1,56'dır, yani metni okuyan grubun standart sapması okumayan grubunkinden %50 daha büyüktür. Diğer yandan, değişim katsayısı her iki grup için hemen hemen aynı, yani gruplar arasındaki değişim farkları birinci gruptaki yüksek puanlarla açıklanabilir. (Bu arada şans başarısı ortalaması 20 ve standart sapması 4 olan puanlar üretir. Metni okumasalar bile öğrencilerin puanı sadece mantıklı seçimler yaparak şans başarısının üzerindedir.)

Değişim katsayısını kullanırken ölçülen değişkenin doğasını göz önünde bulundurmak önemlidir. Eğer ölçek keyfî bir ölçek ise bu katsayıya fazla güvenmemek gerekir. Fakat varyansa da fazla güvenmemek iyi olur. Bu biraz içgüdünün özellikle faydalı olduğu bir yerdir.

## Yansız Kestiriciler

Bölüm 1'de de belirttildiği gibi, biz ortalama ve varyans gibi ölçüleri genellikle evrendeki değerlerine karşılık gelen kestirimler olarak kullanmak için hesaplarız. Örneklemlerin özelliklerine istatistik deriz ve Roman harfleriyle gösteririz (örneğin, $\overline{X}$). Evrenin özelliklerine parametre denir ve Grek harfleri ile gösterilirler. Evren ortalaması $\mu$ sembolü ile gösterilir. Genellikle istatistikler parametrelerin kestirimi olarak kullanılırlar.

Eğer istatistiği hesaplamanın amacı onu parametre kestirimi olarak kullanmaksa, o zaman istatistik seçimimiz, kısmen o istatistiğin, parametrenin iyi bir kestiricisi olup olmadığına dayanır. Aslında ortalamanın diğer merkeze eğilme ölçülerine göre genellikle daha fazla tercih edilmesinin nedeni $\mu$'nun iyi bir kestiricisi olmasındandır. Örneklem varyansı ($s^2$) paydasındaki ($N - 1$) ile tanımlanır, bunun özel nedeni de $s^2$'nin evren varyansı ($\sigma^2$) kestiricisi olarak kullanılmasının sağladığı faydadır.

**yansız kestiriciler**

İstatistikçiler kestirimlerin birkaç farklı özelliğini (yeterlik, verimlilik, kararlılık ve yansızlık) tanımlarlar. Bunların ilk üçü bu kitapta bizim için fazla önemli değildir, bu yüzden onlar ele alınmayacaktır. Fakat yanlı ve **yansız kestiriciler**in arasındaki fark önemlidir. Çünkü genellikle kestiricinin yansız olmasını arzu ederiz.

Varsayalım bir şekilde ortalamasını ($\mu$) bildiğimiz bir evrenimiz var; örneğin, NBA'deki bütün basketbol oyuncularının boylarının uzunluklarının ortalaması. Eğer bu evrenden bir örneklem seçip örneklem ortalamasını hesaplayacak olursak, $\mu$'nun kestiricisi olması nedeniyle özellikle örneklem büyüklüğü $N$ büyük olduğunda $\mu$'ya yakın olmasını bekleriz. Bu evren için boy uzunluklarının ortalaması 7,0 inç ise ($\mu = 7{,}0'$), seçilen 10 oyuncudan oluşan örneklemin ortalamasının da tam olarak olmasa bile yaklaşık olarak $7{,}0'$olmasını bekleriz. (Bunu $\mu \approx 7{,}0'$ olarak yazarız, $\approx$ sembolü "yaklaşık olarak eşit" anlamına gelir.) Şimdi başka bir örneklem seçtiğimizi ve ortalamasının $\overline{X}_2$ olduğunu düşünelim. (İndis ardı ardına seçilen örneklemleri temsil etmek için kullanılır. Yani, 43'üncü örneklemin ortalaması ($\overline{X}_{43}$) ile temsil edilir.) Bu $\overline{X}_{43}$ ortalamasının da muhtemelen $\mu$'ya oldukça yakın

olmasını bekleriz, fakat bu değerin ne $\mu$'ya ne de $\bar{X}_1$'e tam olarak eşit olmasını beklemeyiz. Eğer bu işlemi sonsuz kez tekrar edersek, örneklem ortalamalarının ortalamasının tam olarak $\mu$'ya eşit olduğunu buluruz. Bu yüzden örneklem ortalamalarının **beklenen değeri** (örn. çok sayıda örneklemin ortalaması) $\mu$'ya eşittir. Beklenen değeri kestirilecek parametreye eşit olan kestiriciye yansız kestirici denir ve bu bir istatistiğin sahip olmasını istediğimiz önemli bir özelliğidir. Hem örneklem ortalaması hem de örneklem varyansı onlara karşılık gelen parametrelerin yansız kestiricileridir. (Örneklem varyansı formülünün paydasında $N - 1$ kullanmamızın sebebi yansız bir kestirici elde etmek istememizdir.) Genel olarak yansız kestiriciler yansız insanlar gibidirler - yanlı olanlara göre onlarla çalışılması daha hoştur.

**beklenen değer**

## Evren Varyansının Kestiricisi Olarak Örneklem Varyansı

Örneklem varyansı yansızlık konusunda söz edilenler için mükemmel bir örnek teşkil eder. Daha önce varyans ve standart sapmanın hesabında bölen olarak $N$ yerine $N - 1$ kullanıldığından söz setmiştik. Şimdi bunun nedenini açıklamanın tam zamanı. (Siz $N - 1$'e bölmeye razı olabilirsiniz ama ben bu konuda birçok soru alıyorum, siz açıklamayı istersenız okur, istemezseniz bir sonraki bölüme geçebilirsiniz.)

Örneklem varyansının paydasında neden $N - 1$ olduğunu açıklamak için birçok yol vardır. Belki de en basiti örneklem varyansının ($s^2$) evren varyansının ($\sigma^2$) yansız bir kestiricisi olması hakkında söylenenlerdir. Bir dakikalığına evren varyansını bildiğimiz bir evrenden sonsuz sayıda örneklemimiz (her biri $N$ gözlem içeren) olduğunu farz edelim. Ayrıca örneklem varyansını hesaplamak için de aşağıdaki formülü kullandığımızı düşünelim.

$$\frac{\sum (X - \bar{X})^2}{N}$$

(Paydaya dikkat edelim.) Eğer bu örneklem varyanslarının ortalamasını alırsak,

$$\text{Ortalama } \frac{\sum (X - \bar{X})^2}{N} = \text{E}\left[\frac{\sum (X - \bar{X})^2}{N}\right] = \frac{(N-1)\sigma^2}{N}$$

buluruz. Burada E[ ], "parantez içindekinin beklenen değeri"ni belirtir, Böylece $\sum (X - \bar{X})^2/N$'in ortalama değeri $\sigma^2$ değildir. Bu yanlı bir kestiricidir.

Burada "$N - 1$'e bölüyoruz, çünkü bu işe yarıyor" der gibiyiz. Fakat neden işe yarıyor? Bunu açıklamak için öncelikle **serbestlik derecesi**ni göz önüne almalıyız. Varsayalım elimizde 6, 8 ve 10 sayıları var. Bunların ortalaması 8. Şimdi bize ortalamayı 8'de sabit tutma şartıyla bu sayıları değiştirebileceğimizi söyleseler, kaç tane sayıyı serbestçe değiştirebiliriz? Eğer bu sayıların hepsini birden değiştirirsek, ortalamanın 8 olmayacağına hemen hemen şüphe yoktur. Eğer ortalama sabit kalacaksa sayıların sadece iki tanesini serbestçe değiştirebiliriz. Mesela eğer 6'yı 7 ile ve 10'u 13 ile değiştirirsek, geri kalan sayı kendiliğinden saptanır; eğer ortalamayı 8 olarak tutacaksak bu sayının 4 olması gerekir. Eğer elimizde 50 sayımız ve koşullar da aynı olsaydı, yine sayıların sadece 49'unu serbestçe değiştirebiliriz ve 50'nci sayı saptanmış olur.

**Serbestlik derecesi (*sd*)**

Şimdi tekrar örneklem ve evren varyanslarına geri dönelim ve örneklem varyansını hesaplarken niye bir serbestlik derecesi kaybettiğimizi görelim:

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N} \qquad s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

$\sigma^2$ durumunda, $\mu$ biliniyor ve veriden kestirilmesine gerek yok. Bu yüzden hiç serbestlik derecesi kaybedilmez ve payda $N$'dir. Ancak $s^2$'de $\mu$ bilinmiyor ve örneklem ortalamasından kestirilmesi gereklidir. $\mu$'yu $\bar{X}$'dan kestirdiğimizde, onu değişkenliği kestirme amacı için sabitlemiş oluruz. Bu nedenle, sözü edildiği gibi, "1" serbestlik derecesi kaybederiz ve sadece $N - 1$ serbestlik derecemiz kalır. ($N - 1$ puanı serbestçe değiştirebiliriz.) Ortalamayı kestirdiğimiz zamanlarda bir serbestlik derecesi kaybederiz. Bu yüzden payda (kestirimimizin yapıldığı puanların sayısı) bu sınırlamayı yansıtmalıdır. Payda bağımsız değişebilen veri sayısını gösterir.

## 2.9 Kutu Grafikleri: Dağılımın ve Uç Değerlerin Grafik Gösterimleri

Daha önceden gövde-ve-yaprak gösteriminin, verinin birkaç özelliğini aynı anda betimlediğini görmüştük. Bu tür gösterimler, hem her bir verinin değerini ayrı ayrı *görmemizi sağlar hem de veri kümesini* histogram benzeri bir biçimde bir araya toplar. Gövde-ve-yaprak gösterimine ek olarak John Tukey, veri kümesinin dağılımını ön plana çıkaran veriyi görselleştirme yöntemi de geliştirdi. Bu yöntem **kutu grafiği** veya **kutu**-ve-**bıyık grafiği** olarak adlandırılır.

**kutu grafiği**

**kutu-ve-bıyık grafiği**

Tablo 2.7'de verilen gövde-ve-yaprak grafiği ve eşlik eden veri kümesi, Vermont Üniversitesi'nde normal ve düşük ağırlıklı yeni doğanların katılımıyla yapılan yeni doğan gelişimi çalışmasındaki 38 normal ağırlığa sahip bebeğin hastanede kalma sürelerini göstermektedir. Üç yeni doğana ait veri kayıptır ve bunlar (*) işaretiyle gösterilmiştir. (Yıldız işareti kayıp veriyi göz ardı etmememiz gerektiğini vurgulamak için konulmuştur.) Veri kümesindeki sayılar 1 ile 10 arasında değiştiği için, iki durum hariç tüm yapraklar sıfırdır. Sıfırlar sadece histograma benzer bir dağılım oluşturmak için yer işgal ederler. Gövde ve yaprak gösterimi incelendiğinde dağılımın ortancasının 3 gün olduğu ve sağa çarpık olduğu görülür. Gövdenin alt kısmına doğru HI yazısı ve 20 ile 33 değerleri görülebilir. Bunlar uç değerlerdir ve varlıklarını vurgulamak için bu biçimde gösterilirler. Değerlerinin yeterince büyük olup olmadığı sorusu kutu grafiklerinde ele alınacaktır. Gövde-ve-yaprak gösteriminin son satırı kayıp veri sayısını belirtir.

Orijinalinde Tukey kutu grafiklerini kendi geliştirdiği ölçüler ile tanımladı. Çoğu kişi ise bugünlerde kutu grafiklerini daha geleneksel ölçüleri kullanarak çiziyor. Kitabın bu basımında geleneksel ölçüleri kullanan yaklaşıma başvurulmuştur.

N tane puanın bulunduğu veri kümesinde ortancanın yerini ($N$+1)/2'inci sıradaki puan olarak tanımladık. $N$ tek sayı ise, ortancanın yeri tam sayı olacak ve ortanca ise sıralanmış veri kümesinde bu yeri işgal eden sayı olacaktır. Ortancanın yeri kesirli sayı olduğunda ($N$'nin çift sayı olması durumunda) ortanca bu yerin her iki tarafındaki iki değerin ortalaması olacaktır. Tablo 2.7'deki veri kümesinde ortancanın yeri (38 + 1)/2 = 19,5'uncu sıra olup değeri 3'tür. Kutu grafiğini oluşturmak için daha önceden tanımlanan birinci ve üçüncü çeyrekleri belirlemeliyiz. Bunu yapmanın en kolay yolu **çeyreklerin yerini (konumunu)** belirlemektir. Bu da aşağıdaki gibi tanımlanır:

**çeyreklerin yeri**

$$\text{Çeyreğin yeri} = \frac{\text{Ortancanın yeri} + 1}{2}$$

Eğer ortancanın yeri kesirli bir sayı ise, çeyreğin yeri bulunurken payda bulunan kesir atılır ve yeri ortancanın yeri ile aynıdır. Çeyreğin yeri bize sıralanmış veri kümesinde

**Tablo 2.7** Yeni doğan bebeklerin hastanede kalma süresi (gün olarak) verileri ve gövde-ve-yaprak gösterimi

| Veri | Gövde ve Yaprak | |
|---|---|---|
| 2 1 7 | 1 | 000 |
| 1 33 2 | 2 | 000000000 |
| 2 3 4 | 3 | 00000000000 |
| 3 * 4 | 4 | 0000000 |
| 3 3 10 | 5 | 00 |
| 9 2 5 | 6 | 0 |
| 4 3 3 | 7 | 0 |
| 20 6 2 | 8 | |
| 4 5 2 | 9 | 0 |
| 1 * * | 10 | 0 |
| 3 3 4 | HI | 20,33 |
| 2 3 4 | Kayıp = 3 | |
| 3 2 3 | | |
| 2 4 | | |

© Cengage Learning 2013

çeyreklik değerlerinin[13] nerede olduğunu söyler. Hastanede kalma süresi veri kümesi için çeyreğin yeri (19 + 1)/2 = 10'dur. Bu yüzden çeyrekler en alttan ve en üstten onuncu puanlar olacaktır. Bu değerler sırasıyla 2 ve 4'tür. Aynı değere sahip puanları olmayan veri kümelerinde veya büyük örneklemlerde çeyrekler ortadaki puanların %50'sini içine alır.

**çeyrekler arası açıklık**

**iç çit sınır**

**komşu değerler**

Kutu grafiklerini anlamak için gerekli kavramları tamamlamak amacıyla üç veya daha fazla terimi daha açıklamamız gerekir: **çeyrekler arası açıklık**, **iç çit sınır** ve **komşu değerler**. Önceden gördüğümüz gibi, çeyrekler arası açıklık üçüncü çeyrek ile birinci çeyrek arasındaki açıklıktır. Bizim verimiz için çeyrekler arası açıklık 4 – 2 = 2'dir. İç çit (sınır) Tukey tarafından iç çeyrek açıklığının 1,5 katının uygun çeyreğin altında veya üstünde kalan nokta olarak tanımlanmıştır. Bizim veri kümemiz için iç çeyrek açıklığı 2 ve iç çit çeyreklerden 2 x 1,5 = 3 nokta uzaktadır. Bizim çeyreklerimiz 2 ve 4 olduğu için, iç çitler (sınırlar) 2 – 3 = –1 ve 4 + 3 = 7. Komşu değerler iç çitten daha uzakta olmayan veri gerçek değerleridir. Sahip olduğumuz en küçük veri 1 olduğu için bu değer alt iç çite en yakın değerdir ve alt bitişik değerdir. Üst iç çit 7'dir ve veri kümemiz içinde 7 olduğundan bu değer üst bitişik değer olacaktır. Şimdiye kadar sözünü ettiğimiz bütün terimlerin nasıl hesaplandıkları Tablo 2.8'de gösterilmiştir.

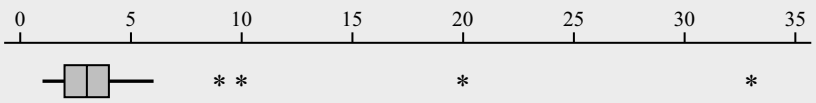
İç çit ve komşu değerler bazen zihin karışıklığına yol açabilir. Bir tarlaya dağılmış inek sürüsü hayal edelim. (Ben hayatımın çoğunu Vermont'da geçirdim, bu yüzden inekler doğal bir örnek.) Tarlanın etrafındaki çit kutu grafiğindeki iç çiti temsil eder. Çite en yakın fakat hâlâ çitin içinde olan inekler komşu değerlerdir. Çitin dışına kaçan ve yolda dolaşan inekler de uç değerler olacaklar.

Şimdi kutu grafiğini çizmek için hazırız. İlk olarak elde edilen bütün değerleri kapsayan ölçeği çizeriz ve numaralandırırız. Bu Tablo 2.8'in altında yapılmıştır. Sonra $Q_1$'den

13 Tukey bu durumdaki çeyreklikleri menteşeler (hinges) olarak adlandırdı. Fakat onları çeyrekler olarak düşünerek çok az şey kaybederiz.

**Tablo 2.8** Tablo 2.7'deki veri için hesaplamalar ve kutu grafiği

Medyanın konumu = $(N + 1)/2 = (38 + 1)/2 = 19{,}5$
Medyan = 3
Çeyreğin konumu = (medyanın konumu† + 1)/2 = (19 + 1)/2 = 10
$Q_1$ = 10. en düşük puan = 2
$Q_3$ = 10. en yüksek puan = 4
Çeyrekler arası açıklık = 4 – 2 = 2
Çeyrekler arası açıklık* = 1,5 = 2 x 1,5 = 3
Alt iç çit = $Q_1$ – 1,5 (çeyrekler arası açıklık) = 2 – 3 = –1
Üst iç çit = $Q_3$ + 1,5 (çeyrekler arası açıklık) = 4 + 3 = 7
Alt komşu değer = en küçük değer ≥ alt çit = 1
Üst komşu değer = en büyük değer ≤ üst çit = 7




†Kesirli değerler düşürülür.

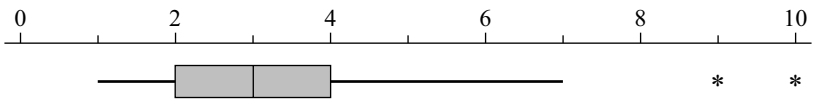
© Cengage Learning 2013

$Q_2$'ye dikdörtgen bir kutu çizeriz, dikine çizgi ortancanın yerini belirtir. Daha sonra çeyreklerden bitişik değerlere kadar çizgi çizeriz (bıyık). Son olarak komşu değerlerin dışında kalan uç değerlerin bulundukları noktaları işaretleriz.

Tablo 2.8'den birkaç önemli noktayı görebiliriz. İlki, dağılımın merkezî kısmı oldukça simetriktir. Bu, ortancanın, kutunun merkezinde olmasından ve ayrıca gövde-ve-yaprak gösteriminden görülebilir. Aynı zamanda sağdaki bıyık soldakine göre oldukça uzun olduğundan dağılımın sağa çarpık olduğunu da görebiliriz. Bu tam açık olarak görülmese de gövde-ve-yaprak grafiğinden de anlaşılabilir. Son olarak, dört tane uç değer görülmektedir; uç değer burada bıyıklardan daha dışarıda olan değerler olarak tanımlanır (ve bu yüzden komşu değerlerden daha uç değerler.) Gövde-ve-yaprak gösterimi uç değerlerin yerini kutu grafiği kadar görsel olarak göstermemektedir.

Burada uç değerler özel ilgiyi gerektirir. Bir uç değer, ölçümdeki, veri kaydındaki, veri girişindeki bir hatayı gösterebilir veya sadece uçta olan bir değeri gösterebilir. Örneğin, bizim verimiz hastanede kalma süresini belirtmektedir, fiziksel bir özürle doğan bebeğin uzun süre hastanede kalması gerekli olmuş olabilir. Bunlar gerçek veri olduklarından, hastane kayıtlarından bu dört uç değer vakasına gidip daha yakından bakmak mümkündür. Kayıtlar kontrol edildiğinde iki verinin yanlış girildiği görüldü ve bu değerler düzeltildi. Diğer iki uç değerin bebeğin fiziksel sorunlarından kaynaklandığı anlaşıldı. Burada proje yöneticisinin yeni doğanların çalışmaya dahil edilip edilmeyecekleri konusunda karar vermesi gerekti (her ikisi de çalışmada yer aldı). Düzeltilen iki değer 33 ve 20 yerine sırasıyla 3 ve 5'ti. Düzeltilen veri kümesi için kutu grafiği Şekil 2.14'te gösterilmiştir. Bu kutu grafiği ile Tablo 2.8'deki en büyük iki değer arasındaki boşluk hariç birbirinin aynısıdır.

Kutu grafikleri veri kümesinin dağılımını incelemek için oldukça yararlı araçlardır. Bu grafikler, yapılacak analizler öncesinde potansiyel sorunları saptamak ve veri hatalarını inceleme açısından yararlıdır. Kitabın geri kalan kısımlarında kutu grafikleri genellikle veri kümesi için görsel rehberler olarak sunulmaktadırlar.

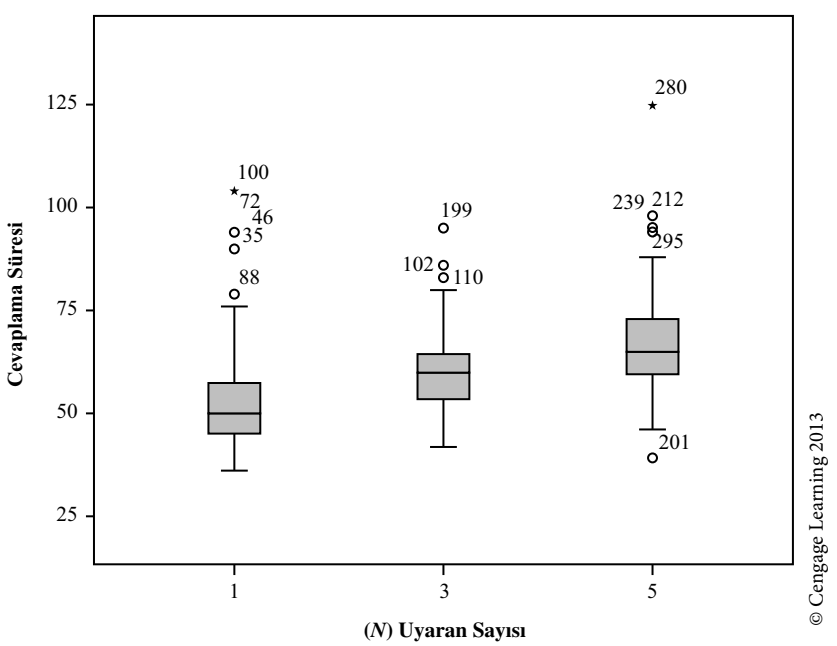



**Şekil 2.14** Tablo 2.7'deki verilerin düzeltilmiş hali için kutu grafiği

© Cengage Learning 2013

**Uyarı:** Farklı istatistik bilgisayar programları kutu grafiğinin çeşitli ögelerini farklı biçimde tanımlayabilir. (Konunun detaylı tartışması için Frigge, Hoaglin ve Iglewicz'e (1989) bakabilirsiniz.) Aynı veri kümesi için birbirinden biraz farklı kutu grafiği çizen iki farklı istatistik programı ile karşılaşabilirsiniz. Hatta bu programlar farklı puanları uç değer olarak belirleyebilirler. Ancak kutu grafikleri normalde formel olmayan incelemeye yönelik araçlar olarak kullanılırlar, bu yüzden de tanımlamalardaki ufak farklar nadiren sorun teşkil eder. Buradaki veri kümelerini kullanarak farklı bir istatistik programı ile kutu grafiği çizerseniz farklı bir grafik ile karşılaşabilirsiniz, bunun sebebini açıklamak için burada bu konudan bahsetmek istedim. (R programında kutu grafiği oluşturmak için kullanabileceğiniz basit kodu Web sitesinde Boxplots.R olarak bulabilirsiniz.)

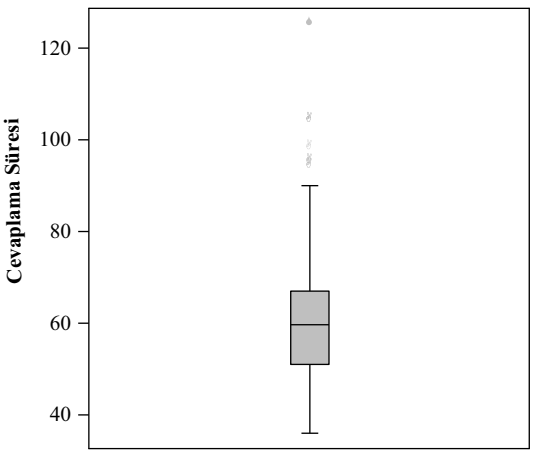
Kutu grafiklerinin gerçek faydası birden fazla grubu birbirleriyle karşılaştırmak istediğimizde ortaya çıkar. Bunun için bölümün başında kullandığımız bireyin belli bir sayının bir önceki slaytta gördüğü sayılar arasında yer alıp almamasına göre butona basma süresinin slaytta bulunan rakam sayısına göre incelendiği örnek veriyi kullanacağız. Şekil 2.15'teki SPSS ile üretilen kutu grafikleri, uyaranın slaytlarda bulunduğu durumlarda cevap süresini uyaran sayısına göre ayrı ayrı belirterek göstermektedir. Uç değerler, uyarıcının belirtildiği deneme numaraları ile betimlenmiştir. Bu grafikten çıkarılabilecek en açık sonuç, slaytlardaki uyaran sayısı arttıkça, cevaplama süresi ve yayılmanın arttığıdır. Ayrıca dağılımların büyük ölçüde simetrik olduklarını da görebiliriz (kutularda ortanca hemen hemen ortadadır ve birkaç uzun cevaplama süresini belirten uç değer vardır).




**Şekil 2.15** Orijinal uyarıcı kümesinde uyaran sayısının fonksiyonu olarak tepki sürelerinin kutu grafiği

**Rapor: Cevaplama Süresi**

| Uyaran sayısı | N | Ortalama | Ortanca | Std. Sapma | Varyans |
|---|---|---|---|---|---|
| 1 | 100 | 53,27 | 50,00 | 13,356 | 178,381 |
| 3 | 100 | 60,65 | 60,00 | 9,408 | 88,153 |
| 5 | 100 | 66,86 | 65,00 | 12,282 | 150,849 |
| **Toplam** | 300 | 60,26 | 59,50 | 13,011 | 169,277 |




SPSS, Inc. çıktısından adapte edilmiştir.

**Gösterim 2.1** Cevaplama süresi verisinin SPSS analizi

## 2.10 Değişme Ölçülerinin SPSS Yardımıyla Elde Edilmesi

Merkeze eğilme ve değişme ölçülerini cevaplama süreleri deneyinden elde edilen veri kümesi için Gösterim 2.1'de görüldüğü gibi SPPS yardımıyla da hesaplayabiliriz. Burada ***Analyze/Compare Means/Means*** menü komutları kullanılmıştır, çünkü her bir uyaran sayısı düzeyi için ayrı betimsel istatistikler elde edilmek istenmiştir. Dikkat edilirse üç grup için bu istatistikler mevcuttur. ***Graphs/Interactive/Boxplot*** komutu yukarıda görülen kutu grafiğini oluşturur. Şekil 2.15'te uyaran sayısına göre kutu grafikleri ayrı ayrı görüldüğü için burada sadece bütün hali sunulmuştur. Uç değerlerin ortaya çıkışları görülebilir.

## 2.11 Yüzde Birlikler (Percentiles), Çeyrekler (Quartiles) ve Onda Birlikler (Deciles)

Bir dağılımın merkezinin konumu ve değişiminden başka birçok özelliği vardır. Bunlardan biri olan ve dağılımı dört eş parçaya bölen çeyreklerden (quartiles) kutu grafikleri işlenirken kısaca söz edildi. Buna göre birinci çeyrek, sıralanmış veri kümesinin alt %25'ini, ikinci çeyrek %50'sini ve üçüncü çeyrek %75'ini ayırır. (Dikkat edilirse ikinci çeyrek aynı zamanda ortancadır.) Bu çeyrekler Şekil 2.10'daki gelişim grafiğinde açıkça gösterilmiştir. Eğer dağılımdaki geçişleri daha duyarlı bir şekilde incelemek istersek, dağılımı on eş parçaya bölen **onda birlikler**e bakabiliriz. Birinci onda birlik dağılımın alt %10'unu, ikincisi %20'sini ve benzeri biçimde devam eder. Son olarak hepimizin bildiği **yüzdelikler (yüzde birlikler)** dağılımı yüz eş parçaya böler. Buna göre 81'inci yüzdelik, sıralanmış veri kümesinde puanların %81'inin altında kaldığı noktadır.

**onda birlikler**

**yüzde birlikler**

**kantiller**

**fraktiller**

Çeyrekler, onda birlikler ve yüzde birlikler genellikle **kantiller (quantiles)** veya bazen **fraktiller (fractiles)** adıyla bilinen genel bir istatistik sınıflamasının en yaygın üç örneğidir. Bu kitapta kantiller hakkında daha fazla bahsetmeyeceğiz, fakat daha geniş bir biçimde giriş seviyesi kitaplarda yer bulurlar (örn. Howell, 2008a). Ayrıca Tukey tarafından savunulan birçok açımlayıcı istatistik tekniklerinde önemli rol oynarlar.

## 2.12 Doğrusal Dönüşümlerin Veri Üzerindeki Etkisi

**doğrusal dönüşüm**

Bazen elimizdeki veriyi bir biçimde dönüştürmek isteriz. Örneğin, metreyi santimetreye, santimetreyi inçe, Fahrenheit'ı Celsius'a, 79 üzerinden puanlanan soruları 100'lü ölçeğe dönüştürmek isteriz. Bütün bunlar, **doğrusal dönüşüm** olarak adlandırılırlar. Bu dönüşümlerde her bir $X$'i bir sabitle çarpar (belki 1) ve bir sabit ekleriz (belki 0): $X_{\text{yeni}} = bX_{\text{eski}} + a$, a ve b sabitlerdir. Üstel, logaritmik, trigonometrik ve benzeri fonksiyonlarla yapılan dönüşümler, doğrusal olmayan dönüşümler olarak sınıflandırılırlar. Doğrusal dönüşümlerin bir örneği Celsius'u Fahrenheit'a dönüştürmektir:

$$F = 9/5(C) + 32$$

Yapılan işlemleri doğrusal dönüşümlerle sınırlandırdığımız sürece, yeni ölçekteki gözlemlerin ortalama ve varyansını eski ortalama ve varyans cinsinden ifade etmeye yarayan basit kurallar vardır:

1. Bir veri kümesine bir sabiti eklemek (veya çıkarmak), o veri kümesinin ortalamasına da o sabiti ekler (veya çıkarır):

   $X_{yeni} = X_{eski} \pm a$ için $\overline{X}_{yeni} = \overline{X}_{eski} \pm a$

2. Bir veri kümesini bir sabit ile çarpmak (veya bölmek) veri kümesinin ortalamasını da aynı sabitle çarpar (veya böler):

   $X_{yeni} = bX_{eski}$ için $\overline{X}_{yeni} = b\overline{X}_{eski}$.

   $X_{yeni} = X_{eski}/b$ için $\overline{X}_{yeni} = \overline{X}_{eski}/b$.

3. Bir veri kümesine bir sabit eklemek veya çıkarmak, veri kümesinin varyansını ve standart sapmasını değiştirmez:

   $X_{\text{yeni}} = X_{\text{eski}} \pm a$ için $s^2_{\text{yeni}} = s^2_{\text{eski}}$

4. Bir veri kümesini bir sabit ile çarpmak (veya bölmek) veri kümesinin varyansını sabitin karesiyle ve standart sapmasını o sabit ile çarpar:

   $X_{yeni} = bX_{eski}$ için $s^2_{yeni} = b^2 s^2_{eski}$ **ve** $s_{yeni} = bs_{eski}$.

   $X_{yeni} = X_{eski}/b$ için $s^2_{yeni} = s^2_{eski}/b^2$ **ve** $s_{yeni} = s_{eski}/b$.

Aşağıdaki örnekte bu kurallar açıklanmaktadır. Her bir durumda sabit olarak 3 kullanılmıştır.

**Bir Sabitin Eklenmesi**

| Eski Veri | $\bar{X}$ | $s^2$ | $s$ | Yeni Veri | $\bar{X}$ | $s^2$ | $s$ |
|---|---|---|---|---|---|---|---|
| 4, 8, 12 | 8 | 16 | 4 | 7, 11, 15 | 11 | 16 | 4 |

**Bir Sabitle Çarpılması**

| Eski Veri | $\bar{X}$ | $s^2$ | $s$ | Yeni Veri | $\bar{X}$ | $s^2$ | $s$ |
|---|---|---|---|---|---|---|---|
| 4, 8, 12 | 8 | 16 | 4 | 12, 24, 36 | 24 | 144 | 12 |

## Merkezlendirme (Centering)

merkezlendirme

**Merkezlendirme** kelimesinin veri ile birlikte kullanılması gittikçe yaygın olarak görülmektedir. Temel fikir oldukça basittir. Bütün gözlemlerden örneklem ortalaması çıkarılır. Bu yeni ortalamanın 0,00 olacağı, fakat standart sapma ve varyansın değişmeyeceği anlamına gelir. Bunun birkaç örneğini göreceğiz, fakat bu araya bir tanesini sıkıştıralım. Varsayalım cinsel çekiciliği boydan yordayacak bir eşitlik yazdık. Bu eşitlikte kesişim olarak adlandırılan ve boy uzunluğu "0" olduğunda çekiciliği yordayan bir terim var. Fakat hiç kimsenin boyu "0" değildir, bu nedenle de anlamsız bir istatistiktir. Eğer boy uzunluğu verisinden boyların ortalamasını çıkararak boyları merkezlendirirsek, boy "0" olduğunda çekiciliği veren kesişim bu durumda ortalama boyda olan birinin çekiciliğini temsil eden anlamlı bir istatistiktir. Merkezlendirme birçok durumda kullanılır, ama bu örnek genel amacını hemen hemen açıklar.

## Bir Dönüşüm Biçimi Olarak Yansıma (Reflection)

Ölçeğin sıralanışını tersine çevirmek çok yaygın ve yararlı bir dönüşümdür. Örneğin, katılımcılardan birkaç maddeyi 5'li likert ölçeği üzerinden katılım derecelerini belirtmelerini istediğimizi varsayalım. Katılımcıların hiç düşünmeden sayfanın başından sonuna kadar aynı noktayı işaretlemelerini önlemek için, maddelerin yarısını olumlu diğer yarısını ise olumsuz cümlelerle ifade edelim. Böylece 5'li ölçekte 5 "tamamen katılıyorum" ve 1 "kesinlikle katılmıyorum"u temsil eder. "Sinemadan nefret ederim" maddesi için 4, "tiyatroyu severim" maddesindeki 2'ye karşılık gelir. Eğer puanların karşılaştırılabilir olmasını istiyorsak, olumsuz maddeleri tekrar puanlamalı (örneğin), 5'i 1'e ve 4'ü 2'ye çevirme gibi. Bu işleme **yansıma** denir ve basit bir doğrusal dönüşüm yoluyla yapılır. Bunun için sadece $X_{\text{yeni}} = 6 - X_{\text{eski}}$ yazarız. Sabit (6) ölçekteki en büyük değere 1 eklenerek bulunur. Şu açıkça görülmelidir ki ölçeği yansıttığımızda ortalamasını da yansıtırız, fakat bunun varyans ve standart sapma üzerinde etkisi yoktur. Bu önceki listedeki kural 3'tür.

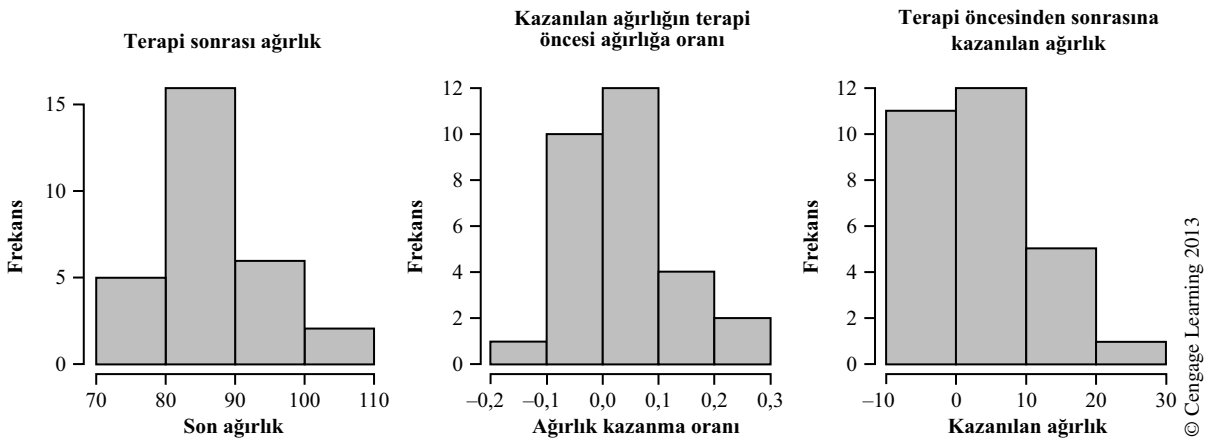
yansıma

## Standartlaştırma (Standardization)

Doğrusal dönüşümlerden yaygın olarak uygulanan biri, her bir gözlemden ortalamanın çıkarılmasıyla verinin yeniden ölçeklendirilmesidir. Bu şekilde dönüştürülen gözlemlere **sapma puanları** denir ve dönüşümün kendisi merkezlendirmeyi içerir, çünkü ortalamayı "0"a merkezlendiriyoruz. Merkezlendirme en çok daha sonraki bölümlerde ele alınacak olan regresyonda kullanılır. Daha yaygın olarak kullanılan bir dönüşüm ise sapma puanları oluşturup daha sonra sapma puanlarını standart sapmaya bölmeyi içerir. Bu puanlara **standart puanlar** denir ve bu işlem de **standartlaştırma** olarak adlandırılır. Temel olarak standart puanlar, standart sapma biriminde ölçülen dönüştürülmüş gözlemlerdir. Örneğin 0,75 standart puanı, ortalamanın 0,75 standart sapma üzerinde olan puandır: –0,43 standart puanı ortalamanın 0,43 standart sapma altında olan puandır. Bölüm 3'te normal

sapma puanları

standart puanlar
standartlaştırma




**Şekil 2.16** Terapinin iştahsız kızların ağırlığına etkisini gösteren alternatif ölçüler

dağılımdan söz ederken standart puanlardan daha çok bahsedeceğiz. Burada bu konudan özellikle söz etmenin sebebi, elimizde normal dağılım olsa da olmasa da standart puanları hesaplayabileceğimizi göstermektir. İnsanlar genellikle standart puanların normal dağıldığını düşünürler, fakat öyle olması için kesinlikle bir neden yoktur. Standartlaştırma ham puanların basit bir doğrusal dönüşümüdür ve bu, dağılımın şeklini değiştirmez.

## Doğrusal Olmayan Dönüşümler

**doğrusal olmayan dönüşümler**

Doğrusal dönüşümler genellikle veriyi 100'lü ölçeğe dönüştürmek veya standart forma sokmak gibi daha anlamlı bir formata dönüştürmek için kullanılırken, **doğrusal olmayan dönüşümler**e genellikle dağılımın şeklini değiştirmek için başvurulur. Görüldüğü gibi, doğrusal dönüşümler dağılımın şeklini değiştirmezler, öte yandan doğrusal olmayan dönüşümler çarpık dağılımları daha simetrik veya tersini yapabilir ve uç değerlerin etkisini azaltabilir.

Bazı doğrusal olmayan dönüşümler o kadar yaygındır ki normalde onları dönüşüm olarak düşünmeyiz. Everitt (Hand, 1994) iştahsızlık bilişsel davranış terapisi gören 29 genç kızın tedavi öncesi ve sonrası ağırlıklarını rapor etmiştir. Mantıklı bir ölçü terapiden sonra bireyin ağırlığı (*Y*) olurdu. Diğeri, tedavi öncesinden sonrasına kazandığı ağırlık *(Y-X)*. Üçüncü alternatif, kazanılan ağırlığı esas ağırlığın fonksiyonu olarak rapor etmektir. Bu da *(Y-X)/X*'tir. Bu ölçüyü kullanabiliriz, çünkü bireyin puanının ne kadar arttığı, başladığında kilosunun ne kadar düşük olduğuyla ilgilidir. Şekil 2.16 aynı veriye dayanan bu üç ölçünün histogramlarını göstermektedir.

Şekil 2.16'da üç alternatif ölçüyü görebilirsiniz, ikinci ve üçüncüde *X* ve *Y*'nin doğrusal olmayan dönüşümlerinin oldukça farklı dağılımlara sahip oldukları görülür. Kazanılan ağırlığın tedavi öncesi ağırlığa bölündüğü durumun grafiğinin, diğerlerinden daha normal dağıldığı açıktır. Kitabın ileriki bölümlerinde diğer doğrusal olmayan dönüşümlerin (örn. karekök ve logaritmik dönüşümler) dağılımın şeklini daha simetrik yapmak için nasıl kullanıldığını göreceksiniz. Örneğin, birkaç yerde stresi psikolojik semptomların yordayıcısı olarak kullanacağız. Fakat semptomlar düzgün bir biçimde dağılmazlar, bu yüzden semptomların $\log_e$'sini alıp değişken olarak LnSemptomu kullanacağız.

## Anahtar Terimler

Frekans dağılımları (2.1)
Histogram (2.2)
Gerçek alt limit (2.2)
Gerçek üst limit (2.2)
Orta nokta (2.2)
Uç değerler (2.2)
Çekirdek yoğunluk grafiği (2.3)
Gövde-ve-yaprak gösterimi (2.4)
Açımlayıcı veri analizi (2.4)
Önemli basamaklar (2.4)
En önemli basamaklar (2.4)
Gövde (2.4)
Sondaki basamaklar (2.4)
Az önemli basamaklar (2.4)
Yapraklar (2.4)
Simetrik (2.5)
İki tepedeğerli (2.5)
Tek tepedeğerli (2.5)
Tepedeğerlilik (2.5)
Sola doğru çarpık (2.5)
Sağa doğru çarpık (2.5)
Çarpıklık (2.5)
Basıklık (2.5)
Normal basık (2.5)
Çok basık (2.5)
Az basık (2.5)
Sigma ($\sum$) (2.6)
Merkeze eğilme ölçüleri (2.7)
Konum ölçüleri (2.7)
Tepedeğer (Mod) (2.7)
Ortanca (Medyan) (2.7)
Ortancanın konumu (2.7)
Ortalama (2.7)
Kırpılmış ortalama (2.7)
Dağılım (2.8)
Açıklık (Ranj) (2.8)
Çeyrekler arası açıklık (2.8)
Birinci çeyrek (2.8)
İkinci çeyrek (2.8)
Üçüncü çeyrek (2.8)
Kırpılmış örneklem (2.8)
Winsorize (2.8)
Ortalama mutlak sapma (2.8)
Standart sapma (*s*) (2.8)
Örneklem varyansı ($\mathbf{s}^2$) (2.8)
Evren varyansı ($\sigma^2$) (2.8)
Değişim katsayısı (2.8)
Yansız kestirici (2.8)
Beklenen değer (2.8)
Serbestlik derecesi (2.8)
Kutu grafikleri (2.9)
Kutu-bıyık grafikleri (2.9)
Çeyrek konumu (2.9)
İç çit (2.9)
Ardışık değerler (2.9)
Bıyıklar (2.9)
Onda birlikler (2.11)
Yüzde birlikler (2.11)
Fraktiller (2.11)
Doğrusal dönüşümler (2.12)
Merkezlendirme (2.12)
Yansıma (2.12)
Sapma puanları (2.12)
Standartlaştırma (2.12)
Doğrusal olmayan dönüşümler (2.12)

## Alıştırmalar

Aşağıdaki soruları ister kâğıt kalem, isterseniz bilgisayar programı kullanarak cevaplayabilirsiniz. Bazı sorularda ADD.dat adlı büyük bir veri dosyasını kullanmanız gerekecek. Veri kümesini www.uvm.edu/~dhowell/methods8/DataFiles/Add.dat adresinden indirebilirsiniz. Bu veri kümeleri (Howell ve Huessy, 1985) araştırmasından elde edilen gerçek veridir. Bu çalışma detaylı bir biçimde Ek: 686'ıncı sayfada Data Set adı altında açıklanmıştır.

2.1 Çocukların anlattıkları hikâyeleri dinleyen herhangi biri, yetişkinlerin aksine çocukların hikâyenin genel hatları yerine hikâyedeki eylemlerin sırasını hatırladıklarını bilir. Çocuklar sinemada izledikleri bir filmi anlatırken "daha sonra…" ifadesini sıklıkla kullanırlar. Çok sabırlı bir araştırmacı 50 tane çocuktan belli bir filmi anlatmalarını istedi. Diğer değişkenler arasında, bağımlı değişken olarak çocukların kullandığı "daha sonra…" ifadelerini saydı. Elde ettiği veri aşağıdaki gibidir:

18 15 22 19 18 20 17 12 16 16 17 21 23 18 20 2120 20 15 18 17 19 20 23 22 10 17 19 19 21 20 18 18 24 11 19 31 16 17 15 19 20 18 18 40 18 19 16

a. Bu veri için gruplanmamış frekans dağılımının grafiğine çiziniz.

b. Dağılımın genel şekli nasıldır?

2.2 Makul sayıda aralık kullanarak Alıştırma 2.1'deki veri için histogram oluştururuz.

2.3 Alıştırma 2.1'de verilen veri için gövde-ve-yaprak grafiği oluştururken ne tür zorluklarla karşılaşabilirsiniz?

2.4 Alıştırma 2.1'de sözü edilen çalışmanın diğer bir kısmında araştırmacı aynı veriyi 50 yetişkinden elde etti. Bu veri kümesi aşağıdaki gibidir:

10 12 5 8 13 10 12 8 7 11 11 10 9 9 11 15 12 17 14 10 9 8 15 16
10 14 7 16 9 1 4 11 12 7 9 10 3 11 14 8 12 5 10 9 7 11 14 10 15 9

a. Sadece bu veriye bakarak ne söyleyebilirsiniz? Yetişkinler ve çocuklar hikâyeleri aynı biçimde hatırlıyor gibi mi görünüyorlar?

b. Alıştırma 2.1'deki çocuk verisinde kullandığınız aynı ölçeği kullanarak bu veri için gruplanmamış frekans dağılım grafiğini çiziniz.

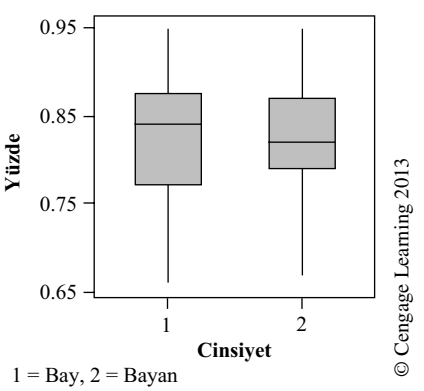
c. Alıştırma 2.1 b'de elde ettiğiniz frekans dağılımıyla bu dağılımı üst üste getiriniz.

2.5 Alıştırma 2.1 ve 2.4'teki veri kümelerini karşılaştırmak için sırt sırta çizilmiş gövde ve yaprak grafiğini kullanınız.

2.6 Sağa çarpık veri kümesi oluşturunuz ve grafiğini çiziniz.

2.7 Gerçekte oluşmuş bir olayı temsil eden çift modlu bir veri kümesi oluşturunuz ve grafiğini çiziniz.

2.8 Benim verdiğim lisans düzeyindeki araştırma teknikleri dersinde, kadınlar genellikle erkeklerden daha başarılıdırlar. Bundan önceki bir senede elde ettiğim sonuçların kutu grafiği aşağıdaki gibidir. Bu kutu grafiklerinden hangi kararlara varabilirsiniz?




1 = Bay, 2 = Bayan

© Cengage Learning 2013

2.9 Alıştırma 2.8'de bayanlar ve baylar için (yaklaşık olarak) birinci ve üçüncü çeyreklerin değerleri kaçtır?

2.10 Aşağıdaki gövde-ve-yaprak grafiği, Alıştırma 2.8'de sözü edilen başarı yüzdeleri kadınlar ve erkekler için ayrı ayrı göstermektedir. Buna göre kadınlar ve erkekler arasındaki farklar için hangi sonuçlara varabilirsiniz?

| Yüzdeler İçin Gövde-ve-Yaprak<br>Cinsiyet = 1 (Bay) $N = 29$<br>Yaprak Birimi = 0,010 | | | Yüzdeler İçin Gövde-ve-Yaprak<br>Cinsiyet = 2 (Bayan) $N = 78$<br>Yaprak Birimi = 0,010 | | |
|---|---|---|---|---|---|
| 3 | 6 | 677 | 2 | 6 | 77 |
| 3 | 6 | | 3 | 6 | 8 |
| 3 | 7 | | 6 | 7 | 000 |
| 5 | 7 | 33 | 10 | 7 | 2233 |
| 7 | 7 | 45 | 15 | 7 | 45555 |
| 7 | 7 | | 15 | 7 | |
| 10 | 7 | 999 | 22 | 7 | 8899999 |
| 12 | 8 | 01 | 34 | 8 | 011111111111 |
| 14 | 8 | 22 | (8) | 8 | 22222233 |
| (4) | 8 | 4455 | 36 | 8 | 445555555 |
| 11 | 8 | 6677 | 27 | 8 | 666777777 |
| 7 | 8 | 8 | 18 | 8 | 888889999 |
| 6 | 9 | | 9 | 9 | 00001 |
| 6 | 9 | 23 | 4 | 9 | 333 |
| 4 | 9 | 4445 | 1 | 9 | 5 |

© Cengage Learning 2013

2.11 Karşılaştığınız bir sonraki 200 kişinin bir ayda izledikleri film sayısının dağılımının şeklinin nasıl olacağını tahmin edersiniz?

2.12 Alıştırmaların başında sözü edilen ve Ekte verilen veri kümesindeki GBD için histogramı çiziniz. (Bu veriye aynı zamanda www.uvm.edu/~dhowell/methods8/DataFiles/Add.dat adresinden ulaşılabilir.)

2.13 Ekte verilen veri kümesindeki ADDSC puanları için bir de yaprak grafiği oluşturunuz.

2.14 Farazi (hipotetik) bir çalışmada bir araştırmacı 10 Avrupalı ve 10 Amerikalının müzik becerilerini 12 dereceli bir ölçek ile değerlendirdi. Avrupalılar için veri kümesi [10 8 9 5 10 11 7 8 2 7]'dir. Bu değişken için $X$ kullanılırsa,

a. $X_3$, $X_5$ ve $X_8$ kaçtır?

b. $\sum X$'i hesaplayınız.

c. b şıkkındaki toplama işlemini, toplam sembolünü kullanarak açıkça yazınız.

2.15 Alıştırma 2.14'teki Kuzey Amerikalılar için veri kümesi [9 9 5 3 8 4 6 6 5 2]'dir . Bu veri için değişken olarak $Y$ sembolü kullanılırsa,

a. $Y_1$ ve $Y_{10}$ kaçtır?

b. $\sum Y$'yi hesaplayınız.

2.16 Alıştırma 2.14'teki veriyi kullanarak,

a. $(\sum X)^2$ ve $\sum X^2$'yi hesaplayınız.

b. $\sum X/N$ ($N$ = gözlem sayısı) hesaplayınız.

2.17 Alıştırma 2.15'teki veri kümesini kullanarak,

a. $(\sum Y)^2$ ve $\sum Y^2$ hesaplayınız.

b. $\dfrac{\sum Y^2 - \dfrac{\left(\sum Y\right)^2}{N}}{N-1}$ hesaplayınız.

c. b şıkkındaki cevabınızın karekökünü hesaplayınız.

d. b ve c şıkkındaki cevaplarınızın birimi nedir?

2.18 Alıştırma 2.14 ve 2.15'teki veri kümelerini iki sütun halinde yan yana yazınız ve her bir sütunu X ve Y olarak adlandırınız. Veri kümelerini eşleşmiş veri olarak ele alınız.

a. $\sum XY$ hesaplayınız.

b. $\sum X \sum Y$ hesaplayınız.

c. $\dfrac{\sum XY - \dfrac{\sum X \sum Y}{N}}{N-1}$ hesaplayınız.

2.19 Alıştırma 2.14 ve 2.15'teki veri kümelerini kullanarak aşağıdaki eşitliklerin doğru olduğunu gösteriniz.

a. $\sum(X + Y) = \sum X + \sum Y$.

b. $\sum X \sum Y \neq \sum XY$.

c. $\sum CX = C \sum X$.

d. $\sum X^2 \neq (\sum X)^2$.

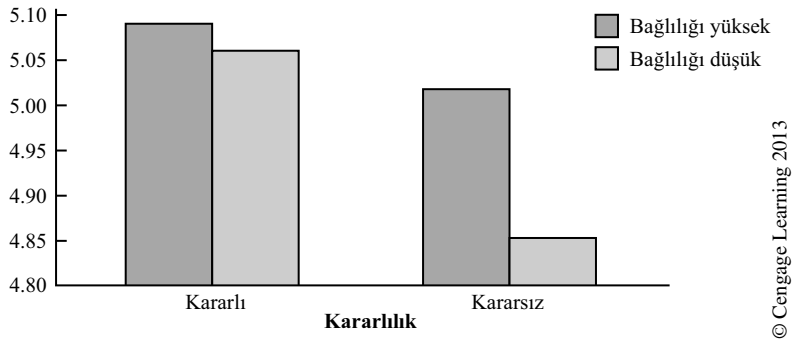
2.20 Tablo 2.1'de (sayfa 17) cevaplama süresi veri kümesi, uyaran karşılaştırması yapabilmek için basamak sayısına göre ayrı ayrı verilmiştir. Her bir veri kümesi için üç ayrı gövde-ve-yaprak gösterimini yan yana oluşturunuz. (Pozitif ve negatif durumları göz ardı ediniz.) Bu üç durum için, cevaplama sürelerinde ne gibi farklılıklar gözlediniz?

2.21 Tablo 2.1'de verilen veri kümesinin elde edildiği orijinal çalışmada Sternberg bireylerin bilgiyi sıralı mı yoksa eş zamanlı mı işlediklerini araştırmak istedi. Sternberg eğer insanlar bilgiyi eşzamanlı işlerlerse onların test uyaranını, karşılaştırma uyaranındaki bütün basamaklarla aynı anda karşılaştıracaklarını ve test sayısının karşılaştırma kümesinin bir parçası olup olmadığına karar verme sürelerinin kümedeki basamak sayısına bağlı olmayacağını düşündü. Eğer insanlar bilgiyi sıralı olarak işlerlerse, karar verme süresi karşılaştırma kümesindeki sayının basamak sayısıyla artacaktı. Buna göre alıştırma 2.20'de çizilen gösterimler sizce hangi hipotezi destekliyor?

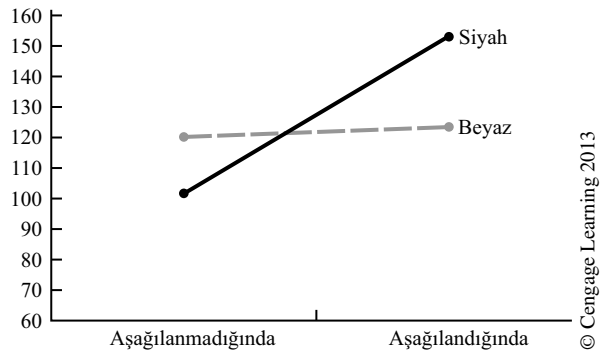
2.22 Alıştırma 2.20'deki cevaplama sürelerinin üç gösterimine ek olarak, insanların bilgiyi nasıl işlediklerini incelemek için Tablo 2.1'deki veriyi daha başka nasıl kullanırdınız?

2.23 İstatistiksel analizlerde en sık kullanılan varsayımlardan biri gözlemlerin birbirinden bağımsız olduğudur. (Bir gözlemin değerinin ne olduğunu bilmeniz size diğer gözlemin büyüklüğü hakkında bilgi vermez.) Tablo 2.1'deki veri kümesi hakkındaki bilginize ve nasıl toplandıkları bilgisine dayanarak bu veriyi nasıl karakterize edersiniz?

2.24 Aşağıdaki şekil kararlı ve kararsız sosyal gruplarda yetişen insan olmayan primatların bağışıklık yanıtlarını inceleyen Cohen, Kaplan, Cunnick, Manuck ve Rabin (1992) makalesinden uyarlanmıştır. Her bir grupta hayvanlar diğer hayvanların fiziksel olarak yakınında bulunma sürelerine göre bağlılığı yüksek ve düşük olanlar olarak gruplanmışlardır. Bağışıklık ölçeğindeki yüksek puanlar hastalıklara karşı daha bağışık olmayı gösterir. Buna göre bu sonuçları nasıl yorumlarsınız?




2.25 Roger ve Prentice-Dunn'ın (1981) biyo-geri bildirim çalışmasında denekler gruptaki diğer deneklere şok uyguladılar. Deneklerin araştırmacı tarafından aşağılandığı ve aşağılanmadığı durumlarda beyaz ırktan ve siyah ırktan deneklere uyguladıkları şok miktarı ölçüldü. Sonuçlar aşağıdaki şekilde verilmiştir. Buna göre sonuçları yorumlayınız.




2.26 Aşağıdaki veri kümesi 1976'dan 2007'ye kadar Amerika'da üniversiteye kayıt olan öğrenci sayısını genel nüfus sayımında kullanılan kategorilere göre vermektedir. Veri kümesinin grafiğini etnik kökene göre Amerika'da üniversite öğrenci sayısındaki değişimi gösterecek biçimde çiziniz. (Veri 1000'li sayılarla ifade edilmiştir.) Veri, http://nces.ed.gov/fastfacts/display.asp?id=98 adresinden alınmıştır.

Derece almak için öğrenim kurumlarına başvuran, kayıt yaptıran öğrencilerin ırk/etnik kökene göre yüzdelerinin dağılımı: Seçilen yıllar, güz dönemi 1976-2007.

| | Yıl | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Irk/etnik köken | 1976 | 1980 | 1990 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
| Beyaz | 82,6 | 81,4 | 77,6 | 68,3 | 67,1 | 66,7 | 66,1 | 65,7 | 65,2 | 64,4 |
| Toplam azınlık | 15,4 | 16,1 | 19,6 | 28,2 | 29,4 | 29,8 | 30,4 | 30,9 | 31,5 | 32,2 |
| Siyah | 9,4 | 9,2 | 9,0 | 11,3 | 11,9 | 12,2 | 12,5 | 12,7 | 12,8 | 13,1 |
| Meksikalı | 3,5 | 3,9 | 5,7 | 9,5 | 10,0 | 10,1 | 10,5 | 10,8 | 11,1 | 11,4 |
| Asyalı | 1,8 | 2,4 | 4,1 | 6,4 | 6,5 | 6,4 | 6,4 | 6,5 | 6,6 | 6,7 |
| Amerikalı yerliler | 0,7 | 0,7 | 0,7 | 1,0 | 1,0 | 1,0 | 1,0 | 1,0 | 1,0 | 1,0 |
| Amerikalı olmayanlar | 2,0 | 2,5 | 2,8 | 3,5 | 3,6 | 3,5 | 3,4 | 3,3 | 3,4 | 3,4 |

Yükseköğretimdeki öğrencilerin zamana göre değişimini gösteren grafiği çiziniz.

2.27 Aşağıdaki veri kümesi Amerika Birleşik Devletleri'nde 1981'den 1990 yılına kadar 13-29 yaş arası bireylerde rastlanan AIDS vakalarının sayısını göstermektedir. Bu yıllar AIDS'in ilk defa yaygın olarak gözlendiği zamanlardı. Zaman içindeki eğilimi gösteren bir grafik çiziniz. (Her sayı 1000 vakayı gösterir ve iki farklı veri kaynağından alınmıştır.)

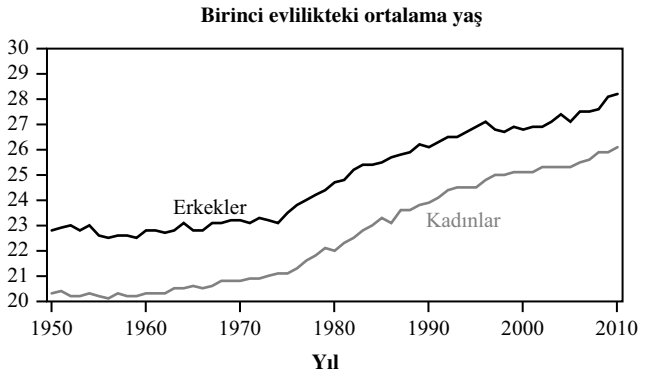
| Yıl | Vakalar |
|---|---|
| 1981–1982 | 196 |
| 1983 | 457 |
| 1984 | 960 |
| 1985 | 1685 |
| 1986 | 2815 |
| 1987 | 4385 |
| 1988 | 6383 |
| 1989 | 6780 |
| 1990 | 5483 |

© Cengage Learning 2013

(Amerika Birleşik Devletleri'nde AIDS/HIV görülme sıklığı düşmektedir. 2006 yılında ABD'de 39,5 milyon insan AIDS/HIV virüsü taşımaktadır. Bu, editörün küçük bir yorumudur.)

2.28 Dünyadaki AIDS/HIV vakalarının daha güncel verisine http://data.unaids.org/pub/EpiReport/2006/2006_EpiUpdate_en.pdf adresinden ulaşılabilir. Buna göre Amerika Birleşik Devletleri'ndeki vaka oranları dünyanın geri kalanıyla karşılaştırıldığında nasıl değişmektedir?

2.29 Aşağıdaki http://www.bsos.umd.edu/socy/vanneman/socy441/trends/marrage.html adresinden alınan veri kümesi, 50 yıllık bir sürede toplumdaki evlilik yaşının nasıl değiştiğini göstermektedir. Veride nasıl bir eğilim görüyorsunuz? Bunun nedeni ne olabilir?




2.30 Ortalamanın ortancadan daha büyük olduğu bir veri kümesi oluşturunuz.

2.31 Alıştırma 2.6'da oluşturduğunuz sağa çarpık veri kümesinde ortalama, ortancanın altında mı yoksa üstünde mi?

2.32 Ortalama ve ortancanın eşit olduğu, fakat tepedeğerin bu değerlerden farklı olduğu tek tepedeğerli bir veri kümesi oluşturunuz.

2.33 Düz koridorları olan bir labirentte koşan 15 farenin önceden belirlenen ölçüte ulaşması için yapmaları gereken deneme sayıları aşağıda verilmiştir:

| Ölçüte ulaşmak için gereken deneme sayısı: | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|---|---|---|---|---|---|---|---|
| Fare sayısı (frekans): | 1 | 0 | 4 | 3 | 3 | 3 | 1 |

Deneme sayılarının tepedeğerini ve ortancasını hesaplayınız.

2.34 Aşağıda verilen veri kümesinde her bir puandan bir sabit (örn. 5) çıkarıldığında merkeze eğilme ölçülerinin hepsinin bu sabit kadar azalacağını gösteriniz.

[8, 7, 12, 14, 37]

2.35 Aşağıdaki veri kümesinde her bir puan bir sabit ile çarpıldığında merkeze eğilme ölçülerinin hepsinin bu sabitle çarpılacağını gösteriniz.

[8 3 5 5 6 2]

2.36 Ortalaması 8,6 olan 10 sayıdan oluşan bir örneklem oluşturunuz. Bu daha önceden bahsedilen serbestlik derecesiyle nasıl ilişkilidir?

2.37 Aşağıdaki betimsel istatistik çıktısı Ek'te açıklanan veri kümesinden GBO ve ADDSC'ye aittir. Burada verilen merkeze eğilme ölçüleri, Alıştırma 2.12 ve 2.13'teki cevaplarda tahmin ettiklerinizle karşılaştırıldığında nasıldır?

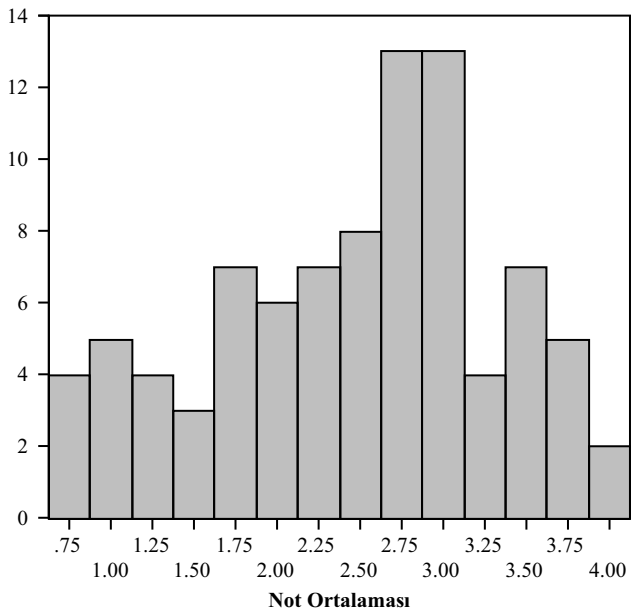
Betimsel İstatistikler

| | ADDSC | GBO | Geçerli N (listeli) |
|---|---|---|---|
| N | 88 | 88 | 88 |
| Minimum | 26 | 1 | |
| Maksimum | 85 | 4 | |
| Ortalama | 52,60 | 2,46 | |
| Stan. Sapma | 12,42 | ,86 | |
| Varyans | 154,311 | ,742 | |

© Cengage Learning 2013

ADDSC ve GBO İçin Betimsel İstatistikler

2.38 Bir veya iki cümleyle aşağıdaki grafiğin örneklemimizdeki öğrencilerin not ortalaması hakkında ne söylediğini açıklayınız.




© Cengage Learning 2013

2.39 SPSS programını kullanarak bir önceki alıştırmada verilen histogramın üzerine normal dağılımı çiziniz. (İpucu: Bu, grafik işlemler menülerinden kolaylıkla yapılabilir.)

2.40 Alıştırma 2.1'deki veri kümesi için ranj, varyans ve standart sapmayı hesaplayınız.

2.41 Alıştırma 2.4'teki veri kümesi için ranj, varyans ve standart sapmayı hesaplayınız.

2.42 Alıştırma 2.40 ve 2.41'deki cevaplarınızı karşılaştırınız. Çocuklar için elde ettiğiniz standart sapma yetişkinlerinkiyle karşılaştırıldığında oldukça farklı mıdır?

2.43 Alıştırma 2.1'deki puanların yüzde kaçı ortalamanın artı-eksi iki standart sapma aralığındadır?

2.44 Alıştırma 2.4'teki puanların yüzde kaçı ortalamanın artı-eksi iki standart sapma aralığındadır?

2.45 Alıştırma 2.34 ve 2.35'te gösterilen sonuçları kullanarak, aşağıda verilen veri kümesini standart sapması 1.00 olacak biçimde dönüştürünüz.

[5 8 3 6 9 9 7].

2.46 Alıştırma 2.4'teki veri kümesi için kutu grafiği oluşturunuz.

2.47 Ek'te verilen veri kümesindeki ADDSC için kutu grafiği oluşturunuz.

2.48 Alıştırma 2.1 ve 2.4'te çocukların ve yetişkinlerin "ve daha sonra…" kullanımlarındaki değişkenliği karşılaştırmak için değişim katsayısını hesaplayınız.

2.49 Ek'teki veri kümesindeki GPO'nun ortalaması 2,456 ve standart sapması 0,8614'tür. Bu bölümde tanımlanan değişim katsayısını hesaplayınız.

2.50 Google'da araştırıp değişim katsayısının rapor edildiği bir araştırma bulunuz.

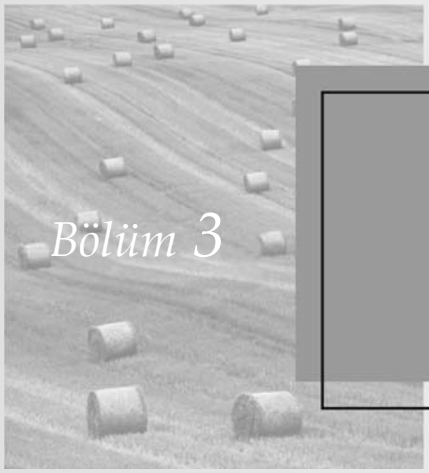
2.51 Tablo 2.6'da verilen veri kümesi için %10 kırpılmış ortalamayı hesaplayınız.

2.52 Tablo 2.6'da verilen veri kümesi için %10 Winsorize edilmiş standart sapmayı hesaplayınız.

2.53 Tablo 2.1'de verilen pozitif ve negatif cevaplama süresi olayları arasındaki farkları göstermek için kutu grafiği çiziniz.

2.54 Hangi durumlarda bir dönüşüm, dağılımın şeklini değiştirir?

2.55 SAS veya S-Plus ile çekirdek yoğunluk grafiğinin nasıl çizileceğini internette Google'da arayınız.



*Bölüm 3*

# Normal Dağılım

## Hedefler

Normal dağılım kavramını geliştirme ve bir örneklemin normalliğine nasıl karar verilebileceği; ayrıca normal dağılımın gözlemler hakkında yorum yapmakta nasıl kullanılabileceği.

## Konular


**3.1** Normal Dağılım
**3.2** Standart Normal Dağılım
**3.3** Normal Dağılım Tablosunun Kullanılması
**3.4** Bir Gözleme Ait Olası Sınırların Belirlenmesi
**3.5** Veri Kümesinin Dağılımının Normal Olup Olmadığının Değerlendirilmesi
**3.6** z ile İlgili Ölçüler

ÖNCEKİ BÖLÜMLERDEKİ tartışmalardan da açıkça anlaşılıyor ki dağılımlar -veri kümelerinin dağılımları, evrenlerin teorik dağılımları ve örneklem dağılımları- ile bir hayli ilgileniyor olacağız. Dağılımların alabileceği birçok şekil içerisinde, bizim amaçlarımız için en önemlisi **normal dağılım** olarak bilinen dağılım türüdür.

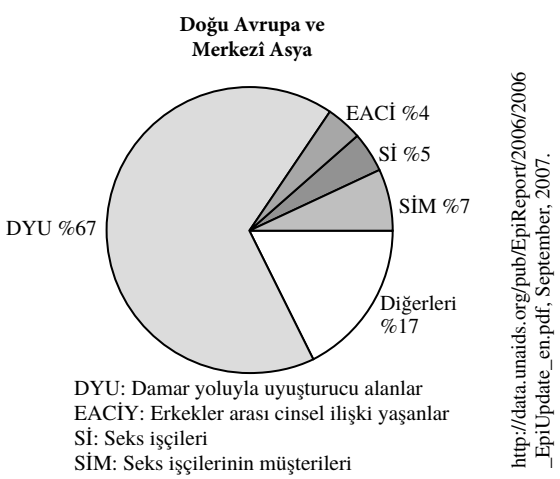
**normal dağılım**

Normal dağılımın ayrıntısına girmeden önce, sadece normal dağılımla değil, genel olarak dağılımlarla neden bu kadar çok ilgilendiğimizi açıklamak için kısa bir parantez açalım. Dağılımlar ile olasılıklar arasındaki önemli ilişki, burada en önemli faktördür. Eğer olayların (veya örneklem istatistiklerinin) dağılımı hakkında bir şeyler biliyorsak bu olayların (veya istatistiklerin) olması olasılıkları hakkında da bir şeyler biliyoruz demektir. Konuyu en sade haliyle görmek için aşağıdaki pasta grafiğini ele alalım. (Bu kitapta pasta grafiği sadece burada görülecektir, çünkü farklı açılarda duran pasta dilimlerinin büyüklüklerinin kıyaslanması çok zordur. Veri kümesini sunmak için çok daha iyi yöntemler var, ancak burada pasta grafiği işe yarar bir amaca hizmet ediyor.)

Şekil 3.1'de görülen pasta grafiği AIDS/HIV Birleşmiş Milletler Ortak Programı tarafından hazırlanan bir rapordan alınmıştır. (http://data.unaids.org/pub/EpiReport/2006/2006_EpiUpdate_en.pdf adresinden 2007 Eylül'ünde erişilmiştir.) Grafik, Doğu Avrupa ve Orta Asya'da AIDS/HIV'in bulaşma kaynaklarını gösteriyor. Dikkate değer biçimde, görülüyor ki dünyanın bu bölgesinde AIDS/HIV vakalarının büyük çoğunluğu damar yoluyla uyuşturucu kullanımından kaynaklanıyor. (Latin Amerika, ABD, Güney ve Güneydoğu Asya'da durum böyle değildir; o bölgelerde ilgili yüzde yaklaşık olarak %20'dir. Ancak biz buradaki veriye odaklanacağız.)

Şekil 3.1'de gördüğünüz gibi, HIV'li insanların %67'si bu virüsü damar yoluyla uyuşturucu alarak (DYU) kapmıştır. Vakaların %4'ü erkekler arası cinsel ilişki (EACİ) yoluyla bulaşmıştır. Vakaların %5'i seks işçilerinden (Sİ), %7'si ise bu sektörün müşterilerinden (SİM) oluşmaktadır. Sınıflandırılmayan ve diğer vakaların oranı ise %17'dir. Ayrıca her bir kategorideki vakaların yüzdesinin, bu kategorileri gösteren pasta dilimlerinin alanlarında da doğrudan yansıtıldığı görülüyor. Her bir kısımdaki kişi yüzdesi, o kısmın işgal ettiği alan ile doğru orantılıdır. Ayrıca, pastanın toplam alanını 1,00 birim kabul edersek, her bir kısmın alanı o kısımdaki gözlemlerin oranına eşittir.

Alanlar hakkında konuşmaktan olasılıklar hakkında konuşmaya geçiş kolaydır. Olasılık kavramı Bölüm 5'te açıklanacaktır, ancak tam bir tanım vermeden de pasta grafiğindeki alanlarla ilgili önemli bir noktaya değinebiliriz. Şimdilik olasılığı sadece günlük kullanım-

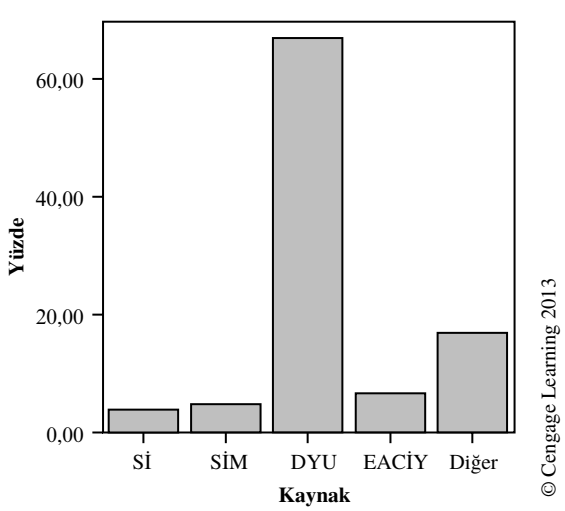



**Şekil 3.1** Farklı evrenlerdeki HIV bulaşma kaynaklarını gösteren pasta grafiği

daki anlamı ile düşünün (yani bir olayın gerçekleşmesinin ne kadar muhtemel olduğu). Bu bakış açısından şu sonucu çıkarmak mantıklıdır: HIV/AIDS hastalarının %67'si bu hastalığı damar yoluyla uyuşturucu kullanmaktan kaptığına göre, eğer HIV/AIDS hastaları listesi içinden tesadüfi olarak bir hastayı seçseydik, seçtiğimiz hastanın virüsü damar yoluyla uyuşturucu kullanmaktan kapmış olması olasılığı 0,67 olacaktı. Farklı terimlerle ifade edilirse; DYU için kullanılan kısmın alanı pastanın alanının %67'si olduğuna göre, bir kişinin bu kısımda olması olasılığı 0,67'dir.

Bu pasta grafiği bize alanların toplanmasını da inceleme imkânı veriyor. Vakaların %5'i Sİ, %7'si SİM ve %4'ü EACİ olarak sınıflandığına göre vakaların 5 + 7 + 4 = %16'sı hastalığı cinsel faaliyetten kapmıştır. (Dünyanın bu bölgesinde HIV/AIDS kaynakları, Batı dünyasında yaşayan insanların bekleyeceğinden çok farklıdır. Dolayısıyla HIV/AIDS önleme programları ona göre şekillendirilmelidir.) Bir başka deyişle, birden çok kategorideki kişilerin yüzdesini, her bir kategorinin yüzdesini toplayarak bulabiliriz. Aynısı, alanlar için de geçerlidir; cinsel faaliyetle ilişkili bulaşmaların yüzdesini Sİ, SİM ve EACİ'yi gösteren alanları toplayarak bulabiliriz. Son olarak, alanları toplayarak yüzdeleri bulabiliyorsak, olasılıkları da alanları toplayarak bulabiliriz. Dolayısıyla, Doğu Avrupa veya Orta Asya'da yaşıyorsanız, HIV/AIDS'i cinsel faaliyet sonucu kapmış olma olasılığınızı ilgili üç kısmın alanlarını (veya olasılıklarını) toplayarak bulabilirsiniz.

Veriyi sunmanın pasta grafiği dışında yolları da vardır. Bunlardan iki tanesi (Bölüm 2'de açıklanmış olan) histogram ve onun yakın kuzeni **çubuk grafiği**dir. Şekil 3.2, Şekil 3.1'in çubuk grafiği biçiminde yeniden çizilmiş halidir. Her ne kadar bu şekil yeni bir bilgi içermiyorsa da bu şeklin pasta grafiğine karşı iki avantajı vardır. İlki, kategorileri karşılaştırmak daha kolaydır, çünkü farklı açılarda duran yayların uzunluğunu kıyaslamaya çalışmak yerine, bakmamız gereken tek şey çubukların boyudur. İkincisi, çubuk grafiği sıklıkla ilgileneceğimiz diğer dağılımlara görsel olarak benziyor. Yatay eksende değişik düzeyler veya kategoriler, düşey eksende ise bu kategorilere karşılık gelen yüzdeler (veya frekanslar) bulunuyor. (Ancak, burada da olduğu gibi, çubuk grafiğinde *X* ekseni, sınıflama ölçeğindedir. Bu, histogram için doğru değildir.) Burada yine dağılımdaki farklı alanla-




**Şekil 3.2** Farklı HIV/AIDS kaynaklarının yüzdelerini gösteren çubuk grafiği

rın olasılıklarla ilişkili olduğu görülebilir. Dahası, tıpkı pasta grafiklerinde yaptığımız gibi alanları anlamlı bir biçimde toplayabiliriz. Sıklıkla kullanılan dağılımlara, özellikle normal dağılıma geçtiğimizde de alanlar, yüzdeler ve olasılıklara dair bu ilkeler hemen hemen hiç değişmeden geçerli olacaktır.

## 3.1 Normal Dağılım

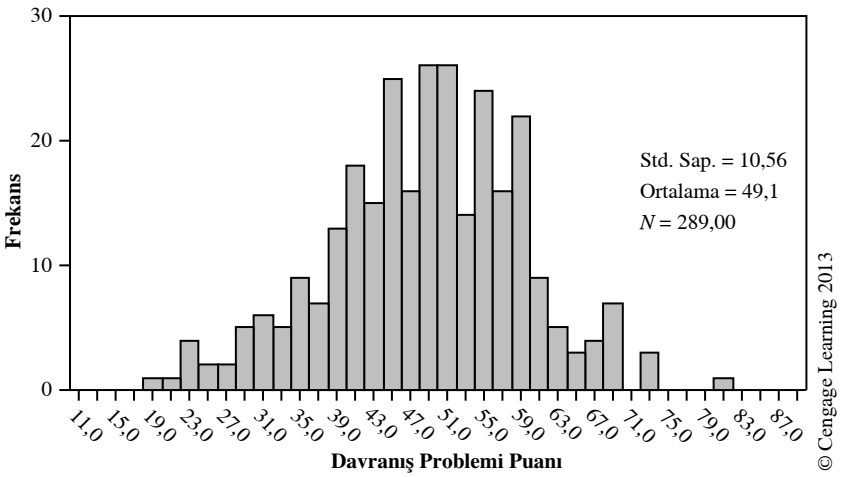
Şimdi normal dağılıma daha yakından bakacağız. Normal dağılımın karşılaşacağımız en önemli dağılımlardan biri olduğu daha önce belirtilmişti. Bunun birçok nedeni vardır:

1. İlgilendiğimiz pek çok bağımsız değişkenin evrende normal dağıldığı varsayılır. Yani, evrenin tamamını gözlemler olarak alsaydık, elde edilecek dağılımın normal dağılıma çok benzeyeceğini söyleyebilirdik.
2. Eğer bir değişkenin normal dağıldığını varsayabilirsek bu bölümde tartışılan teknikler, bu değişkenin değerleri hakkında (yaklaşık veya kesin) yorumlar yapmamıza imkân verecektir.
3. Belirli bir evrenden seçilen sonsuz sayıda örneklemin ortalamalarının dağılımının yaklaşık olarak normal olduğu varsayımı pek çok koşul altında gösterilebilir. Bu dağılıma ortalamanın örneklem dağılımı denir. Örneklem dağılımı bu kitabın geri kalan bölümlerinde genişçe ele alınacak ve kullanılacaktır.
4. Kullanacağımız tekniklerin çoğunun türetilmesinin aşamalarından herhangi bir yerinde gözlemlerin evreninin (veya ölçme hatalarının) normal dağıldığı varsayımı yer alır.

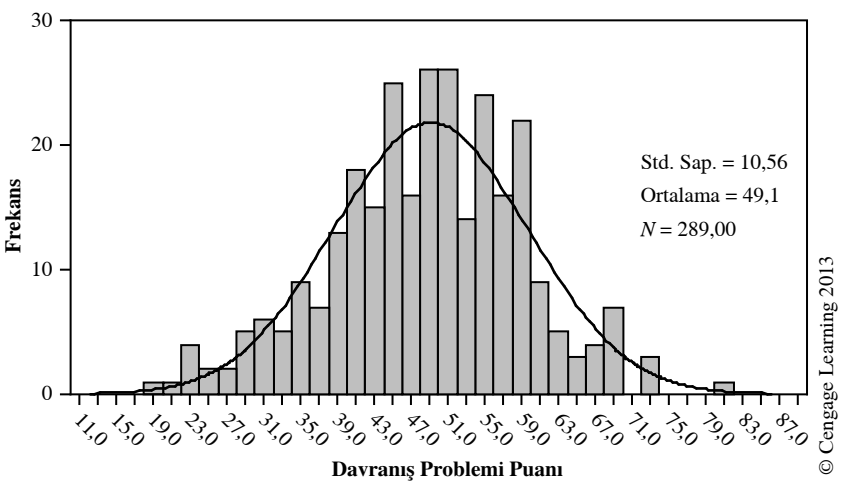
Normal dağılımı tanıtmak için yaklaşık normal dağılan (daha çok gözlemi olsaydı normal dağılıma daha çok benzeyecek olan) bir başka veri kümesine daha bakacağız. Bakacağımız veri kümesi sıkça kullanılan çok boyutlu davranış problemleri ölçeği olan Achenbach'ın Gençler İçin Kendini Değerlendirme Ölçeği (Achenbach, 1991b) ile elde edilmiştir. Çocuk tarafından bildirilen davranış problemlerinin ciddiyetine göre ağırlıklandırılmış bir toplamını ifade eden Davranış Problemleri Toplamı boyutuna odaklanacağız. (Davranış problemi kategorilerinden bazıları şunlardır: "Tartışır", "Dürtüsel", Gösteriş yapar" ve "Sataşır".) Şekil 3.3, 289 ortaokul öğrencisinden elde edilen veri kümesinin histogramıdır. Yüksek puanlar, daha çok davranış problemine işaret etmektedir. Görebildiğiniz gibi dağılımın merkezi 50 civarında olup bu değerin her iki yönünde 25 ile 75 arasında oldukça simetrik dağılmaktadır. Dağılımın standart sapması yaklaşık olarak 10'dur. Dağılım tam düzgün değil –tümsekleri çukurları var– ancak oldukça pürüzsüz, ortada yüksek uçlara doğru düşüyor. (Dağılım ortalaması ve standart sapmasının gerçek değerleri sırasıyla 49,1 ve 10,56'dır.)

Bu dağılımda 52-54 aralığına ve 54-56 aralığına düşen katılımcılar toplanırsa, puanı 52-56 arasında olan öğrencilerin sayısının 54 olduğu görülür. Bu örneklemde 289 gözlem olduğuna göre gözlemlerin 54/289 = %19'u bu aralığa düşmektedir. Bu durum, daha önce alanların toplamıyla ilgili olarak yapılan yorumların bir örneğidir.

Bu dağılımı alıp üzerine bir normal dağılım çizebiliriz. Bu, daha sonra göreceğimiz gibi, normalliği değerlendirmek için en iyi yol olmamasına rağmen, bir örneklemin normalliğini kabaca değerlendirmek için sıklıkla kullanılmaktadır. Şekil 3.4'teki ham veri-




**Şekil 3.3** Davranış Problemi puanlarını gösteren histogram



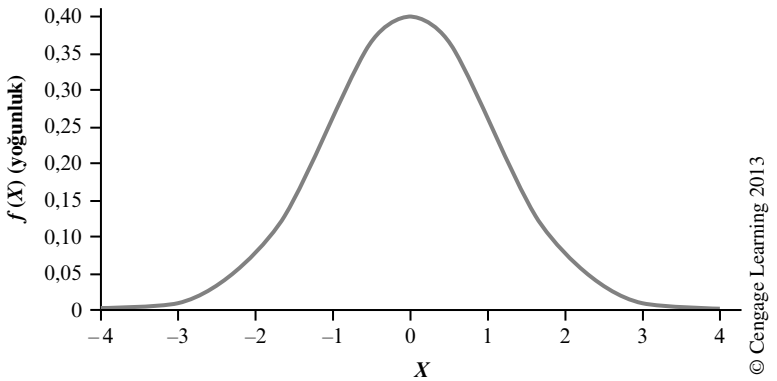

**Şekil 3.4** Davranış Problemi puan dağılımını temsil eden bir karakteristik normal dağılım

**apsis**
**ordinat**

nin üzerine çizilmiş olan düzgün dağılım karakteristik normal dağılımdır. Bu, "çan eğrisi" olarak anılan, tek modlu, simetrik ve sınırları $\pm\infty$ olan bir dağılımdır. **Apsis** veya yatay eksen $X$'in farklı olası değerlerini gösterir. **Ordinat** veya düşey eksen de yoğunluk olarak adlandırılır. Yoğunluk kavramı frekans veya $X$'in olma olasılığı ile ilişkilidir (fakat aynı değildir). Yoğunluk kavramı bir sonraki bölümde ayrıntılı olarak tartışılmaktadır. (Az önce yaptığımız gibi, histogramın üzerine bir normal dağılım eğrisi çizdirmek, dağılımın şeklini değerlendirmeye yardımcı olsa da verinin normal dağılıp dağılmadığını değerlendirmek için daha iyi yollar vardır. Bu bölümün devamında Q-Q (yüzdelik-yüzdelik) grafiklerini tartışacağız ve normalliği değerlendirmenin görece basit bir yolunu göreceğiz.)

Normal dağılım çoğu zaman apsisinde $X$, ordinatında yoğunluk bulunan genel bir dağılım üzerinden tartışılır. Böyle bir dağılım Şekil 3.5'te verilmiştir.

Normal dağılım uzun bir geçmişe sahiptir. İlk olarak şans oyunlarının sonuçlarını araştırmak isteyen DeMoivre (1667–1754) tarafından incelenmiştir. Normal dağılım tam ola-




**Şekil 3.5** Apsisinde *X*, ordinatında yoğunluk bulunan bir karakteristik normal dağılım

rak Pierre-Simon Laplace (1749–1827) tarafından tanımlanmış, bugün sıklıkla kullanılan biçimime Carl Friedrich Gauss (1777–1855) tarafından getirilmiştir. Her ikisi de astronomik gözlemlerdeki hataların dağılımı ile ilgileniyordu. Esasen, normal dağılım Gauss dağılımı ve "normal hatalar yasası" olarak da bilinir. Belçikalı bir astronom, Adolph Quetelet (1796–1874) normal dağılımı ilk kez sosyal ve biyolojik veri kümelerine uygulamıştır. Görünüşe göre yapacak daha iyi bir şeyi yokmuş ki İskoç askerlerin göğüs, Fransız askerlerin boy ölçümlerini toplamıştır. Her iki veri kümesinin de yaklaşık olarak normal dağıldığını bulmuştur. Qutelet, dağılımın ortalamasını doğanın hedeflediği ideal, ortalamanın her iki yanındaki ölçüm sonuçlarını da hata (doğanın idealinden sapma) olarak yorumlamıştır. (Benim gibi 5´8´´ boyundaki bir erkek için daha uzun erkeklerin doğanın hatası olduğunu düşünmek oldukça avutucu.) Artık ortalamayı doğanın ideali olarak düşünmesek de ortalama etrafındaki değişkenliği kavramsallaştırmak için kullanışlı bir yoldur. Aslında, ortalamadan sapmaları tanımlamak için hâlâ *hata* terimini kullanırız. Francis Galton (1822–1911) Quetelet'nin düşüncelerini daha öteye taşımış ve normal dağılıma psikoloji teorisinde, özellikle zihinsel beceriler teorisinde, merkezî bir rol vermiştir. Bazıları Galton'ın bu çabasında gereğinden fazla başarılı olduğunu iddia eder, zira ölçüm sonuçlarının –öyle olmasa bile– normal olduğunu varsaymaya meyilliyiz. Konuyu burada tartışmayacağız.

$\pi$ ve $e$ sabit sayılar ($\pi$ = 3,1416 ve $e$ = 2,7183 ), $\mu$ ve $\sigma$ sırasıyla ortalama ve standart sapma olmak üzere normal dağılım matematiksel olarak

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}(e)^{-(x-\mu)^2/2\sigma^2}$$

ifadesiyle tanımlanır. Eğer $\mu$ ve $\sigma$ biliniyorsa, her $X$ değeri için ordinat, $f(X)$, formülde sayılar yerine konarak kolayca bulunabilir. Göründüğü kadar zor değil ancak pratikte bu hesaplamayı yapmanız muhtemelen hiç gerekmeyecek. Bu dağılımın toplamlı (kümülatif) biçimi tablolanmıştır ve ihtiyacımız olan bilgiyi bu tablodan okuyabiliriz.

Matematiksel analiz dersleri almış olanlarınız hatırlayabilir. $X$'in iki noktası ($X_1$ ve $X_2$ diyelim) arasında eğri altında kalan bölgenin alanı, dolayısıyla tesadüfi olarak seçilen bir puanın bu aralıkta çıkması olasılığı fonksiyonun $X_1$'den $X_2$'ye kadar integrali alınarak bulunabilir. Böyle bir ders almamış olanlarınız rahat olabilir zira bu işlemlerin yapılmış olduğu hazır tablolar kolayca bulunabilir. Böyle bir tablo Ek $z$'de verilmiştir (sayfa 714).

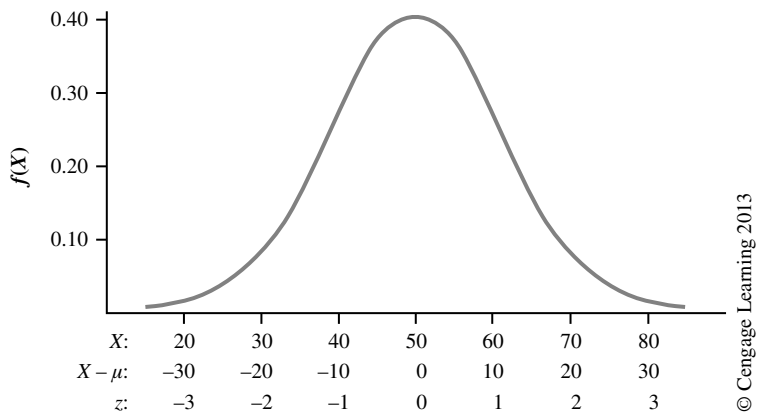
Böyle bir dağılımın neden tablolara döküldüğünü merak etmekten mazur görülebilirsiniz. Bir dağılımın sıklıkla kullanılması (veya varsayılması) bu dağılımdaki bilgileri içeren bir tablonun eklenmesini gerektirmez. Sebep oldukça basittir. Ek $z$'yi kullanarak, evrenden tesadüfi olarak seçilen bir puanın $X_1$ ile $X_2$ arasında çıkması olasılığını kolayca bulabiliriz. Böylece, uygun tablo yardımıyla çeşitli sorulara cevap olarak olasılık cümleleri kurabiliriz. Bu konu, bir çocuğun normal çocuklar evreninden önemli ölçüde farklı olup olmadığını sorduğumuzda veya bir çıktının (mesela ortalamanın) aşırı bir değer olup olmadığını sorduğumuzda ortaya çıkar. Bu konular kitap boyunca farklı bölümlerde yeniden görülecektir.

## 3.2 Standart Normal Dağılım

**standart normal dağılım**

Normal dağılım tablosu yapmaya çalışınca problem çıkar, çünkü normal dağılımın değerleri dağılımın ortalaması ve standart sapmasına ($\mu$ ve $\sigma$) göre farklı değerler alır. İşi düzgün yapmak için $\mu$ ve $\sigma$'nın alabileceği tüm değerler için birer tablo yapmamız gerekirdi, ki bu kesinlikle uygulanabilir değildir. Bu problemin çözümü, ortalaması "0", standart sapması "1" olan **standart normal dağılım** ile çalışmaktır. Bu dağılım $N(0,1)$ olarak gösterilir. $N$ harfi dağılımın normal olduğunu, "0" $\mu$'nün değerini, "1" ise $\sigma^2$'nin değerini ifade eder. Gösterimin genel hali $N(\mu, \sigma^2)$ biçimindedir. Ekteki standart normal dağılım tablosu ile herhangi bir normal dağılımı standart hale getirme (veya tersi) kuralları elimizde olduğu sürece herhangi bir normal dağılım altında kalan alanları Ek $z$'yi kullanarak hesaplayabiliriz.

Şekil 3.6'daki ortalaması 50, standart sapması 10 (varyansı 100) olan dağılımı ele alalım. Bu dağılım Achenbach'ın Gençler İçin Kendini Değerlendirme Ölçeği, Davranış Problemi toplam puanları evrenini göstermektedir. Şekil 3.3 ve 3.4'teki dağılım bu evrenin bir örneklemidir. Şekil 3.6'daki eğrinin altında kalan alanlara dair bir şeyler biliyorsak, Davranış Problemi puanlarının olasılıklarına dair bir şeyler söyleyebiliriz. Örneğin şu kadar yüksek puan, evrenin yalnızca %5'i veya %10'u tarafından alınmıştır. Böyle bir ifadeyi neden kullandığımızı merak edilebilirsiniz. Ancak genellikle teşhis için aşırı puanlar ile tipik puanların ayırt edilmesi önemlidir.




**Şekil 3.6** Apsis ekseninde farklı dönüşümlerin uygulandığı bir normal dağılım

Kolayca ulaşılabilen normal dağılım tabloları sadece ***standart*** normal dağılım tablolarıdır. Dolayısıyla, herhangi bir bireyin belirli bir puandan yüksek alma olasılığına dair sorulara cevap verebilmek için önce Şekil 3.6'daki dağılımı veya en azından onun üzerindeki bazı noktaları standart normal dağılıma dönüştürmeliyiz. Yani demek isteriz ki ortalaması 50, varyansı 100 olan bir normal dağılımdaki –genellikle $\boldsymbol{N}(50, 100)$ olarak gösterilir– $X_i$ puanı; ortalaması 0, varyansı ve standart sapması 1 olan bir normal dağılımda –$\boldsymbol{N}(0, 1)$ olarak gösterilir– $z_i$ puanına denk gelir. Böylece, $z_i$ için doğru olan her şey $X_i$ için de doğrudur. (İstatistikçiler $z$'yi bazen **odak puan**lar olarak adlandırırlar, çünkü $z$'nin *dağılımı* $\mu$ ve $\sigma^2$'den bağımsızdır. Dağılım bu iki parametreden bağımsız olarak her zaman normal olacaktır. Ancak Bölüm 9'da göreceğimiz korelasyon katsayısı esas bir miktar değildir çünkü evrendeki korelasyonun değerine göre farklı şekillerde dağılıma sahip olur.)

**odak puan**

Alıştırma 2.34'ten biliyoruz ki her bir puandan belli bir sabit sayının çıkarılması ortalamayı da o sayı kadar düşürür. Dolayısıyla $X$'in her değerinden 50 (ortalamayı) çıkarırsak yeni ortalama 50 – 50 = 0 olur. Bir başka ifadeyle dağılımın merkezini 0'a aldık. (Genel olarak, $(X - \mu)$ dağılımının ortalaması 0'dır. $(X - \mu)$ puanlarına **sapma puanları** denir çünkü ortalamadan sapmayı gösterir.) Bu dönüşümün etkisi Şekil 3.6'da apsis ekseninin ikinci satırında gösterilmiştir. Henüz yolun yarısındayız, çünkü ortalama 0 olsa da standart sapma ($\sigma$) hâlâ 10'dur. Bir dağılımdaki tüm değerleri bir sabit sayıyla (örneğin 10) çarpar veya bir sabit sayıya bölersek standart sapmayı da aynı sabitle çarpmış veya bölmüş oluruz. Dolayısıyla eğer tüm puanları 10'a bölersek, standart sapma 10'a bölünmüş ve standart sapma 10/10 = 1 olur, ki bu tam bizim istediğimizdir. Bu dönüştürülmüş dağılımı $z$ dağılımı olarak tanımlayacağız ve aşağıdaki gibi yazacağız:

**sapma puanları**

$$z = \frac{X - \mu}{\sigma}$$

Bizim örneğimizde $\mu = 50$ ve $\sigma = 10$ olduğuna göre

$$z = \frac{X - \mu}{\sigma} = \frac{X - 50}{10}$$

Şekil 3.6'nın apsis ekseninin üçüncü satırındaki değerler bu dönüşümün etkisini göstermektedir. Sayısal değerlerin doğrusal dönüşümü dışında verinin başka bir biçimde değişmediğine dikkat ediniz. Dağılımın şekli hâlâ aynıdır ve puanların birbirine göre durumları değişmemiştir. Kişilerin haftalık alkol tüketimlerini ons yerine mililitre üzerinden ölçmek kişilerin sıralamasını değiştirmez. Sadece apsis eksenindeki sayıları değiştirir. (Mahallenin ayyaşı hâlâ mahallenin ayyaşıdır. Yalnızca içkisi mililitre birimiyle ölçülmüştür.) $X$ puanlarını $z$ puanlarına dönüştürmenin ne işe yaradığını anlamak önemlidir. Önceden 60 olan puan şimdi 1'dir. Yani önceden ortalamanın 1 standart sapma (10 puan) üzerinde olan puan hâlâ ortalamanın 1 standart sapma üzerindedir. Fakat artık bu puan 1'dir. Önceden 45 olan puan ortalamanın 0,5 standart sapma *altındaydı***.** Şimdi de ortalamanın 0,5 standart sapma altındadır ve –0,5'tir. Bir başka deyişle $z$ puanı, $X$'in ortalamanın kaç standart sapma üzerinde veya altında olduğunu gösterir. Pozitif $z$ **puanları** ortalamanın üzerinde, negatif $z$ puanları ise altındadır.

**z puanları**

*z* için verilen formül tamamen geneldir. Herhangi bir dağılımı bu formül ile *z puanlarına dönüştürebiliriz. Dağılımın şekli doğrusal bir dönüşümden etkilenmez. Dönüşümden önce normal olmayan bir dağılım, dönüşümden sonra da normal olmayacaktır.* Bazı insanlar puanları *z* puanlarına dönüştürmenin dağılımı normalleştireceğini (normal dağılıma dönüştüreceğini) söylerler. Bu kesinlikle işe yaramaz.

http://surfstat.anu.edu.au/surfstat-home/surfstat.html adresine girip "Hotlist for Java Applets"e tıklayarak normal dağılıma sahip bir evrenden örneklemler alındığında ne olduğunu görebilirsiniz. "Histogram"a bastığınızda değişik biçimlerde oynayabileceğiniz başka bir histogram açacaktır. Temizlemeden tekrar tekrar "start"a basarsanız örnekleme yeni vakalar ekleyebilirsiniz. Gözlem sayısı arttıkça dağılımın nasıl normale yaklaştığını görmek için kullanışlıdır. (Normal olmayan küçük bir örneklemin neye benzeyebileceğini görmek için de.)

## 3.3 Standart Normal Dağılım Tablosunun Kullanılması

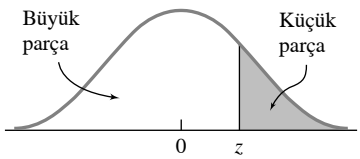
Daha önce de belirtidiği gibi, standart normal dağılım ayrıntılı bir şekilde tablolaştırılmıştır. Böyle bir tablo Ek *z*'de bulunmaktadır. Bunun bir kısmı Tablo 3.1'de verilmiştir.[1] Tabloyu nasıl kullanacağımızı görmek için Şekil 3.7'deki normal dağılımı göz önüne alalım. Bu Şekil 3.6'daki Davranış Problemi toplam puanlarının standartlaştırılmış halini de gösterebilir. Eğrinin altında kalan tüm alan 1,00 kabul edilirse, eğrinin altında kalan alanın ne kadarı ortalamanın en az bir standart sapma üzerindedir. (Alanları önemsiyoruz çünkü doğrudan olasılıklara aktarılabiliyorlar.) Biliyoruz ki z puanları zaten ortalamadan kaç standart sapma uzakta olduklarını gösteriyor. Dolayısıyla eğrinin altında kalan alanın ne kadarının z = 1 ve üzerinde olduğunu bilmek istiyoruz.

Sadece pozitif kısım tabloya alınmıştır. Dağılım simetrik olduğu için pozitif bir *z* ile ilgili bilgi negatif olan *z*'ye aktarılabilir. (Ek *z*'deki tabloda bir de "*y*" sütunu vardır. Bu sadece eğrinin o *z* noktasındaki yüksekliğini [yoğunluğunu] veriyor. Bunu buradaki kısma almadım çünkü hem yer dardı hem de zaten nadiren kullanılıyor.) Tablo 3.1 veya Ek *z*'den *z* = 1,00'a karşılık gelen satırı buluyoruz. Bu satırdan okuyabildiğimize göre *ortalamadan z* = 1*'e kadar* olan alan 0,3413, *büyük bölgenin* alanı 0,8413 ve *küçük bölgenin* alanı 0,1587 ediyor. *Alan* ve *olasılık* terimlerini eşitlediğimize göre evrenden tesadüfi olarak seçilecek bir çocuğun Davranış Problemi puanının ortalamanın en az 1 standart sapma üzerinde (60'ın üzeri) olması olasılığının, eğer puanlar normal dağılıyorsa, 0,1587 olduğunu şimdi söyleyebiliriz. Dağılım simetrik olduğu için bir öğrencinin en az 1 standart sapma kadar ortalamadan düşük puan alma olasılığının da 0,1587 olduğunu biliyoruz.

Şimdi öğrencinin ortalamadan her iki yönde de en az 1 standart sapma (10 puan) uzakta olması olasılığını bulmak istediğimizi varsayalım. Bu yalnızca alanların toplanması konusudur. Normal dağılım simetrik olduğu için z = –1'den küçük değerlere karşı gelen alan ile

[1] Elektronik tabloları tercih ediyorsanız İnternet'te birçok küçük Java programı mevcuttur. *z* olasılıklarını hesaplamakta kullanılanlar arasında benim favorim http://psych.colorado.edu/~mcclella/java/zcalc.html adresinde yer alıyor. Normal dağılımın özelliklerini gösteren çevrim içi bir video da http://huizen.dds.nl/~berrie/normal.html adresinden izlenebilir.

**Tablo 3.1** Normal dağılım (Ek z'nin kısaltılmış kısmı)

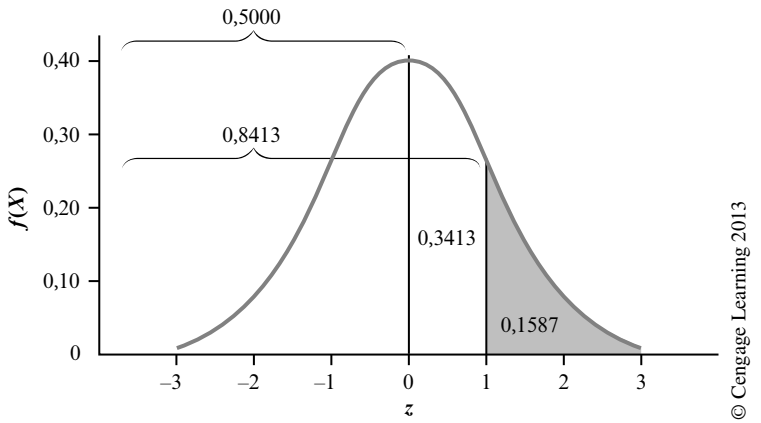



| z | Ortalamadan z'ye kadar | Büyük parça | Küçük parça | z | Ortalamadan z'ye kadar | Büyük parça | Küçük parça |
|---|---|---|---|---|---|---|---|
| 0,00 | 0,0000 | 0,5000 | 0,5000 | 0,45 | 0,1736 | 0,6736 | 0,3264 |
| 0,01 | 0,0040 | 0,5040 | 0,4960 | 0,46 | 0,1772 | 0,6772 | 0,3228 |
| 0,02 | 0,0080 | 0,5080 | 0,4920 | 0,47 | 0,1808 | 0,6808 | 0,3192 |
| 0,03 | 0,0120 | 0,5120 | 0,4880 | 0,48 | 0,1844 | 0,6844 | 0,3156 |
| 0,04 | 0,0160 | 0,5160 | 0,4840 | 0,49 | 0,1879 | 0,6879 | 0,3121 |
| 0,05 | 0,0199 | 0,5199 | 0,4801 | 0,50 | 0,1915 | 0,6915 | 0,3085 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 0,97 | 0,3340 | 0,8340 | 0,1660 | 1,42 | 0,4222 | 0,9222 | 0,0778 |
| 0,98 | 0,3365 | 0,8365 | 0,1635 | 1,43 | 0,4236 | 0,9236 | 0,0764 |
| 0,99 | 0,3389 | 0,8389 | 0,1611 | 1,44 | 0,4251 | 0,9251 | 0,0749 |
| 1,00 | 0,3413 | 0,8413 | 0,1587 | 1,45 | 0,4265 | 0,9265 | 0,0735 |
| 1,01 | 0,3438 | 0,8438 | 0,1562 | 1,46 | 0,4279 | 0,9279 | 0,0721 |
| 1,02 | 0,3461 | 0,8461 | 0,1539 | 1,47 | 0,4292 | 0,9292 | 0,0708 |
| 1,03 | 0,3485 | 0,8485 | 0,1515 | 1,48 | 0,4306 | 0,9306 | 0,0694 |
| 1,04 | 0,3508 | 0,8508 | 0,1492 | 1,49 | 0,4319 | 0,9319 | 0,0681 |
| 1,05 | 0,3531 | 0,8531 | 0,1469 | 1,50 | 0,4332 | 0,9332 | 0,0668 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 1,95 | 0,4744 | 0,9744 | 0,0256 | 2,40 | 0,4918 | 0,9918 | 0,0082 |
| 1,96 | 0,4750 | 0,9750 | 0,0250 | 2,41 | 0,4920 | 0,9920 | 0,0080 |
| 1,97 | 0,4756 | 0,9756 | 0,0244 | 2,42 | 0,4922 | 0,9922 | 0,0078 |
| 1,98 | 0,4761 | 0,9761 | 0,0239 | 2,43 | 0,4925 | 0,9925 | 0,0075 |
| 1,99 | 0,4767 | 0,9767 | 0,0233 | 2,44 | 0,4927 | 0,9927 | 0,0073 |
| 2,00 | 0,4772 | 0,9772 | 0,0228 | 2,45 | 0,4929 | 0,9929 | 0,0071 |
| 2,01 | 0,4778 | 0,9778 | 0,0222 | 2,46 | 0,4931 | 0,9931 | 0,0069 |
| 2,02 | 0,4783 | 0,9783 | 0,0217 | 2,47 | 0,4932 | 0,9932 | 0,0068 |
| 2,03 | 0,4788 | 0,9788 | 0,0212 | 2,48 | 0,4934 | 0,9934 | 0,0066 |
| 2,04 | 0,4793 | 0,9793 | 0,0207 | 2,49 | 0,4936 | 0,9936 | 0,0064 |
| 2,05 | 0,4798 | 0,9798 | 0,0202 | 2,50 | 0,4938 | 0,9938 | 0,0062 |

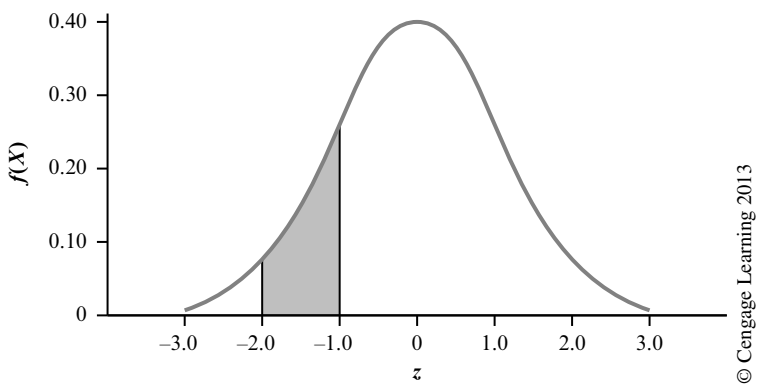
© Cengage Learning 2013

z = +1'den büyük değerlere karşı gelen alana eşittir. Tabloda negatif $z$'lerin bulunmama sebebi budur – gerekli değiller. Artık biliyoruz ki ilgilendiğimiz her iki alan da 0,1587 ediyor. Dolayısıyla $z = \pm 1$'in dışında kalan bölgenin alanı 0,1587 + 0,1587 = 0,3174 eder. Bunun tersi de doğrudur. Eğer $z = \pm 1$'in dışında kalan bölgenin alanı 0,3174 ise $z = -1$ ile $z = 1$ arasında kalan bölgenin alanı 1 – 0,3174 = 0,6826 eder. Dolayısıyla bir çocuğun puanının 40 ile 60 arasında olma olasılığı 0,6826'dır.

Bu işlemleri genişletmek için bir puanın 30 ile 40 arasında olması olasılığını ele alalım. Biraz aritmetik ile bu olasılığın, puanın ortalamanın 2 standart sapma altı ile 1 standart sapma altı arasında bulunma olasılığı olduğunu göreceğiz. Bu durum Şekil 3.8'deki grafikte




**Şekil 3.7** Normal dağılım altında kalan alan örnekleri




**Şekil 3.8** Ortalamanın 2 standart sapma altı ile 1 standart sapma altı arasında kalan alan

gösterilmiştir. (*İpucu*: Şekil 3.8'deki gibi bir grafik çizmek her zaman akıllıcadır. Hataları önler ve hangi bölgeleri aradığınızı netleştirmenize yardım eder.)

Ek $z$'den biliyoruz ki ortalamadan $z = -2$'ye kadar olan alan 0,4772, ortalamadan $z = -1$'e kadar olan alan ise 0,3413'tür. Bu iki alan arasındaki fark $z = -2$ ile $z = -1$ arasındaki alanı verecektir. Bu alan 0,4772 – 0,3413 = 0,1359'dur. Dolayısıyla, normal dağılan bir Davranış Problemleri puanları evreninden seçilerek bir puanın 30 ile 40 arasında olması olasılığı 0,1359'dur.

Son iki paragraftır yaptığımız normal dağılımın altında kalan alanların tartışılması, normal dağılımı açıklamanın geleneksel yoludur. Ancak haklı olarak Davranış Problemleri puanının 50 ile 60 arasında olması olasılığını neden bilmek isteyeceğimiz merak edilebilir. Bu sorunun basit cevabını belki önemsemeyebilirsiniz. Ama çocuğunuzu davranışlarından endişe ettiğiniz için bir değerlendirmeye getirdiniz. Varsayalım ki çocuğunuz 75 puan aldı. Birazcık aritmetik gösterecek ki $z = (75 - 50)/10 = 2,5$ ve Ek $z$'den göreceğiz ki normal çocukların yalnızca %0,62'si bu kadar yüksek puan alıyor. Yerinizde olsam endişelenmeye başlardım. 75 gerçekten yüksek bir puan.

## 3.4 Bir Gözleme Dair Olası Sınırların Belirlenmesi

Son bir örnek olarak, tesadüfi biçimde seçilecek bir öğrencinin puanının içinde kalacağını belirli bir güvenle söyleyebileceğimiz sınırları belirlemek istediğimiz bir durumu ele alalım. Yani şöyle bir cümle kurmak istiyoruz: "Bu evrenden tesadüfi olarak bir öğrenci seçilirse, denemelerin %95'inde öğrencinin puanı _____ ile _____ arasında olacaktır." Şekil 3.9'da aradığımız sınırlar görülmektedir - evrendeki puanların %95'ini kapsayan sınırlar.

Eğer puanların %95'inin içinde kalacağı sınırları arıyorsak aynı zamanda puanların %5'inin dışında kaldığı sınırları arıyoruzdur. Dışarıda kalan bu %5'lik bölümü ayırmak için dağılımın her iki ucundan –veya kuyruğundan– %2,5'luk bölümü kesen $z$ puanlarını bulmalıyız. (Simetrik sınırları kullanmak zorunda değiliz ancak genelde en mantıklı ve en kısa aralığı oluşturan sınırlar simetrik sınırlardır.) Ek $z$'den görüyoruz ki bu değerler $z = \pm 1.96$'dır. Dolayısıyla diyebiliriz ki denemelerin %95'inde öğrencinin puanı ortalamanın 1,96 standart sapma altı ile 1,96 standart sapma üstü arasında olacaktır.

Genellikle cevaplarımızı ham Davranış Problemi puanları cinsinden vermek istediğimiz için bir parça daha işimiz var. Ham puan sınırlarını elde etmek için z puanı formülünde $X$'i yalnız bırakırız. Böylece denklemi $z$ yerine $X$ için çözmüş oluruz. Dolayısıyla, eğer evrendeki puanların %95'ini içine alan sınırları bulmak istiyorsak ortalamanın 1,96 standart sapma altı ve üstündeki puanları bulmalıyız. Bu, şöyle yazılabilir:

$$z = \frac{X - \mu}{\sigma}$$

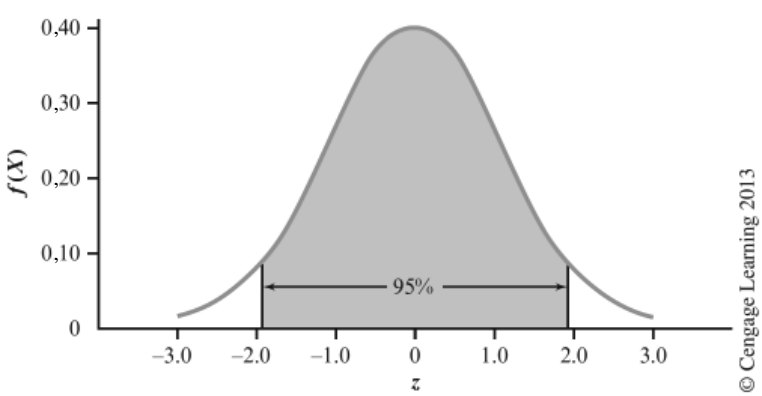
$$\pm 1,96 = \frac{X - \mu}{\sigma}$$

$$X - \mu = \pm 1,96\sigma$$

$$X = \mu \pm 1,96\sigma$$

Aradığımız sınırlar $(\mu + 1,96\sigma)$ ve $(\mu - 1,96\sigma)$'dır. Dolayısıyla

Sınırlar = 50 ± (1,96)(10) = 50 ± 19,6 = 30,4 ve 69,6'dır.




**Şekil 3.9** Davranış Problemi puanlarının %95'ini kapsayan sınırlar

Dolayısıyla, tesadüfi olarak seçilen bir öğrencinin puanının 30,4 ile 69,6 arasında olması olasılığı 0,95'tir. Düşük puanlar bir soruna işaret etmedikleri için onlarla pek ilgilenmeyebiliriz. Ama 69,6 veya daha üzerinde bir puan alan bir kişi için problemdir. Çocukların yalnızca %2,5'i bu kadar yüksek puan alır.

kestirim aralığı

Burada tartıştığımız konu, daha sonra ele alacağımız güven aralıkları konusu ile –aynı olmasa da– yakından ilgilidir. Temel fark şu ki burada evren ortalamasını biliyoruz ve tek bir gözlemin ($X$) nereye düşebileceğini tahmin etmeye çalışıyoruz. Daha sonra buna benzer bir şeyi **kestirim aralığı** olarak adlandıracağız. Güven aralıklarını tartıştığımızda elimizde bir örneklem ortalaması (veya başka bir istatistik) olacak ve evren ortalamasını (veya ilgili parametreyi) %95 olasılıkla içerecek sınırları kestireceğiz. Şu noktada güven aralıklarıyla ilgili hiçbir şey bilmenize gerek yok. Yalnızca, daha sonra ortaya çıkması muhtemel bir karışıklığa önceden engel olmak için değinilmiştir.

# 3.5 Veri Kümesinin Dağılımının Normal Olup Olmadığının Değerlendirilmesi

Bu kitabın pek çok yerinde verinin normal dağıldığını varsayacağız. Ancak bir dağılıma bakıp bu varsayımın ne ölçüde makul olduğunu belirlemek zordur. İstatistik kitapları normal gibi görünen ama normal olmayan dağılım örnekleriyle doludur. Bu örneklerin peşinden, veri normal olmadığı için, bir teknikle elde edilen sonuçların ne kadar bozulmuş olduğuna dair ifadelere yer verilir. Daha önce de dediğimiz gibi, histogramın üzerine gerçek bir normal dağılım eğrisi çizilebilir. Bu bize uyumun ne kadar iyi olduğuna dair belli ölçüde bir fikir sağlayabilir ama bu çoğu zaman yanıltıcı bir yaklaşımdır. Çok daha iyi bir yaklaşım **yüzdelik-yüzdelik (Q-Q) grafikleri**dir.

yüzdelik-yüzdelik grafikleri (Q-Q)

## Q-Q Grafikleri

Yüzdelik-yüzdelik grafikleri ardındaki düşünce, temelinde çok basittir. Varsayalım ortalaması "0", standart sapması "1" olan 100 gözlemden oluşan, tam olarak normal bir örneklemimiz var. (Ortalama ve standart sapma herhangi sayılar olabilir. "0" ve "1" yalnızca tartışmayı daha basit kılıyor.) Böyle bir dağılımda yüzdelikleri, örneğin en alt %1'lik kısma karşılık gelen puanı kolayca hesaplayabiliriz. Ek $z$'den bu değer –2,33 olarak görülür. Ayrıca –2,054 puan en alt %2'yi kesmektedir. Bu hesaplamayı $0{,}00 < p < 1{,}00$ aralığındaki tüm değerler için tekrarlayabilir ve sonuçlara normal dağılıma ait *beklenen yüzdelikler* adını verebiliriz.

Şimdi kendi verimizde gidelim. Tam olarak normal dağılıma sahip $n = 100$ gözlemimiz olduğu için en alt %1'lik bölüme karşılık gelen puan –2,33 ve benzer şekilde dağılımın en alt %2'lik kısmını kesen puan ise –2,054 olacaktır. Bunlara *elde edilen yüzdelikler* adını verebiliriz, çünkü bunlar doğrudan veriden hesaplandılar. Tam olarak normal olan bir dağılımda bu iki yüzdelik puanlar kümesi tamamen aynı olacaktır. Elde edilen 15'inci yüzdelik değeri, normal dağılımın vereceği değer ile aynıdır.

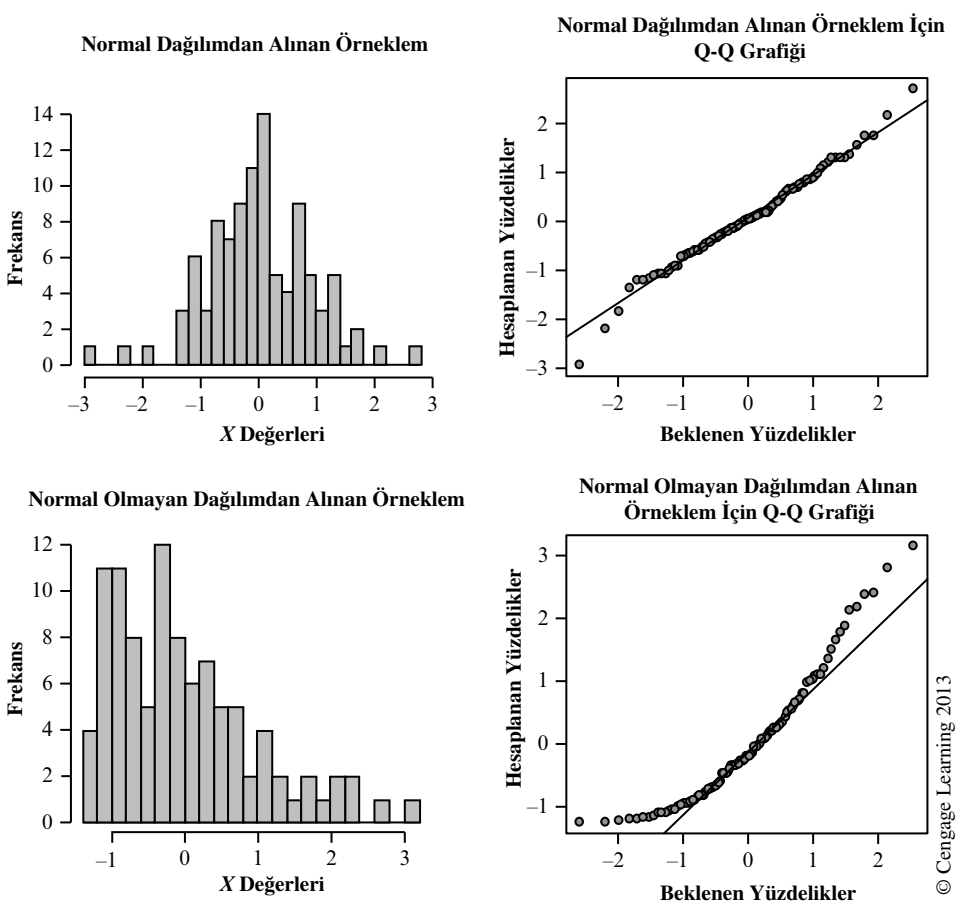
Fakat varsayalım ki örneklemimizin dağılımı normal olmasın. Bu durumda en alt %1'lik kısma karşılık gelen puanı –2,33 yerine –2,8 bulabiliriz. Aynısı diğer yüzdelikler için de olabilir. Burada beklenen yüzdelikler ile verimizden elde edilen yüzdelikler birbirini tutmayacak, uyumsuz olacaktır.

Peki, uyumu nasıl ölçeriz? En kolay yolu karşılıklı grafiğe dökmektir –beklenen yüzdelikleri $Y$, elde edilen yüzdelikleri de $X$ ekseninde gösterelim. Eğer dağılım normal ise grafik 45 derecelik düz bir çizgi oluşturacaktır. Bu grafiklere iki örnek Şekil 3.10'da verilmiştir. Birisi normal dağılımdan diğeri ise kesinlikle normal olmayan bir dağılımdan gelmektedir.

Şekil 3.10'da normal dağılıma sahip bir veri kümesi için olan Q-Q grafiğinde noktaların çoğu doğrunun üzerine iyi bir şekilde düşmektedir. Uçlarda çizgiden biraz uzaklaşmaktadır ancak bu çok büyük örneklemleriniz olmadığı sürece yaygındır. Normal olmayan veride noktalar düz çizgiden belirgin ölçüde sapar. Alt bölümlerde yüzdeliklerin –1 olması beklenen yerde verideki değerin –2 olduğu görülüyor. Yani dağılım sol tarafından kırpılmış. Q-Q grafiğinin sağ üst köşesinde beklenen yüzdelik en az 3 iken elde edilen yüzdelik 2'dir. Bir başka deyişle veri alt kısımda ortalamadan yeterince uzaklaşmamış, üst kısımda ise gereğinden fazla uzaklaşmıştır. R dilinde yazılmış bir Q-Q grafiği programı kitabın Web sayfasında mevcuttur.

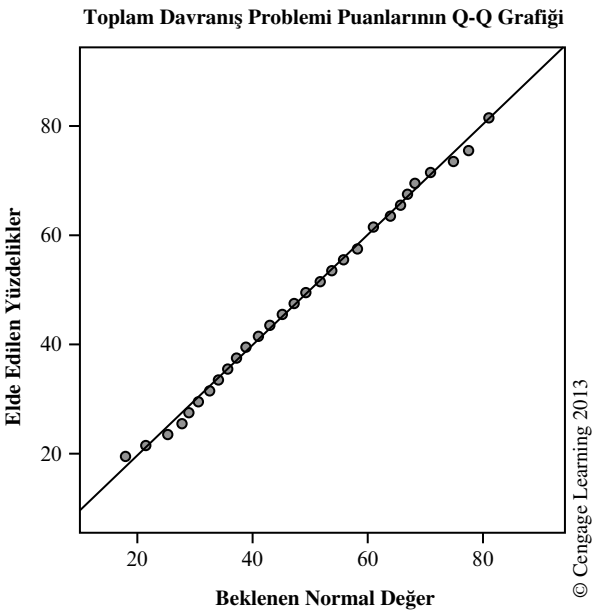
Achenbach'ın Toplam Davranış Problemi puanlarına baktık ve bu puanların normal olduklarını ifade ettik. Şekil 3.11 bu puanların Q-Q grafiğini göstermektedir. Grafikten puanların normal dağıldığı görülmektedir. Bunun sebeplerinden biri Achenbach'ın ölçeğe istenen özellikleri verebilmek için hayli çabalamış olmasıdır.

Ancak kitabın baş kısmındaki tepki süresi verisine bakalım. Şekil 2.4'ü tartışırken verinin soldan kırpılmış olsa da kabaca normal olduğunu söylemiştik. Şekil de aşağı yukarı

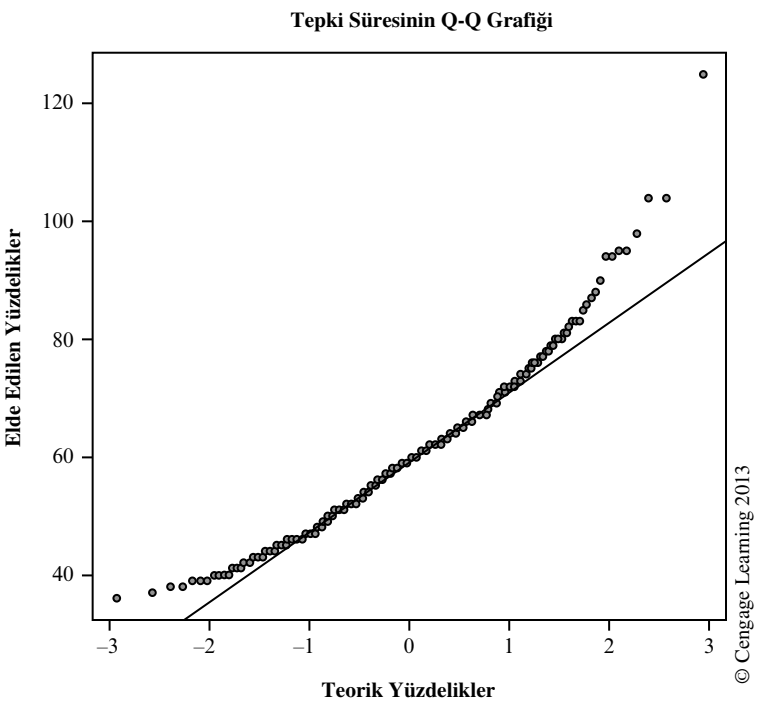



**Şekil 3.10** Normal ve normal olmayan veri kümeleri için histogram ve Q-Q grafikleri

bunu göstermekteydi. Fakat tepki süresi verisine ait, Şekil 3.12'de verilen Q-Q grafiği düz bir çizgi olmaktan uzaktır. Bu, üzerinde normal dağılım eğrisi eklenmiş olsun veya olmasın, bir histograma bakarak dağılımın normal olup olmadığını anlamanın tahmin ettiğinizden daha zor olduğu açıklanmak için yapılmıştır.




**Şekil 3.11** Toplam Davranış Problemi puanlarına ait Q-Q grafiği




**Şekil 3.12** Tepki süresi verisine ait Q-Q grafiği

## Q-Q Grafiğinin Eksenleri

Q-Q grafiğinin temelindeki mantığı açıklarken, değişkenlerden standartlaştırılmışlar gibi söz edildi. (Gerçi bunun şart olmadığı da belirtildi.) Bu, normal dağılım tablosuna gitmekten daha kolay olması nedeniyle yapılmıştır. Fakat bir veya iki eksendeki birimlerin farklı olduğu Q-Q grafikleri ile sıkça karşılaşılacaktır. Bu bir sorun değildir. Önemli olan eksenlerin birimleri değil noktaların doğruya göre olan konumlarıdır. Aslında farklı istatistiksel paket programlar hem farklı birimler kullanır hem de yatay ve dikey ekseni farklı yerleştirirler. Burada verilen grafiklerin (dikeyde) ayna görüntüsü biçimimde bir grafiği görülürse bilinsin ki yalnızca *X* eksenine beklenen değerler değil elde edilen değerler konulmuştur.

## Kolmogorov-Smirnov Testi

**Kolmogorov-Smirnov testi**

En çok bilinen normallik testi SPSS'te "Nonparametric Tests" menüsü altında yer alan **Kolmogorov-Smirnov testi**dir. Bu testin var olduğunun bilinmesi gerekse de *çoğu kişi bu testin kullanılmasını önermez.* Öncelikle küçük örneklemlerin çoğu kesinlikle normal olmamalarına rağmen testi geçerler. Öte yandan, büyük örneklemlerde normallikten sorun teşkil etmeyecek kadar küçük sapmalar olsa da testin normallik hipotezini reddetmesi olasılığı yüksektir. D'Agustino ve Stephens (1986) bu durumu daha güçlü bir biçimde ifade etmişlerdir: "Kolmogorov-Smirnov testi yalnızca tarihî bir değerdir. Hiçbir zaman kullanılmamalıdır." Burada yalnızca başka yerlerde karşılaşılacağı ve SPSS'te görüleceği için değinilmiştir. Testin zayıflıkları bilinmelidir.

# 3.6 z ile İlgili Ölçüler

**standart puanlar**

Daha önce gördüğümüz gibi *z* formülü bir dağılımı, ortalaması "0" standart sapması (ve varyansı) "1" olacak şekilde dönüştürmekte kullanılabilir. Bu dönüştürülmüş puanlara sıklıkla **standart puanlar** denir. Puanların dönüştürülmesine dayalı başka puan sistemleri de vardır. İnsanlar bu sistemleri farkında olmadan her gün kullanır.

**yüzdelik**

Bu puanlama sistemine güzel bir örnek yaygın olarak kullanılan IQ puanlarıdır. IQ testlerinden alınan ham puanlar rutin olarak ortalaması 100, standart sapması 15 (Binet Testi için 16) olan puanlara dönüştürülür. Bu bilgiye dayanarak bir kimsenin IQ puanından (örneğin 120) ortalamanın kaç standart sapma üzerinde veya gerisinde olduğunu anlayabiliriz (yani *z* puanını hesaplayabiliriz). IQ puanları aşağı yukarı normal dağıldıkları için bu *z* puanlarını da Ek *z*'yi kullanarak yüzdeye çevirebilirsiniz. (Bu örnekte, puanların %91'i 120'nin altında kalmaktadır. Bu 91'inci **yüzdelik** olarak bilinir.)

Standart tanı testleri bir başka yaygın örnektir. Bu testlerin puanları da belirlenmiş bir ortalamaya ve standart sapmaya sahip dağılımlara dönüştürülür. Achenbach bunun bir örneğidir. Ham puanlar testin üreticisi tarafından ortalaması 50 standart sapması 10 olan bir dağılıma dönüştürülmüştür. Böyle bir puanlama sisteminin ayarlanması kolaydır. (Ham puan ortalaması ve standart sapmasını kullanarak) Ham puanları *z* puanlarına dönüştürerek başlarız. Daha sonra bu *z* puanlarını da aklımızdaki yeni puan sistemine çeviririz. *z*, kişinin ham puanına karşılık gelen z puanı olmak üzere

Yeni puan = Yeni standart sapma * (*z*) + Yeni ortalama

**T puanları**

Achenbach'ın Gençler İçin Kendini Değerlendirme Ölçeği'nin kullanıldığı, ortalaması 50, standart sapması 10 olan puanlama sistemlerine **T puanları** denir (T her zaman büyük

harfle yazılır). Bu testler psikolojik ölçmelerde kullanışlıdır çünkü ortak bir referans çerçeveleri vardır. Örneğin, insanlar en üst %10'luk bölge için 63 puanı kesme noktası olarak almaya alışmıştır.

## Anahtar Terimler

Normal dağılım (Giriş)
Çubuk grafiği (Giriş)
Apsis (3.1)
Ordinat (3.1)
Standart normal dağılım (3.2)
Odak değer (3.2)
Sapma puanı (3.2)
z puanı (3.2)
Tahmin aralığı (3.4)
Yüzdelik-yüzdelik (Q-Q) grafikleri (3.5)
Kolmogorov-Smirnov testi (3.5)
Standart puanlar (3.6)
Yüzdelik (3.6)
T puanları (3.6)

## Alıştırmalar

3.1 Aşağıdaki verinin $\mu = 4$ ve $\sigma = 1{,}63$ olan bir evren olduğunu kabul ediniz.

$X = [1 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 4 \quad 4 \quad 4 \quad 4 \quad 5 \quad 5 \quad 5 \quad 6 \quad 7]$

a. Verilen dağılımın grafiğini çiziniz.

b. (a)'da verilen dağılımı $X - \mu$ dağılımına dönüştürünüz.

c. Bir sonraki adıma geçip (b)'de verilen dağılımı $z$ dağılımına dönüştürünüz.

3.2 3.1'deki dağılımı kullanarak $X = 2{,}5$; 6,2 ve 9 puanlarına ait $z$ puanlarını hesaplayınız. Sonuçları yorumlayınız.

3.3 Basit bir görevdeki hataları incelemek istediğimizi varsayınız. Bir grup hakemden bir sabah bir mağazaya giren insanları saymalarını istiyoruz. Bazı hakemler bazı kişileri eksik sayacak, bazıları ise bazı kişileri iki kere sayacaktır. Bu nedenle her hakemin aynı sayıyı vermesini beklemiyoruz. Rapor edilen kişi sayılarının ortalamasının 950, standart sapmasının 15 olduğunu kabul ediniz. Rapor edilen kişi sayılarının dağılımının normal olduğunu varsayınız.

a. Rapor edilen kişi sayılarının yüzde kaçı 960 ile 990 arasında olacaktır?

b. Rapor edilen kişi sayılarından yüzde kaçı 975'in altında olacaktır?

c. Rapor edilen kişi sayılarından yüzde kaçı 990'ın altında olacaktır?

3.4 3.3'teki örneği kullanarak;

a. Sonuçların ortadaki %50'lik kısmını kapsayan bölgenin $X$ değerlerini (rapor edilen kişi sayıları) bulunuz.

b. Rapor edilen kişi sayılarının %75'i _____ değerinden az olacaktır.

c. Rapor edilen kişi sayılarının %95'i _____ ile _____ arasında olacaktır.

3.5 Alıştırma 3.3'teki projenin başındaki kişi mağazaya sadece 950 kişinin geldiğini saymıştır. Özenle saydığı düşünülürse bu makul bir sayı mıdır? Neden?

3.6 Bir grup dördüncü sınıf öğrencisinin okuma puanlarının ortalaması 25, standart sapması 5'tir. Bir grup dokuzuncu sınıf öğrencisinin okuma puanları ortalaması ve standart sapması ise sırasıyla 30 ve 10'dur. Dağılımları normal kabul ediniz.

a. İki grubu da aynı şekil üzerinde yerleştirerek kabaca grafiğini çiziniz.

b. Dördüncü sınıfların yüzde kaçı ortalama bir dokuzuncu sınıf öğrencisinden daha iyi okuma puanına sahiptir?

c. Dokuzuncu sınıfların yüzde kaçı ortalama bir dördüncü sınıf öğrencisinden daha kötü okuma puanına sahiptir? (Bu düşünceye Bölüm 8'de güç konusunu işlerken geri döneceğiz.)

3.7 Hangi koşullar altında Alıştırma 3.6'nın (a) ve (b) şıklarının cevapları aynı olurdu?

3.8 Bir tanı testi, çocuk yalnızca en alt %10'lük dilime girerse (10'uncu yüzdelik) bir soruna işaret etmektedir. Ortalama 150, standart sapma 30 olduğuna göre tanısal olarak anlamlı olan kesme puanı kaçtır?

3.9 Bir dekan, fakültesindeki öğretim üyelerinin bir sonraki yıla ait maaş artışlarını saptayacaktır. Ortalama maaşın 2000 $ ve standart sapmanın 400 $'a yükseltilmesine ve dağılımının normal olmasına karar vermiştir.

a. Öğretim üyelerinin en verimli %10'u en az _____ $ artış alacaktır.

b. Öğretim üyelerinin en verimsiz %5'i en çok _____ $ artış alacaktır.

3.10 Bir dersteki tüm öğrencilere insanların ne kadarının emniyet kemeri taktıklarını inceleme ödevi verdik. Her bir öğrenciden 100 otomobile bakması ve kaçında emniyet kemeri takıldığını rapor etmesi istendi. Öğrencinin rapor ettiği sayı öğrencinin puanı olarak ele alındı. Sınıftaki puanların ortalaması 44, standart sapması 7'dir.

a. Puanların normal dağıldığını varsayarak grafiğini çiziniz.

b. Bütün yıl boyunca görevlerini pek yerine getirmeyen bir öğrenci 100 arabadan 62'sinde emniyet kemeri takıldığını rapor etmiştir. Öğrencinin ödevi gerçekten yapmak yerine rastgele bir sayı attığını düşünmek için yeterli sebebimiz var mı?

3.11 Bir süre önce bir arkadaşım dil problemlerinin teşhisi için bir test geliştirdi. Testi uygulayan kişi, çocuğa bir uyaran veriyor ve çocuktan bu uyarana dair çoğul, olumsuz, edilgen gibi dil yapılarını üretmesi bekleniyor. Ölçeğinden elde edilen puanlar çocuğun kaç tanesini doğru yaptığını gösteriyor. Testin ortalaması 48, standart sapması 7'dir. Veliler testin puanlarını anlamakta güçlük çekince arkadaşım puanları (veliler daha kolay anlayabilsin diye) ortalaması 80, standart sapması 10 olan puanlara dönüştürmek istedi. Bu işi nasıl yapabilir?

3.12 Maalesef bütün dünya normal dağılım prensibi üzerine kurulu değil. Bir önceki örnekte gerçek dağılım fena halde çarpık çünkü çocukların çoğunda dil problemi yok ve yapıları hatasız bir şekilde oluşturabiliyorlar.

a. Dağılımın olası şeklini çiziniz.

b. Normal olmayan bir dağılımda, alttan %10'luk bölümü işaretlemekte kullanılacak kesme puanını bulmakta nasıl bir yol izlersiniz?

3.13 1981 Ekim ayında Graduate Record Exam (GRE) testine giren tüm adayların puan ortalaması 489, standart sapması ise 126'ydı. Sınava katılanların yüzde kaçının 600 veya daha az almış olması beklenir? (Buna 600'ün yüzdelik sırası denir.)

3.14 Alıştırma 3.13'te, puanların %75'ine eşit veya onlardan daha yüksek olan puan kaçtır? (Buna 75'inci yüzdelik denir.)

3.15 1981 Ekim ayında GRE'ye giren son sınıf veya mezun öğrencilerin ortalaması 507, standart sapması 118'dir. Bu, Alıştırma 3.13 ve 3.14'teki soruların cevaplarını nasıl değiştirir?

3.16 Alıştırma 3.15, referans grubun önemi hakkında ne ifade etmektedir?

3.17 Ek'teki veri kümesindeki GPA değişkeni için 75. yüzdelik kaç puandır? (Gözlemlerin %75'inin bu noktanın altında olması beklenir.)

3.18 Davranış Problemi puanlarının ortalamasının 50, standart sapmasının 10 olduğunu varsayınız. En üst %2'de yer alan çocukları belirlemek için kullanılacak kesme puanı kaçtır?

3.19 Kitabın 3.6 numaralı kesiminde T puanlarının ortalaması 50, standart sapması 10 olacak şekilde tasarlandığını ve Achenbach'ın Gençler İçin Kendini Değerlendirme Ölçeği'nin T puanları ürettiğini söyledik. Şekil 3.3'teki verinin ortalaması tam 50, standart sapması tam 10 değil. Sizce neden?

3.20 SPSS, OpenStat veya R gibi standart bir bilgisayar programı kullanarak normal dağılan bir değişkenden 20'şer gözlemden oluşan 5'er örneklem üretiniz. (SPSS için ilk örneklemin oluşturulması amacıyla kullanılacak kodlar şöyledir:
**COMPUTE norm1 = RV.NORMAL(0,1)**. R için olanı ise **norm1<– rnorm(20, 0, 1)** biçimindedir.

a. Her birisi için Q-Q grafiklerini oluşturunuz ve grafiklerdeki farkları gözleyiniz. Bu size tam olarak normal bir dağılıma sahip bir değişkenden alınan örneklemelerin 45 derecelik düz çizgiye ne kadar yaklaşacağı hakkında fikir verecektir. Farkları nasıl nitelersiniz?

b. Bu alıştırmayı $n = 50$ alarak tekrarlayınız.

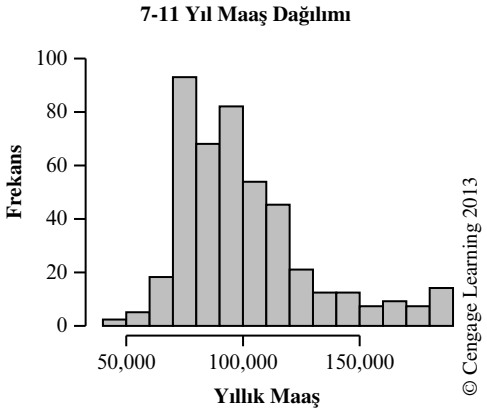
3.21 Bölüm 2, Şekil 2.16'da, Everitt'in iştahsızlık tedavisi örneğindeki üç bağımlı değişken için histogramlar oluşturulmuştu. Bu veriye www.uvm.edu/~dhowell/methods8/DataFiles/Fig2-16.dat adresinden ulaşılabilir. (Değişken adları dosyanın ilk satırında yer almaktadır.) Şekil 3.6'daki her grafik için ilgili Q-Q grafiğini çiziniz. Şekillere bakarak ulaşacağınız sonuçlar ile Q-Q grafiklerine bakarak ulaşacağınız sonuçlar birbirini tutuyor mu? (Not: Bu üç dağılımın hepsi Kolmogorov-Smirnov normallik testi sonucunda normal bulunurdu. Gerçi hiçbir normallik testi küçük örneklemlerde pek iyi değildir.)

## Tartışma Soruları

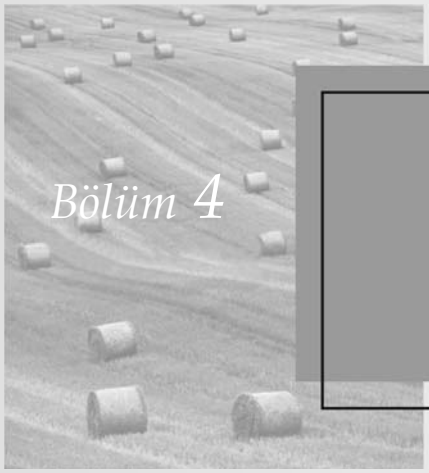
3.22 Tablo 2.2 ve Şekil 2.1'de frekans dağılımları sunulan tepki süresi verisine bakılırsa verinin tamamının normal dağılmadığı görülür. Bu veri için ortalama 60,26, standart sapma 13,01'dir. Basit hesaplamalarla verinin yüzde kaçının ortalamanın ±1,0; ±1,5; ±2,0; ±2,5 ve ±3,0 standart sapma altında ve üstünde olduğu hesaplanabilir. Eğer veri normal dağılsaydı puanların yüzde kaçının bu sınırların altında ve üstünde olacağı da normal dağılım tablosundan elde edilebilir. Bu değerleri hesaplayıp karşılıklı olarak bir grafiğe dökünüz. (Böylece kısmi bir Q-Q grafiği oluşturmuş olacaksınız.) Bu çizdiğiniz kısmi Q-Q grafiğini veya tam bir Q-Q grafiğini kullanarak verinin normalden ne ölçüde saptığını değerlendiriniz. Cevabınız örneklemin çok küçük ya da çok büyük olması durumunda nasıl değişirdi?

3.23 Yakın zamanda bana sorulan bir soruya cevap vermek için 16 puandan oluşan ortalaması 16,3, standart sapması 4,25 olan ve aşağı yukarı normal dağılan bir dağılım oluşturdum. Alıştırma 3.20'deki yaklaşım, bu değerlere yakın ortalamaya ve standart sapmaya sahip bir dağılım oluşturmak için kullanılabilir, ama ben değerlerin tam olarak aynı olmasını istiyorum. Böyle bir veriyi nasıl oluşturabilirim?

3.24 Aşağıda grafiği sunulan veri 2008–2009 yıllarında psikoloji profesörlerine ödenen yıllık maaşları göstermektedir. Veri Web sayfasında Ex3-24.dat dosyasında bulunabilir. Ayrıca dosyada 24 yıllık veya daha fazla tecrübeye sahip profesörlerin maaşları da yer almaktadır. Verinin sağa çarpık olduğu aşikar olsa da bu veriye normal gibi muamele ettiğinizde ne olmasını beklersiniz? Maaş değerleri yerine bu değerlerin logaritmalarını –log(Yıllık Maaş)– alırsanız ne olur? (Bazı $X$ değerlerin logaritmalarını hesap makinanızla alınız ve nasıl değiştiğini gözleyiniz.) Uç değerlerin sebebine nasıl bir açıklama getirirsiniz?



3.25 SAT.dat dosyasında 50 eyaletin SAT puanları, eğitime harcanan para, SAT sınavına giren öğrenci sayıları yer almaktadır. (Veri, Ek Veri kümelerinde açıklanmıştır.) Toplam SAT puanlarının histogramını çiziniz. Bu dağılım normal midir? Veri kümesinde yer alan adjcomb değişkeni toplam SAT puanlarının sınava giren öğrenci oranları üzerinden ayarlanmış bir halini içermektedir. Bu değişken hakkında ne diyebilirsiniz? Ayarlanmış SAT puanları ile ayarlanmamış toplam SAT puanları dağılımları nasıl farklılıklar göstermektedir?



Bölüm 4

# Örneklem Dağılımları ve Hipotez Testi

## Hedefler

Hipotez test etmenin genel teorisini inceleyip belirli kavramları hipotez test etmedeki anlamlarıyla açıklama ve bu kitapta tartışılan yöntemlerin temellerini sunma.

## Konular


4.1 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dair İki Basit Örnek
4.2 Örneklem Dağılımları
4.3 Hipotez Test Etmenin Teorisi
4.4 Sıfır Hipotezi
4.5 Test İstatistikleri ve Bunların Örneklem Dağılımları
4.6 Sıfır Hipotezi Hakkında Karar Verme
4.7 I. ve II. Tip Hatalar
4.8 Tek ve Çift Yönlü Testler
4.9 Sıfır Hipotezini Reddetmek Ne Anlama Gelir?
4.10 Hipotez Test Etmeye Alternatif bir Bakış
4.11 Etki Büyüklüğü
4.12 Son Bir İşlenmiş Örnek
4.13 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dönüş

İKİNCİ BÖLÜM'de bazı istatistikleri inceleyip bunların bir olayın kaç kere görüldüğünün gösterilmesi için nasıl kullanılabileceklerini gördük. Verinin betimlenmesi, her analiz için temel ve önemli olsa da karşımıza çıkan ilginç problemlerin çoğunu cevaplamaya yetmez. Tipik bir deneyde bir grup kişiye belirli bir işlem uygulayıp bu kişilerin puanlarının genel olarak insanların puanlarından farklılık gösterip göstermediğine bakabiliriz. Veya bir gruba bir deneysel girişimi uygulayıp tedavi uygulanmayan kontrol grubuyla puan ortalamalarını kıyaslamak isteyebiliriz. Betimsel istatistikler, örneğin, örneklem ortalaması ile kuramsal bir evren ortalaması arasındaki farkın veya iki örneklemin ortalamaları arasındaki farkın sadece şans ile açıklanabilecek kadar küçük bir fark olup olmadığını veya deneysel girişimimizin etkisine atfedilebilir gerçek bir fark olup olmadığını söyleyemez.

İstatistikçiler sıklıkla "şans değişkenliği" veya "örnekleme hatası" diye terimler kullanır ve ne olduğunu bildiğinizi varsayarlar. Belki de biliyorsunuzdur. Ancak bilmiyorsanız, eğer bunu açıklamak için bir dakika ayırmazsak, kitabın geri kalanında zihnimiz karışacaktır. Basit bir örnekle başlayacağız.

Bölüm 3'te Achenbach'ın Gençler İçin Kendini Değerlendirme Ölçeği davranış problemleri toplam puanlarını ele almıştık. Toplam davranış problemi puanları evrende normal dağılmaktadır (diğer bir deyişle, bu puanların tüm evrenininde yaklaşık olarak normal dağılmaktadır). Evren ortalaması ($\mu$) 50, evren standart sapması ise ($\sigma$) 10'dur. Biliyoruz ki farklı çocuklar farklı davranış problemleri gösterir. Dolayısıyla farklı puanlar alırlar. Yine biliyoruz ki 50 çocuktan oluşan bir örneklem alsak, bu örneklemin ortalaması muhtemelen tam 50 olmayacaktır. Bir örneklemdeki çocukların ortalaması 49 iken başka bir örneklemin ortalaması 52,3 olabilir. Örneklem ortalamaları, o örnekleme hangi çocukların denk geldiğine bağlı olacaktır. Örneklemden örnekleme görülen bu beklendik değişkenlik, "şansa bağlı değişkenlik" veya "hata varyansı" dediğimizde kast ettiğimizdir. Bu ifade, örneklemden elde edilen istatistiklerin (bu örnekte ortalama) örneklemden örnekleme doğal olarak değişkenlik gösterdiği gerçeğine işaret eder.

örnekleme hatası

Aynı şekilde, **örnekleme hatası** terimi bu bağlamda şansa bağlı değişkenlik ile eş anlamlı olarak kullanılır. Bu terim, bir örneklem istatistiğinin sayısal değerinin, örnekleme dâhil olan gözlemlere bağlı olarak, kestirdiği evren parametresinden muhtemelen farklı (dolayısıyla hatalı) olacağına gönderme yapar. Bu bağlamda "hata", dikkatsizlik veya yanlış yapmak anlamında değildir. Davranış problemleri örneğinde, bir örneklemde aşırı yaramaz bir çocuk varken bir diğer örneklemde çok sayıda nispeten yaramaz olmayan çocuk bulunabilir.

## 4.1 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dair İki Basit Örnek

Korelasyon ve regresyonu inceleyeceğimiz zaman ele alacağımız örneklerden biri öğrencilerin ders değerlendirme sonuçlarıyla o dersten bekledikleri not arasındaki ilişkiye bakmaktadır. Pek çok öğretim üyesi bu konuyu hayli önemser çünkü yarıyıl ders değerlendirme sonuçlarını –tıpkı bir öğrencinin ders notlarını açtığı gibi– kaygıyla açarlar. Bazı öğretim üyeleri, ders değerlendirme sonuçlarının, öğrencilerin o dersten kaç beklediklerinden bağımsız bir şekilde iyi ya da kötü olduğunu düşünür. Bazıları ise derse pek gelmeyen ve muhtemelen başarısız olacak öğrencilerin ders değerlendirme yaparken haksız yere derse

düşük puan verdiklerini düşünür. Son olarak, dersten iyi bir not dışında bir şeyler de alabilmiş olan başarılı öğrencilerin ders değerlendirmesinde yüksek puan verebileceğini iddia edenler vardır. Ancak, ders değerlendirme notları ile öğrenci performansı arasındaki ilişki deneysel bir sorudur ve böyle bir soru, ilgili veriye bakarak cevaplanabilir. Diyelim ki 50 farklı dersten oluşan tesadüfî bir örneklemde yüksek not bekleyen öğrencilerin ders değerlendirmelerinde yüksek, düşük not bekleyenlerin ise düşük not söyleme eğilimi olduğunu gördük. Bu küçük veri kümemizdeki bulgunun, öğrencilerin genelinde var olan bir eğilime mi işaret ettiğini, yoksa çalışmayı yeniden yaptığımızda kaybolacak, eldeki veri kümesine has bir gariplik mi olduğuna nasıl karar veririz? (İlginize göre tahmininizi, ne tür bir sonuç bulacağımızı kestiriniz. Bu konuya tekrar döneceğiz.)

İkinci örnek Doob ve Gross (1968) tarafından yapılan ve algılanan sosyal statünün etkisini inceleyen bir çalışmadan gelmektedir. Trafik lâmbalarında bekleyen, eski (düşük statülü) bir otomobil yeşil yandığında çalışmazsa, vakaların %84'ünde arkadaki arabanın korna çaldığını bulmuşlardır. Ancak çalışmayan araba pahalı, yüksek statülü bir otomobil olduğunda arkadaki sürücü vakaların yalnızca %50'sinde korna çalmıştır. Bu sonuçlar iki farklı şekilde açıklanabilir:

- Bir örneklemdeki %84 ile diğerindeki %50 arasındaki fark örnekleme hatasına (örneklemler arasındaki tesadüfi değişkenlik) atfedilebilir. Dolayısıyla, algılanan sosyal statünün korna çalma davranışını etkilediği sonucuna ulaşamayız.
- %84 ile %50 arasındaki fark büyük ve güvenilirdir. Fark örnekleme hatasıyla açıklanamaz. Dolayısıyla, insanların yüksek statüdeki otomobillere daha az korna çalma eğiliminde oldukları sonucuna ulaşabiliriz.

Bu soruya verilecek cevap için yapılacak istatistiksel işlemler bu örnek ile ders değerlendirme örneği arasında farklılık gösterecektir (çünkü birisi korelasyonlarla diğeri ise oranlarla ilgili). Yine de altında yatan mantık temelde aynıdır.

hipotez test etme

Ders değerlendirmeleri ve korna çalma örnekleri hipotez test etme başlığı altına düşmektedir. Bu bölümün amacı, **hipotez test etme**nin teorisini en genel haliyle, testlerin kendilerine has özelliklerine girmeden sunmaktır. Korelasyonları değil, farkları içeren durumlara odaklanacağım ama mantık temelde aynı. (Bölüm 9'da korelasyonlarla ilgili daha fazla içerik bulacaksınız.) Hesaplamaların detaylarını kasten irdelemiyorum çünkü amacım, en basit teknik detaylar dışında hiçbir konuya girmeden hipotez test etmenin kavramlarını incelemektir.

Problemin ne olduğu konusunda açık olmalıyız. Hipotez test etmenin sebebi verinin belirsiz olmasıdır. Kalabalık sınıfların düşük puanlar alıp almadığına karar vermek istediğimizi varsayalım. Hepimiz biliriz ki bazı kalabalık sınıflar çok iyidir, bazıları ise kötüdür. Benzer şekilde az öğrencinin bulunduğu iyi ve kötü sınıflar da vardır. Bu durumda, kalabalık derslerden veri toplarsak, bu sınıflara ait ortalamalar hangi sınıfların örneklemde olduğuna göre belirlenecektir. Veri toplama işlemimizi yine kalabalık sınıflardan oluşan başka bir tesadüfî örneklem üzerinde tekrar edersek neredeyse kesinlikle ortalama farklı çıkacaktır. Benzeri durum az kişinin olduğu sınıflar için de geçerlidir. Kalabalık sınıflar ile az kişinin bulunduğu sınıflar arasında bir fark bulduğumuzda biliyoruz ki yeniden veri toplasak, fark biraz farklı çıkacaktır. Dolayısıyla, ortalamalar arasındaki fark muğlâktır. Fark, kalabalık sınıflar, az kişinin olduğu sınıflardan daha iyi olduğu için mi sıfırdan büyük; yoksa bizim örneklemimizdeki sınıflar yüzünden mi böyle? Eğer fark hayli büyükse, muhtemelen kalabalık sınıflarla az kişinin olduğu sınıflar arasındaki farkı yansıtıyordur. Eğer hayli küçükse, muhtemelen rasgele gürültüyü yansıtıyordur. Ama ne kadar büyük olunca büyük, ne kadar küçük olunca küçük diyeceğiz? İncelemeye başladığımız problem ve bu bölümün konusu budur.

Yukarıdaki iki örneğe veya daha sonra vereceğim üçüncü örneğe bakacak olursak, bulduğumuz farkın, tesadüfi bir değişkenlik mi, yoksa bize insanların yüksek statüden olduğunu düşündükleri otomobillere daha az korna çalma eğiliminde olduklarını düşündürecek kadar büyük bir fark mı olduğuna karar vermek için bir yol bulmalıyız.

Eğer farklar şansla ilgili olduğu kadar küçük ise, endişe etmeyebiliriz. Diğer taraftan, farklılık için şansla ilgili bir kural koyduysak, muhtemelen daha ileriye bakmalıyız. Şansla ilgili karar hipotez testi konusudur.

## 4.2 Örneklem Dağılımları

Ders değerlendirmeleri ve korna çalmaya üçüncü bir örnek ekleyeceğiz. Girmek için beklediğimiz park yerinden çıkarken çok fazla zaman harcayan sinir bozucu sürücüleri içeriyor. Ruback ve Junieng (1977), sürücüleri, birinin boşalttığı park yerine girmek için bekleyen bir kimsenin olduğu, diğerinde ise olmadığı yüzer kişilik iki gruba ayırıp basit bir çalışma yapmıştır. Ardından sürücülerin park yerinden çıkmak için harcadıkları zamanı ölçmüşlerdir. Boşalttığı park yerine girmek için bekleyen bir kimse olmadığında, sürücülerin park yerini boşaltmaları ortalamada 32,15 saniye, bekleyen bir kimse olduğunda ise ortalamada 39,03 saniye almıştır. Her iki grup için bekleme sürelerinin standart sapması 14,6 saniye olmuştur. Bekleyen birisi olduğunda sürücülerin 6,88 saniye daha fazla zaman harcadıklarına dikkat edin. (Düşünecek olursanız, 6,88 saniye aslında uzun bir zamandır, özellikle de bekleyen sizseniz.)

Burada iki olası açıklama var. İlki, bekleyen birinin olması, park yerinden çıkmak için gereken süreyi etkilemiyor olması ve bekleyen birisi varken ve yokken park yerinden çıkmak için gereken sürenin ortalamada aynı olması mümkündür. Bu durumda, bulduğumuz fark yalnızca örneklemlerimizde hangi sürücülerin olduğunun bir sonucudur. Söylediğimiz şu ki her iki durumdaki sürücülerin tüm evrenlerine ulaşmış olsak, evren ortalamaları ($\mu_{\text{bekleyen yok}}$ ve $\mu_{\text{bekleyen var}}$) aynı olurdu. Bulduğumuz fark ise örnekleme hatası. Alternatif açıklama ise evren ortalamalarının gerçekten farklı olması ve sürücülerin biri bekliyorken park yerinden çıkmalarının daha uzun zaman almasıdır. Örneklem ortalamaları 32,15 ve 32,18 olsaydı, siz de ben de ilk açıklamadan yana olurduk – en azından reddetmek istemezdik. Eğer ortalamalar 32,15 ve 59,03 olsaydı, ikinci açıklamadan yana olurduk – bekleyen birinin olması fark yaratıyor. Ancak bulunan fark arada bir yerlerde ve hangi açıklamanın makul olduğuna karar vermeliyiz.

örneklem dağılımları

"Elde edilen fark tesadüfe atfedilemeyecek kadar büyük müdür?" sorusunu cevaplamak isteriz. Bunu yapmak için **örneklem dağılımları**nı kullanmak zorundayız. Bunlar bize, örnekleme hatasının bir fonksiyonu olarak bir örneklemden diğerine ne ölçüde bir değişkenlik bekleyebileceğimizi gösterir.

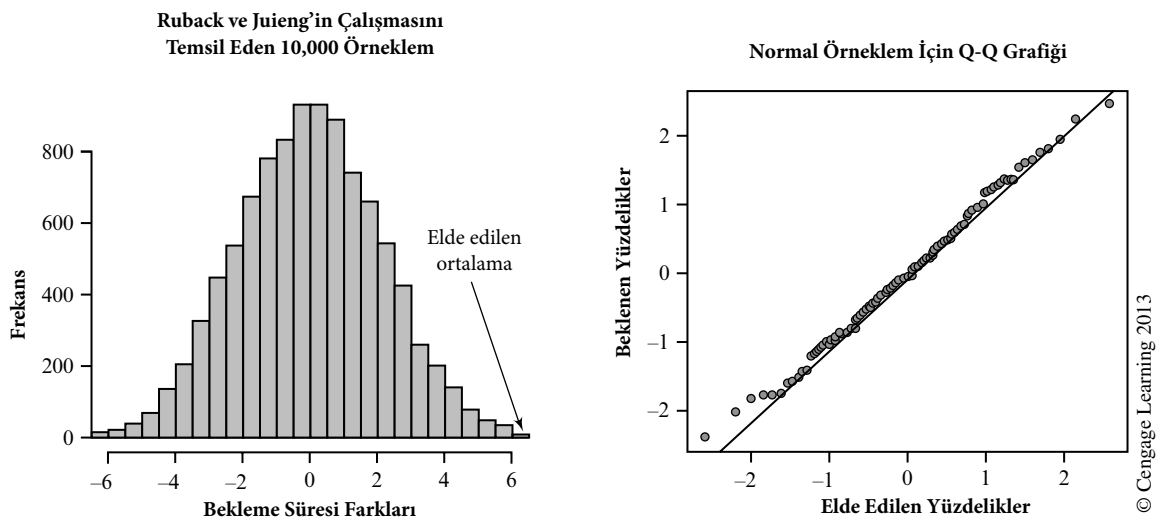
İstatistiksel testlerin tamamının altında yatan en temel kavram bir istatistiğin örneklem dağılımıdır. Denebilir ki örneklem dağılımları olmasaydı, istatistiksel testler de olmazdı. Kabaca, örneklem dağılımları belirli koşullar altında belirli bir istatistiğe dair hangi değerleri elde etmeyi bekleyebileceğimizi (veya bekleyemeyeceğimizi) söyler. Örneklem ortalamaları arasındaki farkların dağılımının standart sapması (dağılımın **standart hata**sı olarak bilinir) deneyin tekrarlarında bu istatistikte (ortalamalar arasındaki fark) ne ölçüde bir değişkenlik bekleyebileceğimizi gösterir. Örneklem dağılımları (örneklem istatistiğinin değeri biliniyorsa) sözü edilen belirli koşulların gerçekten var olma olasılıklarının değerlendirilmesine imkân verir.

standart hata

Temelde, bir istatistiğin örneklem dağılımı, bu istatistiğin tekrar tekrar örneklem alınması (deneyin sonsuz kez yeniden yapılması veya örneklemin sonsuz kez yeniden seçilmesi) durumunda elde edilecek değerlerinin dağılımı olarak düşünülebilir. Örneklem dağılımları hemen hemen her zaman matematiksel olarak belirlenir. Ancak, teoride deneysel olarak nasıl elde edilebileceklerini ele alırsak, neyi gösterdiklerini anlayabilmek daha kolay olur.

**ortalamalar arası farkların örneklem dağılımı**

Örnek olarak **ortalamalar arası farkların örneklem dağılımı**nı ele alacağız çünkü park yerindeki bekleme süreleri örneğiyle doğrudan ilişkili. Ortalamalar arası farkların örneklem dağılımı belirli koşullar (evren ortalamalarının aynı olması koşulu) altında seçilen sonsuz sayıda örneklem çiftlerinin ortalamaları arasındaki farkların dağılımıdır. Ortalamaları ve standart sapmaları bilinen iki evrenimiz olduğunu varsayalım. (Buradaki argümanımızın mantığı için sayısal değerleri önemli olmasa da her iki evrenin ortalamalarının 35, standart sapmalarının ise 15 olduğunu varsayacağız. Genel olarak, evren standart sapmasını nadiren biliriz ama burada bildiğimizi varsayıyoruz.) Yine varsayalım ki elimizde bu örneklemlerden alınmış çok büyük sayıda (teorik olarak sonsuz) her biri yüzer kişilik örneklem çiftleri olsun. Her bir örneklem çifti için iki örneklemin ortalamasını ve bu ortalamalar arasındaki farkı hesaplayacağız. Bunu bütün örneklem çiftleri için yaptıktan sonra farkların dağılımını çizeceğiz. Böyle bir dağılım, ortalamalar arası farkların örneklem dağılımı olacaktır. Tanımladığımız örneklemeyi gerçekleştirmek, yani ortalaması 35, standart sapması 15 olan bir evrenden 100'er kişilik 10.000 tane örneklem çifti seçip her bir çiftteki örneklemlerin ortalamaları arasındaki farkı hesaplamak için R yazılımında dokuz satırlık bir program yazdım. (Kitabın Web sayfasında Sampling Distribution. R adıyla bulunuyor.) Bu dağılımın bir histogramı Şekil 4.1'in sol kısmında, Q-Q grafiği ise sağ kısmında görülüyor. Dağılımın normal olduğundan şüphe duyulacağını sanmıyorum. Dağılımın merkezi 0,0 noktasında bulunuyor. (Ortalamaların kendileri 35 civarında olacaktır.) Şekilden görüldüğü gibi birbiriyle aynı evrenlerden örneklem aldığımızda ortalamalar arası farkların genelde –3 ile +3 arasında olması hayli muhtemeldir. Aynı şekilden yine görülüyor ki bu evrenlerden örneklem alındığında ortalamalar arasındaki farkın 10 veya daha fazla olması pek muhtemel değildir. Bu evrenlerden alınan örneklemlerin ortalamaları arasındaki farkların nasıl değerler alabileceğini biliyor olmamız soruyu tersine çevirip sormamıza imkân verecek: Örneklem ortalamaları arasında elde edilmiş olan bir fark, gerçekte aynı evrenden –ortalamaları aynı evrenlerden– örnekleme yaptığımız hipotezini destekleyen bir bulgu olabilir mi?




**Şekil 4.1** Her biri 25 gözleme dayalı olan ortalamalar arası farkların dağılımı

Ruback ve Juieng (1997) iki durumdaki bekleme süreleri arasında 6,88 saniyelik bir fark bulmuştur. Eğer iki evren ortalaması aynı olsaydı bu kadar büyük bir farkın elde edilmesinin pek muhtemel olmadığı Şekil 4.1'den net olarak görülmektedir. Esasen küçük örnekleme çalışmamda 10.000 vakanın yalnızca 6 tanesi 6,88 kadar veya daha uç bir fark gösterdi. Bu da olasılığı 0,0006 yapıyor. Sürücülerin, park yerinden çıkmalarının birisi beklerken –artık ne sebeple oluyorsa– daha uzun sürdüğü sonucuna ulaşmak için kesinlikle gerekçemiz var.

## 4.3 Hipotez Test Etmenin Teorisi

### Giriş

Davranış bilimlerinde istatistik konusunda yapılan tartışmaların en önemlilerinden biri, hipotez test etmenin teorisi üzerinedir. Hipotez test etmenin mantığı ve teorisi 75 yıldır tartışılagelmektedir, son zamanlarda bu tartışmalar hayli yoğunlaşmıştır. Bu konudaki tartışmalarda sarf edilen ifadeler her zaman yapıcı olmamıştır (karşınızdakinin fikrine "dangalakça yanlış", "bilimsel metodun saptırılması" veya "saçma" demek, genelde insanları tarafınıza katmakta pek işe yaramaz) ancak sonuçta bazı gerçek ve olumlu değişimler elde edilmiştir. Bu değişimler öyle önemlidir ki bu bölüm ve kitabın geri kalanının önemli bir bölümü bu değişimlere uyum sağlamak amacıyla yeniden yazılmıştır.

Hipotez test etmenin rolüne dair görüşler birkaç konuyla ilişkilidir. İlk ve en temel olarak, bazıları hipotez test etmenin en başta akla uygun bir yol olup olmadığını sorgulamaktadır. Ben akla uygun olduğunu düşünüyorum. Ancak akla uygun olsa da olmasa da, bu konunun mantığı yaptığımız işle çok fazla ilgilidir ve deneysel literatürde göreceklerinizin temeline yerleşmiştir. Dolayısıyla, mantıklı bulun ya da bulmayın, bunu anlamak zorundasınız. İkincisi hipotez testi için kullanacağımız mantıkla ilgilidir. Yaygın kullanılan mantık 1920'lerde ve 1930'larda Fisher ve Neyman ile Pearson tarafından ortaya konulan görüşlerin bir karışımıdır. (Fisher ve Neyman ile Pearson görse, bu karışım hakkında derin tereddütler ifade ederlerdi. Yine de pek çokları, özellikle davranış bilimciler tarafından kullanılmaya başlanmış ve kullanılagelmiştir.) Önce bu yaklaşımı tartışacağız. Ardından kabaca aynı yere çıkan ama bir çoğuna göre daha mantıklı ve akla uygun olan, daha güncel versiyondan bahsedeceğiz. Üçüncü ve belki de en önemli olarak, geleneksel hipotez test etmeye ek olarak neyi dikkate almalıyız? İstatistiksel bir test yapıp farkın istatistiksel olarak "$p \leq 0,05$ düzeyinde" manidar olduğunu ilân etmek artık yeterli değildir. Bir hipotez testi, sadece bir ilişkinin güvenilir olup olmadığını veya farkın tesadüf ile açıklanabilir olup olmadığını ortaya koyabilir. Hipotez testi yapmanın yanında okuyucularımıza farkın kendisi, farka dair güven aralığı ve testin gücü hakkında da bilgi sunmalıyız. Bu, kitapta önceki baskılara göre vurgulanan yerlerde ve bulguları tanımladığım bölümlerde değişikliklere yol açacaktır. Bence temel sonuç şu ki hipotez testi –nasıl yaparsanız yapınız– önemlidir ancak yeterli değildir. Bu konudaki tartışmaların, başka hiçbir fayda sağlamasa bile, bizi bir noktaya getirmiş olması hayli önemlidir. İşlenecek çok konu olduğunu görüyorsunuz. Ancak görüşleri ve önerileri anladığınız zaman, bu konuları, alanınızdaki pek çok kimseden daha iyi kavramış olacaksınız.

1990'ların ortalarında Amerikan Psikoloji Derneği hipotez testleri konusunu gözden geçirmesi için bir görev ekibi oluşturdu (Wilkinson, 1999). Rapora https://www.apa.org/science/leadership/bsa/statistical/tfsi-followup-report.pdf adresinden ulaşılabilir. Konunun daha ayrıntılı incelemesi, Nickerson (2000) tarafından yazılan mükemmel makalede bulunabilir. Bu iki doküman, alandaki güncel bakış açısını çok iyi özetlemektedir. Bu öneriler, elinizdeki kitapta hangi konulara yer verileceği konusunu etkilemiştir. Dolayısıyla, güven aralıklarına ve etki büyüklüğü ölçülerine önceki baskılara göre daha sık rastlayacaksınız.

## Hipotez Test Etmeye Geleneksel Yaklaşım

Önümüzdeki birkaç sayfa boyunca hipotez test etmeye geleneksel yaklaşımı ele alacağız. Buradaki açıklama, hemen her istatistik kitabında bulabileceğiniz ve tam olarak anlamanız gereken bilgidir. Buradaki kavramlar hipotez testi denince –kim diyor olursa olsun– kastedilenin temelidir.

Az önce örneklem dağılımlarını tartışıyorduk ki bu dağılımlar araştırmaların veri analizlerinin kalbinde yer alır. Deneysel veya matematiksel fark etmeksizin, örneklem dağılımlarını ilginç bulduğumuzdan değil, önemli sebeplerimiz olduğu için elde ederiz. Olağan sebep hipotez test etmek istememizdir. Sürücülerin park yerinden çıkmak için daha uzun zamana ihtiyaç duydukları örnekteki ortalama sürelerin arasındaki farkların örneklem dağılımına dönelim. "Veri kümesindeki örneklemde gözlediğimiz kadar bir ortalamalar arası fark, aynı ortalamaya sahip evrenlerden örneklem almış olsaydık meydana gelirdi" hipotezini test etmek isteriz. Bu, birisi beklerken park yerinden ayrılış süresi ortalamasının bekleyen biri yokken ayrılış süresi ortalamasından farklı olup olmadığını bilmek istediğimizi söylemenin bir başka yoludur. Böyle bir hipotezi test etmenin bir yolu, eğer gerçekten aynı ortalamaya sahip evrenlerden örneklem almış olsaydık, örneğin 6,88 saniye kadar uç bir değeri elde etme olasılığına dair bir fikir sahibi olmaktır. Örneklem dağılımları tam olarak böyle bir soruya cevap vermek için tasarlanmıştır.

Şekil 4.1'deki örneklem ortalamaları farkı dağılımını elde ettiğimizi (oluşturduğumuzu) varsayalım. Yine varsayalım ki örneklem ortalama farkı 6,88 değil de 2,88 olsun ve 2,88 kadar büyük bir fark bulma olasılığımızın örneklem dağılımımıza bakarak 0,092 olduğunu belirlemiş olalım. (Burada, bu olasılığı nasıl hesapladığımız önemli değil.) Bu durumda düşüncemiz şöyle olacaktı: "Eğer gerçekten aynı ortalamaya sahip evrenlerden örneklem almış olsaydık, 2,88 kadar büyük bir örneklem ortalaması farkı elde etme olasılığı 0,092 olurdu. Çok çok yüksek bir olasılık değil ama düşük de sayılmaz. 2,88 kadar bir örneklem ortalamaları farkı aynı ortalamaya sahip evrenlerden elde edilen örneklemlerde sıklıkla karşılaşıldığına göre, elimizdeki iki örneklemimizin farklı evrenlerden geldiğini düşünmek için bir sebebimiz yoktur."

Esasen örneklem ortalamaları arasındaki fark 6,88 idi. Örneklem dağılımlarından hesapladığımıza göre aynı ortalamaya sahip evrenlerden seçilen örneklemler için 6,88 kadar büyük bir fark elde etme olasılığı sadece 0,0006'dır. O zaman düşüncemiz şöyle olur: Eğer, aynı ortalamaya sahip evrenlerden örneklem alınmış olsaydı, örneklem ortalamaları arasındaki farkın 6,88 kadar büyük çıkması olasılığı sadece 0,0006'dır (pek mümkün olmayan bir olay). Böyle evrenlerden alınan örneklemlerin ortalamaları arasındaki farkın bu kadar büyük çıkması pek muhtemel değil. Dolayısıyla, veri kümemizdeki örneklemlerin farklı ortalamaya sahip evrenlerden geldiği sonucuna ulaşmamız gayet makuldür. Sürücülerin park yerinden çıkması, bekleyen biri olduğunda daha çok zaman alır.

Bu örnekteki adımların farkına varmak önemlidir çünkü çoğu hipotez testindeki mantık bu şekildedir. Test fiilen belli adımlardan oluşmaktadır:

araştırma hipotezi

sıfır hipotezi

1. Sürücüler, bekleyen biri olduğunda park yerinden daha uzun sürede çıkarlar hipotezini, çoğu zaman **araştırma hipotezi** olarak anılır, test etmek istedik.
2. İki durum için davranışların tesadüfî birer örneklemini elde ettik.
3. Gerçekte bu örneklemlerin aynı ortalamaya sahip evrenlerden alındığı hipotezini (**sıfır hipotezi**, $H_0$ olarak anılır) oluşturduk. Bu hipotez park yerinden çıkma süresinin bekleyen birinin olup olmamasına bağlı olmadığını ifade eder.
4. $H_0$'ın (sıfır hipotezinin) doğru olduğu varsayımı altında ortalamalar farkının örneklem dağılımını elde ettik. (Evren ortalamaları aynı olduğunda ortalamalar farkının örneklem dağılımını elde ettik.)
5. Örneklem dağılımını göz önüne alarak, iki örneklemimiz arasında gözlediğimiz fark kadar veya daha büyük bir farkın elde edilmesi olasılığını hesapladık.

6. Bu olasılığa dayanarak $H_0$'ı ret ettik veya ret edemedik. $H_0$ evren ortalamalarının aynı olduğunu ifade ettiği için $H_0$'ın ret edilmesi bu ortalamaların farklı olduğuna dair bir inancı yansıtır. Ancak bu ortalamalar arasındaki fark belirtilmemiştir.

Yukarıdaki tartışma belli ölçüde basite indirgenmiştir, ancak ayrıntılarıyla vakti geldiğinde ilgilenebiliriz. Yaklaşımın mantığı, hepsi olmasa bile çoğu istatistiksel testin mantığını yansıtmaktadır.

1. Araştırma hipoteziyle başla.
2. Yokluk hipotezini oluştur.
3. $H_0$'ın doğru olması koşulu altında ilgili istatistiğin örneklem dağılımını oluştur.
4. Veri topla.
5. Örneklem istatistiğini bu dağılımla karşılaştır.
6. $H_0$'ın doğru olması koşulu altında örneklemde gözlediğimiz kadar uç bir istatistiğin elde edilmesi olasılığına göre $H_0$'ı reddet veya elde tut.

### İlk Engel

Muhtemelen size belli etmeden bir şey anlatmış oldum. En azından ne olduğunu fark etmelisiniz. Bu, hipotez test etme konusundaki kavganın en önemli sebeplerinden biridir. Bu konuda yapabileceğiniz bir şey olmasa da anlamanız gerekir. Sanırım şunu bilmek istiyorsunuz: "Eldeki veri böyle olduğuna göre [verinin böyle olması koşulu altında] $H_0$'ın (Bekleyen birisi varken sürücülerin park yerinden çıkması daha uzun sürmez.) doğru olma olasılığı kaçtır?" Ancak size verdiğim bu değil, gelecekte vereceğim de bu olmayacak. Size farklı bir sorunun cevabını verdim: "Sıfır hipotezinin doğru olması koşulu altında, eldeki verinin elde edilmesi olasılığı kaçtır?" Sizin bilmek isteyeceğiniz sorunun cevabı nasıl verilebilir bilmiyorum. Kötü bir istatistikçi olduğumdan değil, o sorunun cevabı çoğu durumda aşırı zor, sıklıkla da imkânsızdır. Ancak size verdiğim şey hâlâ kullanışlıdır - her zaman da kullanılır. Polis, düz çizgide süremeyen, düzgün konuşamayan bir sürücüye alkollü araç kullanma cezası yazdığı zaman "*Eğer ayık olsaydı,* böyle davranmazdı." demektedir. Bu mantık, hipotez test etme yaklaşımlarının çoğunun merkezindedir.

## 4.4 Sıfır Hipotezi

Gördüğümüz gibi sıfır hipotezi, hipotezlerin test edilmesinde önemli bir rol oynar. Göstermeyi umut ettiğimizin tam zıttı bir hipotez oluşturuyor olmamız insanların sıklıkla zihnini karıştırır. Örneğin, 'üniversite öğrencileri özgüven puan ortalaması 100 olan bir evrenden gelmez' araştırma hipotezini kanıtlamak istediğimizde, hemen puan ortalaması 100 olan bir evrenden geldiklerini söyleyen sıfır hipotezini oluştururuz. Veya birer örneklemin çekildiği iki evrenin ortalamalarının farklı olduğunu ($\mu_1$ ve $\mu_2$) söyleyen araştırma hipotezinin geçerli olduğunu göstermeyi umut ediyorsak evren ortalamalarının aynı (veya bir başka ifadeyle $\mu_1 - \mu_2 = 0$) olduğunu söyleyen sıfır hipotezini oluştururuz. "Sıfır hipotezi" teriminin adını nereden aldığı bu ikinci örnekte kolayca görülüyor. İki evren ortalaması arasındaki farkın sıfır veya yok olduğuna gönderme yapıyor. Bazıları sıfır hipotezine yokluk hipotezi bazıları ise "nul hipotezi" adını verir. Ancak bu sonuncusu mevzuyu çok karmaşıklaştırır, üstelik pek işimize yaramaz. Sıfır hipotezini bazı sebepler yüzünden kullanırız. Fisher tarafından bu kavram ilk oluşturulduğunda, ortaya koyduğu felsefi argümana göre bir şeyin doğru olduğunu hiçbir zaman ispatlayamayız ama yanlış olduğunu kanıtlayabiliriz.

"Her insanın iki kolu vardır." önermesini 3000 kişiyi gözlesek bile ispatlamış olmayız. Ancak tek kollu bir kişiyi bile bulmak önermenin yanlış olduğunu şüpheye yer bırakmaksızın kanıtlar. Fisher'ın temel düşüncesine karşı çıkılabilir olsa da –ki pek çokları çıkmıştır– sıfır hipotezi, istatistikteki baskın konumunu korumaktadır.

Sıfır hipotezini kullanmak için ikinci ve daha pratik bir sebep ise her istatistiksel test için bir başlangıç noktası sağlamasıdır. Üniversite öğrencilerinin özgüven puan ortalamalarının 100'den yüksek olduğunu göstermek istediğiniz durumu ele alın. Diyelim ki bir şeyin doğru olduğunu gösterebilme ayrıcalığını da elde ettiniz. Hangi hipotezi test edeceksiniz? $\mu = 101$ hipotezini mi test edeceksiniz? $\mu = 112$'yi mi? Yoksa $\mu = 113$ hipotezini mi? Anlatmaya çalıştığım şu ki davranış bilimlerinde özel bir **alternatif (araştırma) hipotezi** hemen hemen hiçbir zaman yoktur. Bu olmadan da bize gerekli olan örneklem dağılımını oluşturamayız. (Bu, ilk Neyman-Pearson yaklaşımına karşı dile getirilen ilk argümandı, çünkü sıfır hipotezinin yanlışlanması yerine sıklıkla sanki test edilecek özel bir alternatif hipotez varmış gibi konuşuyorlardı.) Ancak, eğer $\mu = 100$ ile başlarsak, hemen $\mu = 100$ için örneklem dağılımını elde edebiliriz ve ardından eğer veri kümemiz ikna edicisiyse bu sıfır hipotezini ret edip, en baştan göstermeyi umut ettiğimiz sonuca, yani üniversite öğrencilerinin özgüven ortalamasının 100'den büyük olduğu sonucuna ulaşabiliriz.

**alternatif (araştırma) hipotezi**

## İstatistiksel Kararlar

Verimiz, sıfır hipotezinin doğru olması durumunda beklenenden belirgin ölçüde farklılık gösteriyorsa basitçe sıfır hipotezini ret ederiz ve bunun ne anlama geldiği konusunda pek bir görüş ayrılığı yoktur - sıfır hipotezinin yanlış olduğunu kast ederiz. (Bu, okuyucularımıza ne bulduğumuz hakkında daha fazla açıklama yapmamıza gerek olmadığı anlamına gelmez.) Veri, sıfır hipotezini ret etmeye yönlendirmediğinde, durumun anlamlandırılması daha belirsiz ve problemlidir. Sıfır hipotezinin ret edilemeyişini nasıl yorumlayacağız? Sıfır hipotezinin doğru olduğunu "ispatladığımızı" mı söyleyeceğiz? Veya sıfır hipotezini "kabul" edebileceğimizi mi söyleyeceğiz? Veya "elde tutacağımızı" mı, yoksa "kararı sonraya bırakacağımızı" mı?

Reddedilememiş bir sıfır hipotezinin nasıl yorumlanacağı, 75 yıldır istatistik dersi öğrencilerinin başının belası olmuştur ve muhtemelen olmaya da devam edecektir. (Ancak Bölüm 4.10'a bakınız.) Bir şey yanlış değilse doğru olmalıdır düşüncesi sağduyuda kolayca dışlanamayacak kadar kökleşmiştir.

İstatistikçilerin üzerinde anlaştıkları bir konu, sıfır hipotezinin doğru olduğunun "ispat edildiğinin" iddia edilemeyeceğidir. Gözleyeceğimiz 3000 kişinin iki tane kolunun olması her insanın iki kolu olduğunu söyleyen sıfır hipotezini ispatlamayacaktır. Aslında biliyoruz ki birçok normal insanın ikiden az kolu vardır. Sıfır hipotezini reddetmemek çoğu zaman yeterince veri toplamadığımız anlamına gelir.

Konu somut bir örnekle daha kolay anlaşılacaktır. Wagner, Compas ve Howell (1988) lise öğrencilerinde stresle başa çıkma eğitimi programının etkililiğini değerlendiren bir çalışma yapmıştır. (Kitabın başındaki ilk örnek budur.) Eğer çalışma sonucunda programa katılan öğrencilerin stres kaynaklı problemleri, programa katılmamış olan kontrol grubundaki öğrencilerden manidar ölçüde az bulunsaydı, pek tartışma olmadan, programın etkili olduğu sonucuna ulaşırdık. Ancak, eğer gruplar arasında, önceden belirlenmiş istatistiksel manidarlık düzeyi kadar fark olmazsa, ne sonuca ulaşacağız?

Manidar olmayan bir farka dayanarak deney ve kontrol gruplarının evren ortalamalarının eşit olduğunu gösterdiğimiz sonucuna ulaşamayacağımızı biliyoruz. Deney ve kontrol gruplarına uygulanan bu iki girişim, bizim küçük veri kümemiz ile kesin olarak belirleyemeyeceğimiz kadar az bir farka yol açıyor olabilir.

Fisher'ın iddiasına göre manidar olmayan bir sonuç açık olmayan bir sonuçtu. Fisher için tercih sıfır hipotezini reddetmek ile kararı geciktirmek arasındaydı. Fisher, iki grup arasında manidar bir fark bulamamayı, programa katılan öğrencilerin kontrol grubuna göre stresle çok az farkla daha iyi ya da daha kötü başa çıkabilmelerine veya gruplar arasında fark olmamasına bağlardı. Fisher'a göre $H_0$'ın reddedilememesi, sadece verinin bu üç sonuç arasından birini seçmek için yetersiz olduğu anlamına gelir bu nedenle de kararı geciktirmeliyiz. Bu iddiayı Jones ve Tukey (2000) tarafından getirilen bir öneriye değindiğimiz zaman yine göreceksiniz.

Neyman ve Pearson (1933) konuya biraz farklı yaklaşmış ve çok daha pragmatik bir bakış benimsemiştir. Bir okul yönetimi tarafından, bir stres yönetimi programını değerlendirmek üzere işe alındığımızı düşünün. Varsayalım ki ilk çalışmamızın sonuçları manidar çıkmadı. Fisher, program hakkındaki kararı geciktirmemizi önerirdi. Neyman ve Pearson konuya bundan farklı bakardı. Okul yönetiminin karşı karşıya olduğu problem ile ilgilenirlerdi: Bu program için para harcamaya devam etmeli miyiz? İlk çalışma sonuçlarının belirsiz olduğu ve programın faydalı (veya zararlı) olup olmadığına dair güvenebileceğimiz bir karar verebilmek için programı devam ettirip daha çok veri toplamak üzere okul yönetiminden daha çok para istediğimizde, pek etkilenmiş olmayacaklardır. Neyman–Pearson yaklaşımında sıfır hipotezi ya kabul edilir ya da reddedilir. Ancak kabul etmekten kasıt, sıfır hipotezinin doğru olduğunu ispatlamış olmak değildir. Basitçe şunu diyoruz: En azından daha uygun bir veriye ulaşıncaya kadar sıfır hipotezi doğruymuş gibi davranıyoruz. Fisher'ın bakış açısında, ideal okul yönetimi biz program hakkındaki kararımızı verinceye kadar programı desteklemeye devam ederdi. Neyman–Pearson perspektifinde ise eldeki bulguların programı desteklemeye yeterli olmadığı sonucuna ulaşıp, program için para vermeyi keserdi.

Buradaki açıklama Neyman–Pearson yaklaşımını aşırı basitleştiriyor olsa da bakış açılarının özünü içeriyor. Fisher ile Neyman ve Pearson arasındaki tartışma canlı (ve nadiren medenî) bir tartışmaydı. Şimdiki uygulamalar ise her iki bakış açısından da öğeler barındırıyor. Çoğu istatistikçi reddedememenin geçici, belirsiz doğasını yansıtan, "sıfır hipotezi elde tutulmuştur" veya "sıfır hipotezi reddedilememiştir" gibi ifadeler kullanır. Bu ifadelerde Fishercı bir hava vardır. Öte yandan, Neyman–Pearson ekolünün önemli bir özelliği Bölüm 4.7'de işleyeceğimiz II. Tip hatalara (yanlış olan bir sıfır hipotezini reddetmemek) yaptığı vurgudur. Eğer reddetmek veya kabul etmek arasında bir seçim yapacaksanız, hem yanlış bir sıfır hipotezini kabul etme hem de doğru bir sıfır hipotezini reddetme olasılıklarıyla ilgilenmelisiniz. Fisher'ın en başından zaten hiçbir zaman sıfır hipotezini kabul etmeyeceği için yanlış bir sıfır hipotezini kabul etmek hatasıyla ilgili endişe duymasına gerek yoktu.[1] Bazı diğer konuları açıkladıktan sonra Kesim 4.10'da bu konuya geri döneceğiz. Ama önce aynı dilden konuşabilmek için hipotez test etmeye dair bazı temel bilgiyi ve ileride kullanmak üzere bir örneği ele almamız gerekiyor. Bu temel bilgi, hipotez test etme ile ilgili önerilen her yaklaşım için önemlidir.

---

[1] Bir yanda Fisher, diğer yanda Neyman ve Pearson teorilerinin farkları üzerine mükemmel tartışmaları şu kaynaklarda bulabilirsiniz: Gigerenzer, Swijtink, Porter, Daston, Beatty ve Krüger (1989) dördüncü bölüm; Lehman (1993) ve Oakes (1990). Esas meseleler arasında olasılık kavramı, sonsuz evren veya sonsuz yeniden örnekleme düşüncesi, kritik değerin seçimi vardır. Görüş ayrılığı, basit olmaktan çok uzaktır.

## 4.5 Test İstatistikleri ve Bunların Örneklem Dağılımları

örneklem istatistikleri

test istatistikleri

Ortalamanın örneklem dağılımını ele aldık, ancak bunun yerine ortancayı, varyansı, açıklığı, korelasyon katsayısını (ders değerlendirme örneğinde olduğu gibi), oranları (korna çalma örneğinde olduğu gibi) veya önemseyebileceğimiz herhangi bir istatistiği ele alsaydık da konu temelde aynı olacaktı. (Teknik olarak dağılımların şekli farklı olacaktı ama bu bölümde kasıtlı olarak o konuları görmezden geliyorum.) Az önce bahsedilen istatistikler genelde **örneklem istatistikleri** olarak anılırlar, çünkü örneklemin özelliklerini tanımlarlar. Başka türden bir istatistik sınıfı daha vardır: **test istatistikleri**. Bunlar belirli istatistiksel yöntemlerle ilgili ve kendi örneklem dağılımları olan istatistiklerdir. Test istatistiklerine bazı örnekler *t, F* ve $\chi^2$'dir. Bunların bazılarını daha önce görmüş olabilirsiniz. (Görmediyseniz endişelenmeyin. İlerideki bölümlerde ayrı ayrı inceleyeceğiz.) Burası herhangi bir test istatistiğini ayrıntılı olarak açıklamak için bir bölüm değil. Bu bölümü olduğu yere okuyucular teknik ayrıntılarla ilgilenmeleri gerektiğine dair bir endişe yaşamasınlar diye koydum. Bu bölüm, test istatistiklerinin örneklem dağılımlarının, temelde ortalamanın örneklem dağılımı gibi elde edilip kullanıldığını açıklamak için uygun bir yer.

Örnek olarak Bölüm 7'de ele alınacak olan ***t*** istatistiğinin örneklem dağılımını ele alalım. *t* testini daha önce duymamış olanlar için, başka şeylerin yanında, iki örneklemin aynı ortalamaya sahip evrenlerden çekilip çekilmediğinin test edilmesinde kullanıldığını söylemek yeterli olacaktır. $\mu_1$ ve $\mu_2$ iki örneklemin çekildiği iki evrenin ortalamalarını gösteriyor olsun. Sıfır hipotezi evren ortalamalarının eşit olduğu, bir başka ifadeyle $H_0 : \mu_1 = \mu_2$ (veya $\mu_1 - \mu_2 = 0$) hipotezidir. Çok sabırlı olsaydık, birbirinin aynı iki evrenden sonsuz sayıda örneklem çiftleri seçip her bir çift için *t* istatistiğini hesaplayıp (nasıl hesaplandığı sonra gelecek) çıkan ***t*** değerlerinin grafiğini çizerek $H_0$ doğru iken ***t***'nin örneklem dağılımını deneysel olarak elde edebilirdik. Bu durumda $H_0$ doğru olmak zorundadır, çünkü birbirinin aynı evrenlerden örneklemler alarak $H_0$'ı doğru olmaya zorladık. Ortaya çıkan dağılım, $H_0$ doğru iken *t*'nin örneklem dağılımıdır. Daha sonra elimizde *t*'nin belirli bir değerini veren iki örneklemimiz olsaydı, örneklem *t*'sinin değeri ile örneklem dağılımını kıyaslayıp sıfır hipotezini test edebilirdik. Gözlenen *t*'nin değeri, $H_0$ doğru iken *t*'nin örneklem dağılımının bize görmeyi beklememizi söylediği değerlere benzemiyorsa sıfır hipotezini reddederdik.

Yukarıdaki paragrafı, *t* yerine *F*, $\chi^2$ veya herhangi bir test istatistiği koyduğumda, sadece test istatistiğinin nasıl hesaplanacağını içeren küçük değişiklikler yaparak yeniden yazabilirdim. Dolaysıyla, görüyorsunuz ki tüm örneklem dağılımları temelde aynı şekilde elde edilir (birbirinin aynı olan evrenlerden sonsuz sayıda örneklem alarak bunların test istatistiklerini hesapla ve grafiğini çiz).

## 4.6 Sıfır Hipotezi Hakkında Karar Verme

Kesim 4.2'de park yerinden ayrılma süreleri verisini ele aldığımız zaman aslında bir hipotez test ettik. Hatırlayacağınız gibi, ortalaması 35, standart sapması 15 olan bir evrenden çifter halinde örneklemler çektik. (Evren parametrelerinin böyle olduğunu nereden bildiğimiz hakkında endişelenmeyin - Ben uydurdum.) 10.000 tekrarın her biri için örneklem çiftlerinin ortalamaları arasındaki farkların grafiğini çizdik. Sonra gördük ki bu koşullar altında Ruback ve Juieng'in bulduğu kadar büyük bir fark, 10.000 denemenin yalnızca

6 tanesinde ortaya çıkmış (olasılık 0,0006). Bu, öylesi muhtemel olmayan bir sonuç ki elimizdeki ortalamaların, aynı ortalamaya sahip evrenlerden gelmediği sonucuna vardık.

karar varma

Bu noktada hipotez test etmenin **karar verme** süreciyle ilgilenmek durumundayız. 0,0006 olasılığa sahip bir olayın $H_0$'ı redddetmemiz için yeterince düşük bir olasılığa sahip olup olmadığına karar vermeliyiz. Burada yıllar önce geliştirilmiş olan keyfî teamüllere başvuracağız. Bu teamüllerin altında yatan mantık daha sonra daha iyi anlaşılacak fakat şimdilik bunların sadece gelenek olduğunu aklınızda bulundurun. Bir geleneğe göre $H_0$ koşulu altındaki olasılık 0,05'e eşit veya daha küçükse ($p \leq 0,05$), $H_0$'ın reddedilmesini gerektirir. Daha ihtiyatlı bir başka geleneğe göre ise, $H_0$ koşulu altındaki olasılık 0,01'e eşit veya daha küçükse $H_0$ reddedilir. Buradaki 0,05 ve 0,01 değerleri testin **reddedilme düzeyi** veya **manidarlık düzeyi** olarak bilinir. (Farkın 0,05 düzeyinde manidar olduğunu söylediğimizde şunu kast ederiz: Sıfır hipotezi doğru olsaydı bu kadar büyük bir fark denemelerin %5'inde veya daha azında ortaya çıkardı.) $H_0$ koşulu altındaki olasılık önceden belirlenmiş manidarlık düzeyinin altında olduğu zaman $H_0$'ı reddedeceğiz. Bir başka ifadeyle, $H_0$ koşulu altında, manidarlık düzeyinden düşük bir olasılığa sahip her bulgu, **ret bölgesi**ne düşer, çünkü bu bulgu $H_0$'ı reddetmemize yol açar.

reddedilme düzeyi
manidarlık düzeyi

ret bölgesi

Kitabın geri kalan bölümünde standart bir ret düzeyi belirlemek için 0,05 manidarlık düzeyini kullanacağız. Bazılarının bu düzeyi fazla gevşek bulduğunu da aklımızda tutalım.[2] Buradaki örneğimizde elde ettiğimiz olasılık p = 0,0006 olarak hesaplanmıştır. 0,05'ten küçük olduğu aşikâr. $H_0$ koşulu altında, eldeki verinin olasılığı 0,05'ten küçük olduğunda $H_0$'ı reddeceğimizi belirtmiştik. Bu nedenle iki grubun puanlarının, ortalamaları farklı olan evrenlerden geldiğine karar vermek için gerekçemiz vardır.

## 4.7 I. ve II. Tip Hatalar

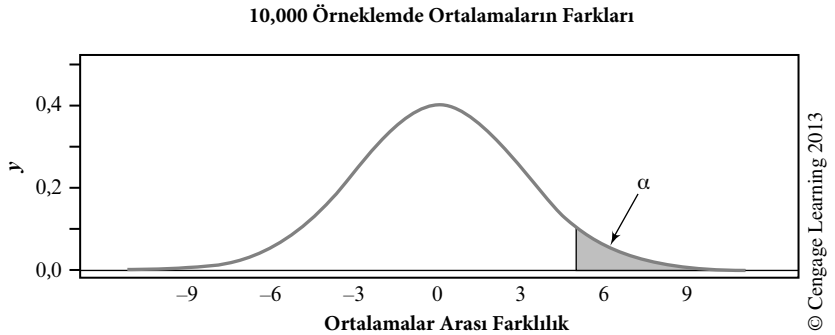
İstatistiksel bir testte karara vardığımızda, yanılıyor olma olasılığımız her zaman vardır. Bu, istatistiksel olsun veya olmasın tüm kararlar için geçerlidir. Ancak istatistikçinin diğer karar vericilerde olmayan bir avantajı vardır. Hem rasyonel bir süreçle karar verir hem de yanılmasına dair koşullu olasılıkları belirleyebilir. Günlük hayatta yalnızca hangisinin doğru olabileceğine dair öznel duygularımızla karar veririz. Ancak istatistikçi aslında doğru olan bir $H_0$ hipotezini reddetme olasılığını tam olarak hesaplayabilir. Bunun hesaplanabilmesi, doğrudan hipotez testinin mantığından gelir.

Park yeri örneğini hatırlayın ama bu sefer Ruback ve Juieng'in bulduğu sonucu görmezden gelin. Sıfır hipotezinin doğru olması durumunda farkların örneklem dağılımı Şekil 4.2'de verilmiştir. Boyalı bölge dağılımın üst %5'lik bölümünü gösterir. En yüksek %5'lik bölümü kesen puana **kritik değer** denir. Kritik değerler ret bölgesinin (veya bölgelerinin)

kritik değer

[2] Burada açıklanan hipotez test etme görüşüne göre sıfır hipotezi doğru olduğunda, çalışmadaki veriyi elde etme olasılığı önceden belirlenmiş manidarlık düzeyinin altında olursa sıfır hipotezi reddedilir, üstünde olursa reddedilmez. Günümüzde böylesi değişmez kuralların uygun olmadığına ve olasılık değerinin kendisine dikkat edilmesi gerektiğine dair pek çok görüş vardır. Bir başka deyişle, geleneksel yaklaşımda (0,05 manidarlık düzeyinde) $p = 0,051$ ile $p = 0,150$ bulgularının istatistiksel olarak "manidar olmayışının" ölçüsü aynıdır. Benzer şekilde $p = 0,048$ ile $p = 0,0003$ istatistiksel olarak aynı ölçüde "manidardır". Alternatif görüş ise $p = 0,051$ sonucunu "neredeyse manidar", $p = 0,0003$ sonucunu ise "çok manidar" olarak ele alır. Bu yaklaşımın önerilecek yanları varsa da, özellikle yalnızca istatistiksel manidarlığın rapor edilmesinden uzaklaşma eğilimi göz önüne alınarak, bu yaklaşıma burada sadık kalınmamıştır. Çoğu bilgisayar programı tam olasılık değerlerini çıktısında verir. Bu değerler sağduyulu bir şekilde yorumlandığında kullanışlı olabilir. Zorluk, "sağduyulu bir şekilde yorumlamanın" ne anlama geldiğini tanımlamaya çalışınca ortaya çıkar.

sınırını (veya sınırlarını) oluşturan $X$ (değişken) değerleridir. Buradaki örnekte kritik değer 4,94'tür.




**Şekil 4.2** Ortalamaların farklarının üst %5'i

Bir sonuç dağılımın en üst %5'lik bölümüne düştüğünde $H_0$'ı reddetmemizi söyleyen bir karar kuralımız varsa, ortalamalar arası fark boyalı bölgeye düştüğünde $H_0$'ı reddedeceğiz. Yani, evren ortalamaları aynı iken, bulduğumuz fark kadar bir farkın gerçekleşmesi olasılığı 0,05 veya daha az olduğunda $H_0$'ı reddedeceğiz. Ancak sürecimizin doğası gereği, bekleyen kişinin park yerinden ayrılma süresine etkisi yokken bile ortalamalar arasındaki farkların %5'i boyalı bölgeye düşecektir. Dolayısıyla, ortalamalar arasında fark olmadığını söyleyen sıfır hipotezinin gerçekte doğru olduğu bir durumda sonucumuzun boyalı bölgede çıkması ve aslında doğru olan sıfır hipotezini reddetme hatasını yapmamız olasılığı %5'tir. Bu tür (aslında doğru olması koşulu altında sıfır hipotezini reddetmek) hatalara **I. Tip hata** denir ve koşullu olasılık $\alpha$ **(alfa)** ile gösterilir. Aynı zamanda ret bölgesinin alanıdır. ($\alpha$ Şekil 4.2'de gösterilmiştir.) Bundan böyle $\alpha$'yı I. Tip hata olasılığını göstermek için kullanacağız.

**I. Tip hata $\alpha$ (alfa)**

I. Tip hata olasılığının "koşullu" yapısını aklınızda bulundurun. Yani I. Tip hatadan bahsettiğimizde $H_0$ doğru iken $H_0$'ı reddetme olasılığını kast ederiz. Yaptığımız hipotez testlerinin %5'inde $H_0$'ı reddedeceğimizi söylemiyoruz. Önemli ve anlamlı değişkenler üzerinde yaptığımız çalışmalarda sıklıkla $H_0$'ı reddetmeyi umarız. I. Tip hatadan kasıt sadece $H_0$'ın doğru olması durumunda $H_0$'ı reddetmektir.

%5'lik I. Tip hata olasılığını çok yüksek bularak ölçütümüzü daha sıkılaştırıp dağılımın %5'lik bölümünün değil de %1'lik bölümünün kullanılmasının daha iyi olacağını düşünebilirsiniz. Bu işlem de gayet meşrudur, ancak fark etmelisiniz ki ölçütü daha sıkılaştırdıkça başka türden bir hata –gerçekte $H_0$ yanlış, $H_1$ doğru iken $H_0$'ı reddetmemek hatası– yapma olasılığını artırırsınız. Bu tür hatalara ise **II. Tip hata** denir ve olasılığı $\beta$ **(beta)** ile gösterilir.

**II. Tip hata $\beta$ (beta)**

II. Tip hatayla ilgili esas zorluk $H_0$ yanlış, $H_1$ doğru iken ($H_1$ koşulu altında) örneklemimizin geldiği evrendeki dağılımın nasıl olacağını neredeyse hiçbir zaman bilememizden kaynaklanır. Bir başka deyişle, hiçbir zaman $H_0$'ın tam olarak ne ölçüde yanlış olduğunu bilemeyiz. Sadece $H_0$ koşulu altında dağılımın şeklini biliriz. Eldeki bağlama dönersek, bekleyen birinin park yerinden ayrılma süresini etkilemediği durumda ortalamalar arası farkın dağılımının nasıl olduğunu biliyoruz ama bekleyen birinin park yerinden ayrılma süresini etkilemesi durumunda ne kadar etkilediğini, fark ettirdiğini bilmiyoruz. Bu durum Şekil 4.3'te gösterilmiştir. Şekilde $H_0$ ile gösterilen dağılım sıfır hipotezinin doğru olduğu durumdaki dağılımı, $H_1$ ise sıfır hipotezinin yanlış, $H_1$'in doğru olduğu durumdaki kuramsal dağılımı göstermektedir. $H_1$ için çizilen dağılımın kuramsal olduğunu hatırlayın. Aslında bu dağılımın yerini tam olarak bilmiyoruz. Sadece $H_0$ ile gösterilen dağılımın sağında

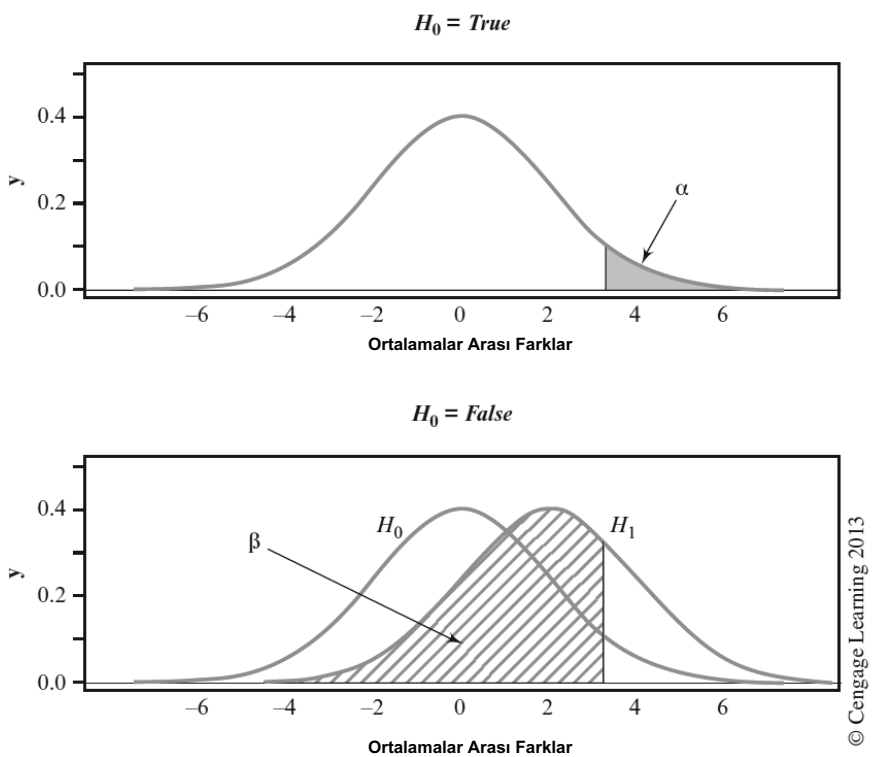
olduğunu biliyoruz. ($H_1$ için olan dağılımı, ortalaması $H_0$ koşulu altındaki dağılımın ortalamasının iki birim sağına gelecek şekilde çizdim. Farkın iki birimi keyfi olarak belirlendi.)

Şekil 4.3'ün üst kısmındaki dağılımın boyanmış bölümü ret bölgesini göstermektedir. O bölgeye (yaklaşık 3,5'in üzeri) düşen bir gözlem sıfır hipotezinin reddedilmesini sağlar. Eğer sıfır hipotezi doğruysa, biliyoruz ki gözlemlerin %5'i bu bölgeye düşer ve tekrarlarımızın %5'inde I. Tip hata yaparız.

Şekil 4.3'ün alt kısmındaki taralı bölge ise II. Tip hata yapma olasılığını ($\beta$) göstermektedir. Bu, bekleyen birinin olmasının park yerinden ayrılma süresini etkilediği ama farkın $H_0$ hipotezini reddetmemiz için yeteri kadar yüksek olmayan (yetmeyen) durumdur.

Şekilde 4.3'te gösterilen, ortalama ve varyansını keyfî olarak aldığım dağılımda $\beta$, normal dağılım tablosundan $\mu = 35$ ve $\sigma = 15$ alınarak 3,5 (kritik değer) ve üzeri puan elde etme olasılığı biçiminde hesaplanabilir. Hesaplamanın nasıl yapıldığı $\beta$'yı anlamanız için önemli değildir. Bu bölüm hesaplamaları içermeyecek şekilde tasarlandığı için doğrudan olasılığı ($\beta$ ile gösterilen bölümün alanını) söyleyeceğim: 0,76. Dolayısıyla bu örnekte, evrendeki bekleme süreleri 3,5 saniye fark ediyorsa ($H_1$ gerçekte doğruysa) evrende deneylerimizin %76'sında $H_0$ yanlış olduğu halde reddetmeyerek II. Tip hata yapacağız.

Şekil 4.3'ten görüleceği gibi, ret bölgesinin sınırını sağa taşıyarak I. Tip hata yapma olasılığını ($\alpha$) 0,05'ten 0,01'e indirirsek I. Tip hata olasılığını azaltmış olsak da II. Tip hata olasılığını artırmış oluruz. $\alpha$'yı 0,01'e çekmek $\beta$'yı 0,92'ye çıkarmak anlamına gelir. Hangi manidarlık düzeyinin kullanılması gerektiği elbette tartışmaya açıktır. Karar, çalışmanız için I. ve II. Tip hatalardan hangisinin daha ciddî olduğunu düşündüğünüze bağlıdır.




**Şekil 4.3** $H_0$ ve $H_1$ altında ortalamalar arası farkların dağılımı

**Tablo 4.1** Karar verme sürecinin olası çıktıları

| | Gerçek Durum | |
|---|---|---|
| Karar | $H_0$ Doğru | $H_0$ Yanlış |
| $H_0$ Ret | I.Tip hata $p = \alpha$ | Doğru karar $p = 1 - \beta$ = Güç |
| $H_0$ Ret değil | Doğru karar $p = 1 - \alpha$ | II. Tip hata $p = \beta$ |

© Cengage Learning 2013

I. Tip hatadan (ortalama sürücüyü kaba olmakla haksız yere suçlamak) kaçınmak daha önemliyse daha sıkı (küçük) bir $\alpha$ düzeyi belirlersiniz. Öte yandan II. Tip hatadan (herkesi, öyle olmadıkları halde, kibar oldukları için takdir etmek) kaçınmak daha önemliyse daha yüksek bir $\alpha$ düzeyi seçersiniz. ($\alpha$ = 0,20 almak bu örnekte $\beta$'yı 0,46'ya indirir.) İnsanlar maalesef pratikte 0,05 veya 0,01'den birini keyfi olarak seçmekte ve $\beta$'yı ise tamamen göz ardı etmektedir. Pek çok durumda yapılabilecek şey bundan ibaret olabilir. (Aslında, muhtemelen öğretim elemanınızın önerdiği $\alpha$ düzeyini kullanacaksınız.) Diğer durumlarda yapılabileceklerin neler olduğunu Bölüm 8'de göreceksiniz.

Şekil 4.3'ün tamamen kuramsal olduğunu vurgulamalıyım. Şekli, ortalamalar arasındaki farkı keyfi olarak 2 ve her iki evren için standart sapmayı 15 aldığım için çizebildim. Farkı 2,5, standart sapmayı 10 alsaydım tüm cevaplar değişecekti. Pratikte genellikle dağılımın ortalamasını ve varyansını bilmeyiz; yalnızca akıllıca tahminlerde bulunabiliriz. Bu nedenle $\beta$'ya dair kestirimlerimiz çok kaba olur. Uygulamada, çalışmamızla tespit edebilmeyi isteyeceğimiz en küçük farkı yansıtan $H_1$ altındaki $\mu$ değerini seçebiliriz çünkü zaten daha büyük farklar daha küçük $\beta$ değerleri üretecektir.

güç

I. ve II. Tip hatalara dair bu açıklamalar sonucunda karar verme sürecini bir tablo ile özetleyebiliriz. Tablo 4.1 bir deneyin dört olası sonucunu göstermektedir. Tablodaki öğeler aslında aşikâr, ancak şimdiye kadar söz etmediğimiz bir kavram var: **güç**. Güç, $H_0$ gerçekte yanlışken $H_0$'ı reddetme olasılığıdır. Yanlış bir $H_0$'ı reddetme olasılığı $\beta$ olduğuna göre güç 1 – $\beta$'ya eşittir. Güç hakkında daha fazla bilgi Bölüm 8'de bulunur.

## 4.8 Tek ve Çift Yönlü Testler

Önceki tartışmalarımız bizi tek ve çift yönlü testler konusuna getiriyor. Park yeri örneğimizde insanların park yerinden çıkarken daha uzun süre harcamalarıyla ilgiliydik. Bu nedenle $H_0$'ı sadece sürücüler daha uzun süre harcadıklarında reddettik. Aslında konuyu daha açık anlatabilmek için bu yaklaşımı seçtim. Varsayalım ki sürücülerimiz birisi bekliyor olduğunda park yerinden 16,88 saniye daha erken ayrılmış olsunlar. Bu, sıfır hipotezi doğruysa hiç muhtemel olmayan bir sonuç olsa da ret bölgesine düşmezdi çünkü ret bölgesinde sadece daha uzun süreler vardı. Sonuçta, kendimizi $H_0$'ın doğru olması durumunda gerçekleşmesi olasılığı çok düşük olan ama beklenenin aksi yönde olan bir veri karşısında $H_0$'ı reddedememiş olarak buluyoruz.

Böyle bir duruma karşı kendimizi nasıl koruyabileceğimiz sorusu ortaya çıkıyor (eğer korunmanın gerekli olduğunu düşünüyorsak). Bir olası cevap, deneyi yapmadan önce hem çok yüksek hem de çok düşük olan belli bir yüzdedeki (öneğin %5) uç puanları ret bölgesine dâhil etmektir. Ancak hem en alt %5'i hem de en üst %5'i reddedersek toplamda ret bölgemiz %5 değil, %10 olur. Bu da manidarlık düzeyimiz ve I. Tip hata olasılığımız olan $\alpha$'nın 0,10 olduğu anlamına gelir. Bu uygun değil çünkü 0,10 kadar yüksek bir $\alpha$ ile çalışmak istemeyiz. $\alpha$'nın en fazla 0,05 olmasını tercih ederiz. Bunu sağlamanın yolu da en düşük % 2,5 ile en yüksek % 2,5'i (toplamda %5) reddetmektir.

**tek-yönlü testler**
**yönlü testler**
**çift-yönlü testler**
**yönsüz testler**

Sadece en düşük (veya sadece en yüksek) ortalama farkları sonucunda $H_0$'ı reddettiğimiz durumlardaki testlere **tek-yönlü** veya **yönlü testler** denir. Bunlarda farkın ne yönde olacağına dair bir tahminde bulunuruz ve ret bölgesi dağılımın yalnızca bir ucunda bulunur. Her iki yöndeki uç puanları reddettiğimiz testlere ise **çift-yönlü** veya **yönsüz testler** denir. Akılda tutmak gerekir ki çift yönlü bir test yaptığımızda bir şey kazanırken (her iki yöndeki uç farklarda sıfır hipotezini reddetme imkânı) bir şey kaybederiz. Tek yönlü bir testin %5'lik bölümüne düşen puanların bir kısmı, buna karşılık gelen çift yönlü testte ret bölgesi dışında kalır çünkü artık o uçta yalnızca %2,5'luk bölüm ret bölgesindedir.

Park yeri örneğinde tek yönlü bir testi seçtim çünkü örneği basitleştiriyordu. Ancak normalde bu, rasyonel bir seçim yapma yolu değildir. Çoğu durumda dağılımın hangi ucunun önemli olduğunu bilmeyiz (Bazen her iki uç da önemlidir.), bu nedenle her iki ucu da kontrol etmemiz gerekir. Bu durum, çocukların sigaraya başlamasını engellemek amacıyla bir kampanya başlattığımızda ortaya çıkabilir. Kampanyanın sigaraya başlama vakalarını azalttığını bulabiliriz. Veya yetişkinler tarafından yürütülen bu kampanya, çocuklar için sigarayı daha cazip, daha heyecanlı hale getirip sigaraya başlama vakalarını artırabilir. Her iki durumda da sıfır hipotezini reddetmek isteriz.

Genelde çift yönlü testlerin tek yönlü testlerden çok daha yaygın olmasının bazı sebepleri vardır. İlk olarak, araştırmacının, verinin nasıl olacağına dair hiçbir fikri olmayabilir. Bu nedenle her duruma hazırlıklı olması gerekir. Bu nadir olsa da bazı keşifsel araştırmalarda görülebilir.

İkinci bir sebep, araştırmacıların, verinin hangi yönde çıkacağına emin olsalar da yanılıyor olma durumlarına karşı hazırlıklı olmak istemeleridir. Bu durum, tahmin edeceğinizden daha sık ortaya çıkar. (Dikkatlice hazırlanmış hipotezlerin, aksi yönde yazılmış olmak gibi can sıkıcı bir huyları vardır. Neden öyle olduğu ise sonradan çok aşikâr görünür.) Sigara karşıtı kampanya bunun bir örneğidir. Gerçekten de kötü hazırlanmış bazı kampanyaların faydadan çok zarar getirdiğine dair kanıtlar vardır. Verinin aksi yönde çıkma durumuna karşı sıklıkla sorulan bir soru vardır: "Neden tek yönlü bir test yapıp veri aksi yönde çıkarsa, testi çift yönlü test olarak yapmıyoruz?" Bu yaklaşım uygun değildir. Teste önce bir yöndeki %5 ile başlayıp sonradan aksi yöndeki %2,5'luk bölümü de ret bölgesine eklerseniz toplamda %7,5 eder. Bu durumda bir yöndeki sonuçların %5'ini (veri beklediğiniz yönde çıkarsa), öbür yöndeki sonuçların da %2,5'unu (veri aksi yönde çıkarsa) reddedeceksiniz. %5 + %2,5 = %7,5. Bunu inkâr etmek mümkün değil. Bir başka deyişle, dondurmasına yazı tura atarken "tura" deyip paranın ne geldiğini gördükten sonra "yazı" diye değiştirme hakkım olduğunu söylesem benimle yazı tura atar mısınız? Veya yazı tura attığımızda sonuç sizin lehinize çıktığında "Üç tanesinden ikisini kazanan dondurmayı alır." desem sizce adil olur mu? Her iki stratejiye de karşı çıkarsınız. Çıkmalısınız da. Aynı sebepten ötürü, tek yönlü veya iki yönlü test arasındaki seçim veri toplanmadan önce yapılır. Bu aynı zamanda, genelde iki yönlü test yapılmasının da bir sebebidir.

Çift yönlü testlerin daha yaygın olmasının üçüncü sebebi, bazı durumlarda tek yönlü bir testin tanımlanamıyor oluşudur. İkiden çok grubun var olması bunun bir örneğidir. Varyans analizini incelerken bu konuyu ele alacağız. İkiden çok grubun olduğu durumlarda tek yönlü bir test tanımlamak hemen hemen imkânsızdır. Bu durumlarda çok yönlü bir test kullanacağız. Bölüm 6'da söz edilecek olan ki-kare testi de tek yönlü bir test yapmayı engeller. Bazı ek adımlarla mümkün olsa da bunu önermem.

Her ne kadar önceki bölümler ve ilerideki Kesim 4.10 çift yönlü testlerin kullanılmasını desteklese de, kitabın genelinde çift yönlü testlere yer versek de çok sıkı, değişmez kurallar yoktur. Son karar, farklı türdeki hataların ne ölçüde ciddi olacağına dair ne bildiğiniz üzerine kurulu olacaktır. Akılda tutulması gereken, dağılımın bir ucunu ele alırsak, tek yönlü

test ile iki yönlü test arasındaki fark, sadece farklı kesme noktaları kullanmasıdır. $\alpha = 0{,}05$ düzeyindeki çift yönlü bir test, $\alpha = 0{,}01$ düzeyindeki tek yönlü bir testten daha liberaldir.

Örneklem dağılımları yoluyla hipotez test etme konusunu iyi anladıysanız bu dersin geri kalanı size daha kolay gelecektir. Yeni karşılaştığınız her istatistik için şu iki soruyu sormanız gerek:

- İstatistik nasıl ve hangi varsayımlar altında hesaplanmıştır?
- $H_0$ altında bu istatistiğin örneklem dağılımı nasıl görünür?

Eğer bu iki sorunun cevabını biliyorsanız, testiniz eldeki veri için test istatistiğini hesaplayıp örneklem dağılımıyla kıyaslayarak yapılır. İlgili örneklem dağılımları tablolarda listelenmiş olduğu için aslında yapmanız gereken, hangi testi kullanacağınızı bilmek ve test istatistiğini hesaplamaktır. (Elbette istatistik dersinde hipotez testinden başka konular da var. Dolayısıyla belki de fazla göklere çıkarıyor olabilirim. İstatistiği anlamak için sadece belirli testlerin hesaplamalarını yapıp değerlendirmek yeterli değil. Hesaplama, özellikle günümüz bilgisayar imkânlarıyla işin kolay kısmı.)

## 4.9 Sıfır Hipotezini Reddetmek Ne Anlama Gelir?

İyi yetişmiş araştırmacılarda bile, sıfır hipoteziyle ilgili, reddetmenin ne anlama geldiğine dair karıştırma vardır. Daha önce sözünü ettiğim gibi, sıfır hipotezinin doğru olması koşulu altında çalışmamızdaki gibi bir veri kümesi elde etme olasılığını hesaplarız, verinin bu olması koşulu altında sıfır hipotezinin doğru olması olasılığını değil. Tekrarladığımın farkındayım zira konu tekrarlanmayı hak ediyor.

Varsayalım ki iki evren ortalamasının aynı olduğunu söyleyen sıfır hipotezini $p = 0{,}045$ düzeyinde reddettik. Şunu söylemek pek çoklarına çekici gelir. Bu sonuca göre sıfır hipotezinin doğru olması olasılığı 0,045'tir. Oysa bu olasılığın anlamı bu *değildir*. *Sıfır hipotezi doğru olsaydı*, bu fark kadar büyük bir fark elde etme olasılığının sadece 0,045 olduğunu gösterdik. Bu, sıfır hipotezinin doğru olması olasılığının 0,045 olduğunu söylemekten çok farklıdır. Hipotez koşulu altında verinin olasılığı ile veri koşulu altında hipotez olasılığını karıştırıyoruz. Bunlara **koşullu olasılıklar** denir. Koşullu olasılıklardan Bölüm 5'te söz edeceğiz. 0,045 olan olasılık, $H_0$'ın doğru olması koşulu altında, [$H_0$ doğru iken] verinin bu şekilde olması olasılığıdır [$p(D \mid H_0)$ biçiminde gösterilir]. Verinin eldeki biçimde olması koşulu altında [veri böyle iken] $H_0$'ın doğru olması olasılığı [$p(H_0 \mid D)$] değildir. Bu konuda okuduğum en iyi inceleme Nickerson (2000) tarafından yazılan bir makaledir.

koşullu olasılıklar

## 4.10 Hipotez Test Etmeye Alternatif Bir Bakış

Şimdiye kadar hipotez testleriyle ilgili sunduklarım geleneksel yaklaşımdır. Her istatistik kitabında vardır ve iyi anlamanız gerekir. Ancak son zamanlarda hipotez test etmeye farklı bakış açılarına olan ilgi arttı. Özellikle Jones ve Tukey (2000) tarafından önerilen yaklaşım geleneksel yaklaşımın bazı problemlerine dikkat çekiyor.

Daha sonra Bölüm 7'de de kullanacağımız, iki evren ortalamasının karşılaştırıldığı bir örnekle başlayacağız. Adams, Wright ve Lohr (1996) bir grup homofobik heteroseksüel erkeğe ve bir grup homofobik olmayan heteroseksüel erkeğe cinsel içerikli, erotik homoseksüel görüntüler içeren bir video göstermiş ve katılımcıların cinsel uyarılmalarını kaydetmiştir. Videoları izleyen iki grup arasında cinsel uyarılma bakımından fark olup olmadığını incelemek istiyorlardı. (Dikkat ederseniz, hangi grubun daha yüksek ortalamaya sahip olacağını beklediklerini söylemedim.)

Geleneksel hipotez test etme yaklaşımı sıfır hipotezini, $\mu_h$ homofobik erkeklerin ortalaması ve $\mu_D$ homofobik olmayan erkeklerin ortalaması olmak üzere, $\mu_h = \mu_d$ biçiminde oluşturmaktır. Geleneksel (iki yönlü) alternatif hipotez ise $\mu_h \neq \mu_d$ şeklindedir. Böyle bir durumda sıfır hipotezinin hiçbir zaman doğru olmayacağını pek çok insan dile getirmiştir. Tüm homofobik erkekler evrenine ve homofobik olmayan erkekler evrenine ulaşsak, bu iki evrenin ortalamalarının virgülden sonra sayısız basamağa kadar aynı olacağına inanmak mantıklı değildir. Ortalamalar kaç olursa olsun, kesinlikle en azından küçücük bir miktar da olsa farklılık göstereceklerdir.[3] Dolayısıyla en baştan biliyoruz ki sıfır hipotezi zaten yanlıştır. Bu durumda zaten yanlış olduğunu bildiğimiz bir hipotezi neden test ettiğimizi sorabiliriz. (Pek çok insan da bu soruyu sormuştur.)

Jones ve Tukey (2000) ile Harris (2005) esasen üç hipotez veya sonucumuzun olduğunu iddia etmiştir - Jones ve Tukey "sonuç" diye bahseder. Biri $\mu_h < \mu_d$, diğeri $\mu_h > \mu_d$, sonuncusu ise $\mu_h = \mu_d$. Bu üçüncüsü geleneksel sıfır hipotezidir ve az önce dediğimiz gibi bunun zaten hiçbir zaman tam olarak doğru olmayacağını biliyoruz. Dolayısıyla bunu eleyebiliriz. Bu üç sonucun her biri farklı hareket tarzına yol açar. Eğer ilkini ($\mu_h < \mu_d$) test edip reddedersek, homofobik erkeklerin daha çok uyarıldıkları sonucuna, ikincisini ($\mu_h > \mu_d$) test edip reddedersek de homofobik erkeklerin daha az uyarıldıkları sonucuna ulaşırız. Eğer her ikisini de reddedemezsek karar vermek için yeterli kanıtımızın olmadığı sonucuna varırız - evren ortalamaları neredeyse kesinlikle farklı ama hangisinin daha büyük olduğunu bilmiyoruz.

Bu yaklaşım ile geleneksel yaklaşım arasındaki fark küçük görünebilir ancak önemlidir. Her şeyden önce Lyle Jones ve John Tukey bize bir şey söylediği zaman kesinlikle dinlemeliyiz. Bunlar doktorayı yeni bitirmiş insanlar değiller, hayli saygın istatistikçilerdir. (Eğer istatistik Nobel Ödülü var olsaydı, John Tukey alırdı.) İkincisi, sıfır hipotezinin hiçbir zaman tam olarak doğru olmadığını ancak bazen eldeki verinin hangi evren ortalamasının büyük olduğu sonucuna ulaşmaya yetmediğini kabullenir. Dolayısıyla, "sıfır hipotezi reddedilememiştir" veya "sıfır hipotezi elde tutulmuştur" gibi muğlak ifadeler kullanmak yerine sıfır hipotezi fikrini kenara kaldırıp şunu söylüyoruz: "$\mu_h$ mi $\mu_d$'den büyük yoksa $\mu_d$ mi $\mu_h$'den büyük karar veremedik." diyoruz. Üçüncüsü, bu iki tane tek yönlü test yapıyormuşuz gibi görünüyor ama önemli bir farkı var. Tek yönlü geleneksel bir testte hangi yönü test ettiğimizi önceden belirleriz. Eğer bulgu, o yönün uç kısmında yer alırsa sıfır hipotezini reddeder $\mu_h < \mu_d$ sonucuna varırız. Eğer bulgu, diğer yönde çıkmışsa ne kadar uçta yer alırsa alsın sıfır hipotezini reddedemeyiz. Ancak bu Jones ve Tukey'nin önerdiği şey değil. Onların yaklaşımı, yönü önceden belirlemeyi gerektirmiyor.

Jones ve Tukey, yönü önceden belirlemeyi değil, veriyi topladıktan sonra bulgunun herhangi bir uçta olup olmadığını belirlemeyi öneriyor. Alt uçtaysa $\mu_h < \mu_d$ sonucuna, üst uçtaysa $\mu_h > \mu_d$ sonucuna ulaşırız. Eğer hiç biri değilse, verinin karar vermek için yeterli

---

[3] Virgülden sonra üçüncü basamağa kadar önemsiz şeyler üzerinde oyalandığımızı düşünebilirsiniz. Ancak homofobi üzerine düşündüğünüzde, hangi grubun arasındaki fark ne yönde olursa olsun az olmayacağını beklemek makuldür. Park yeri örneği için de aynısı geçerlidir. Dünya park yeri için bekleyen birinin varlığı ya da yokluğundan etkilenmeden sadece arabasına binip giden insanlarla doludur. Ama bekleyen kişiye ayak bağı olmamak için hızlıca park yerinden çıkmaya gayret eden çok düşünceli insanlar var olduğu gibi bekleyen biri olduğundan park yerinden kasten yavaş yavaş ayrılan arızalı insanlar da vardır. Son iki gruptan hangisi daha kalabalıktır bilmiyorum ama eşit sayıda olmadıklarından eminim. Fark edilebilir ölçüde bir ayrım olacaktır. Sıfır hipotezinin doğruya çok yakın olabileceği aşırı basit olmayan hiçbir örnek düşünemiyorum.

olmadığını ifade ederiz. (Dikkat ederseniz, hiç "reddetmek" sözünü kullanmadım, "sonuç" dedim. İnce ama bence önemli bir fark.)

Jones ve Tukey bir adım daha ileri gidip manidarlık düzeyini de değiştiriyor. İlk olarak sıfır hipotezinin doğru olma olasılığının 0,00 olduğunu biliyoruz. [Bir başka ifadeyle, $p$ ($\mu_h = \mu_d$) = 0.] Fark küçük olabilir ama yine de bir farktır. Bu yaklaşımda, sıfır hipotezini reddetmemek gibi bir hata yapıyor olamayız çünkü yeterince kanıt bulunamadığını söylemekle bir hipotezi hatalı bir biçimde reddetmek aynı şey değildir. Jones ve Tukey (2000) şöyle ifade etmiştir:

> Bu yöntemle, bir sonuç yalnızca aksi yönde ise hatalı olabilir. Bir yönde bir sonuca ulaşıldığında (aslında bilinmeyen) gerçek aksi yöndeyse hata vardır. Sonucun yönüne karar verilemediğini ifade etmek değerlendirilememiş bir imkân olsa da bir hata değildir. Hatayı kontrol etmek ve aynı zamanda değerlendirilememiş imkânları (belirsiz sonuçları) en aza indirmeyi isteriz. (s. 412)

Dolayısıyla, iki şeyden yalnızca birisi doğrudur - ya $\mu_h > \mu_d$ ya da $\mu_h < \mu_d$. Eğer $\mu_h > \mu_d$ doğru ise (yani homofobik erkekler, homoseksüel videolarla homofobik olmayanlara göre daha fazla uyarılıyorlarsa) yapabileceğimiz tek yanlış var: Hatalı bir biçimde tersi sonuca ulaşmak - $\mu_h < \mu_d$. Eğer dağılımın her iki ucundaki %2,5'lik bölümleri kullanan, geleneksel çift yönlü hipotez testini kullanıyor olsaydık, bu hatanın olasılığı 0,025 olurdu. Öte yandan gerçekte $\mu_h < \mu_d$ doğru ise yapabileceğimiz tek hata $\mu_h > \mu_d$ sonucuna ulaşmak olurdu ki bunun olasılığı da en çok 0,025'tir. Dolayısıyla eğer geleneksel bir çift yönlü testin kesme noktalarını kullanırsak, zıt yönlü hata yapma olasılığımız en çok 0,025 olacaktır. Burada alanları veya olasılıkları toplamamız gerekmiyor çünkü bu hatalardan yalnızca birini yapmak mümkün. Jones ve Tukey bu nedenle, I. Tip hata olasılığını 0,05 yapacak şekilde her bir uçtan %5'lik bölümü kesen noktaları (geleneksel bir çift yönlü testte $\alpha$ = 0,10 manidarlık düzeyi) kullanmayı önermektedir. Bu doğrudur ama yine de dergi hakemleri dâhil, geleneksel eğitim almış pek çok meslektaşınızın artık bu noktada sağı solu belli olmaz. Bu nedenle şansınızı pek zorlamak istemeyebilirsiniz.

Bu noktada öğrencilerin iki itirazla ellerini kaldırdıklarını tahmin edebiliyorum. İlki kılı kırk yardığımız iddiası olacaktır. Cevabım: "Hayır, kılı kırk yarmıyoruz." Bu konular literatürde heyecanla tartışılmıştır. Hatta bu tartışmalar esnasında bazı insanlar hipotez testlerini toptan bırakmamız gerektiğini iddia etmiştir (Hunter, 1997). Jones-Tukey formülasyonları hipotez test etmeye anlam kazandırmıştır. Ayrıca önerilerinin tamamını yerine getirirseniz istatistiksel güç de artar. (İnanıyorum ki "hipotez test etme" yerine "sonuca ulaşma" tabirini tercih ederlerdi.) İkincisi öğrenciler pekâlâ birçok sayfayı geleneksel yaklaşıma ayırıp ardından bir iki sayfada onun tamamen yanlış olduğunu neden söylediğimi sorabilirler. Bunu en başta cevaplamaya çalıştım - geleneksel yaklaşım yaptıklarımıza öylesine yerleşmiş ki onu anlamadan idare edemezsiniz. Okuyacağınız çalışmaların çoğunun arka planında geleneksel yaklaşım var ve meslektaşlarınız geleneksel yaklaşımı anlamanızı bekleyecek. Daha iyi bir alternatifin olması sizi geleneksel yaklaşımı anlamak zorunda olmaktan kurtarmıyor. Jones ve Tukey'nin önerdiği gibi $\alpha$ düzeylerini değiştirmediğiniz sürece hemen hemen aynı işlemleri yapacak ama daha mantıklı sonuçlara ulaşacaksınız. Her zaman çift yönlü hipotez testi yapmanızı şiddetle öneririm. Muhtemelen $\alpha$ = 0,05 düzeyinde yapmalısınız. Ancak aklınızda bulunsun, $\alpha$'yı bu düzeyde tuttuğunuz sürece farkın yönü hakkında yanılıyor olma olasılığınız yalnızca 0,025 olacaktır.

## 4.11 Etki Büyüklüğü

**etki büyüklüğü**

Birinci Bölüm'ün başlarında, manidarlık testinin ötesinde **etki büyüklüğü** ölçülerinin rapor edilmesini öngören bir hareketin başladığından söz ettim. Aslında bazı dergiler çoktan bu konuda ısrar etmeye başladı. İlerleyen bölümlerde bu konuyu genişleteceğim ama bahse

girerim kimse size bir etki büyüklüğü ölçüsünden gizlice söz ettiğimi anlamadı. Park yerinde bekleme süresi hakkında yazarken Ruback ve Juieng (1997) tarafından bulunan farkın 6,88 saniye olduğunu ve bu farkın bekleyen için hayli çok olduğunu ifade ettim. Bir adım öteye gidip bekleme sürelerinin standart sapması 14,6 saniye olduğuna göre neredeyse yarım standart sapmalık bir fark olduğunu da söyleyebilirdim. Bekleme süreleri arasındaki farkın değerini veya "yaklaşık yarım standart sapma olduğunu söylemek etkinin ne kadar büyük olduğuna dair bir ölçü bildirir—hayli muteber bir ölçüdür. Daha formal etki büyüklüğü ölçüleri hakkında söylenecek şeyler var ama bu kadarı size ne oldukları hakkında bir fikir verecektir. Bu konuyu ilerideki bölümlerde tekrarlayıp derinleştireceğim.

Bu konuda bir şey daha söylemeliyim. Hipotez testi üzerine tartışmaları anlamanın önündeki önemli bir zorluk istatistikçilerin terminoloji kullanımı konusunda yıllardır özensiz olmasıdır. Örneğin, sıfır hipotezi reddedildiğinde "manidar bir fark" bulunduğunu rapor etmek hayli yaygındır. Pek çok okuyucu çalışmanın "önemli bir fark" bulduğunu düşünmekte haklıdır. Oysa kastedilen kesinlikle bu değildir. İstatistikçiler ve araştırmacılar "manidar" kelimesini "istatistiksel olarak manidar" teriminin kısa hali olarak kullanır. Farkın, çok az olsa bile, tesadüf ile açıklanamayacağı anlamına gelir. Etki büyüklükleri üzerine son zamanlarda yapılan vurgu, bulgunun ne kadar önemli olduğuna dair az da olsa bir fikir verebilmesi içindir. Bu kitapta "manidar" kelimesinden önce "istatistiksel olarak" ifadesini yazmaya çalışacağım ama her zaman yapabileceğime söz veremiyorum.

## 4.12 Son Bir İşlenmiş Örnek

Bu sefer, ilk örnek gibi örneklem ortalaması yerine tek bir puan barındıran farklı bir örnek ele alalım. Bu örneği, daha açıklayıcı olması için standart normal dağılım formülüyle ilgili bildiklerimizden yola çıkarak ele alacağız. Parmak vuru hızı, nöropsikoloji alanında yaygın olarak kullanılan bir ölçüdür. Çalışmalar, Alzheimer, travmatik beyin hasarı ve şizofreni rahatsızlıkları bulunan insanlardan oluşan evrenlerde düşük parmak vuru performansı olduğunu göstermiştir. Halstead Reitan Bataryası'nın normatif veri kümesi yaygın olarak ulaşılabilir durumdadır. Bize gönderilen bir hastanın puanını, elimizdeki bu veri ile kıyaslayabiliriz. Bu normlar genel olarak geniş evrenler üzerinden hesaplanmıştır. Normal kadınlar için performans ortalaması 47,8, standart sapması ise 5,3'tür. Puanı 35 olan bir kadın hastamız olduğunu varsayalım. Bu hastanın performansı, nörolojik bir hasardan şüphelenmeye yetecek kadar normalin altında mıdır?

Açıklamalarımın devamında hipotez test etmede geleneksel yaklaşımı benimseyeceğim ancak siz Jones ve Tukey yaklaşımı ile arasındaki farkı görebiliyor olmalısınız. Burada, iki olası seçeneğimiz var: Kişide hasar vardır ya da yoktur. Eğer hasar yoksa kişinin puanının geldiği evrenin ortalamasını ve standart sapmasını biliyoruz: Sırasıyla 47,8 ve 5,3. Hastamızda hasar varsa puanının geldiği evrenin ortalamasını ve standart sapmasını bilmiyoruz; bilmemiz de gerekmez. Hatta hastanın puanını herhangi bir evrenden seçilmiş bir puan olarak ele almak zorunda da değiliz. Kişinin durumu hakkında makul bir sonuca ulaşmak için sıfır hipotezini "Kişide hasar yoktur." olarak kuracağız. Daha doğrusu "Kişi, ortalaması 47,8 olan bir evrenden gelmiştir." $H_0$: $\mu = 47{,}8$. $H_1$ hipotezini ise "Kişide hasar vardır." biçiminde oluşturacağız ($\mu \neq 47.8$). (Dikkat ediniz, Jones ve Tukey $H_1$: $\mu < 47{,}8$ ve $H_2$: $\mu > 47{,}8$ hipotezlerini eş zamanlı olarak test eder ve sıfır hipotezi yerine de karar vermek için yeterli veri olmadığı sonucunu kullanırdı.)

Geleneksel yaklaşıma göre, şimdi iki ya da tek yönlü test arasında bir seçim yapmalıyız. Normalden daha hızlı parmak vuruş performansının tanısal bir anlamı olmadığı için bu örnekte tek yönlü testi tercih edeceğiz. Tek yönlü testi seçtiğimiz için de alternatif hipotezi $H_1$: $\mu < 47{,}8$ biçiminde oluşturacağız.

İstatistiksel teknikleri eldeki veriye uygulamadan önce bir karar daha vermemiz gerekiyor. Testimiz için bir manidarlık düzeyi seçmeliyiz. Bu seferki testi %1 yerine %5 düzeyinde yapmayı tercih ediyorum çünkü hem kitapta genel olarak %5'i standart olarak alıyorum hem de II. Tip hata yapmaktan I. Tip hata yapmaya göre daha çok endişe ediyorum. Eğer I. Tip bir hata yaparsam kişide hasar olduğu sonucunu ileri tetkikler yanlışlayacaktır. Ama II. Tip bir hata yaparsam kişide hasar olmadığı sonucuna vararak daha ciddî bir hata yapmış olacağım.

Ardından normal bir kimsenin 35 kadar düşük bir puan alması olasılığını hesaplayarak devam etmeliyiz. Önce 35 ham puana karşılık gelen $z$ puanını buluruz. Bu hesabı Bölüm 3'ten biliyoruz. Her ne kadar normatif veriden elde edilen değerler (geniş de olsa bir örneklemden hesaplandıkları için) gerçek evren parametreleri olmasa da buradaki ihtiyacımızı karşılayacak kadar yakınlar. (Normatif veri kümeleri küçük veri kümeleri olduğunda uygulanacak teknikler Bölüm 7'de yer alıyor.) 35 ham puanın, ortalamadan –yani 47,8'den– kaç standart sapma uzaklıkta olduğunu söyleyecek olan $z$ puanını hesaplayabileceğimizi Bölüm 3'ten biliyoruz. Eğer bu çok uç bir puan ise, hastamızın nörolojik bir hasarı olduğunu düşünme eğiliminde olacağız. Bunu yapmak için aşağıdaki hesaplamayı gerçekleştiriyoruz.

$$z = \frac{X - \mu}{\sigma} = \frac{35 - 47,8}{5,3} = -2,42$$

Hastanın puanı normal kimselerin ortalamasının 2,42 standart sapma altındadır. Ardından, –2,42 ve daha düşük $z$ puanı alma olasılığını görmek için $z$ tablosuna bakarız. Ek $z$'den, bu olasılığın 0,0078 olduğunu görebiliriz.

Bu olasılık değeri, belirlemiş olduğumuz %5 sınırının altında olduğu için sıfır hipotezini reddedeceğiz. Kararımızın gerekçesi, nörolojik hasarı bulunmayan kişilerin evreninden bir kişiyi gözlemiş olsaydık, 35 kadar düşük bir puan alması olasılığının hayli düşük olmasıdır. Karar doğrultusunda, gözlediğimiz kişinin nörolojik hasarı olan bireyler evreninden geldiği ve tedaviye ihtiyacı olduğu sonucuna ulaşacağız.

Sıfır hipotezini reddederek I. Tip hata yapmış olabileceğimizin farkında olmak önemlidir. Nörolojik hasarı bulunmayan kadınlardan örneklem alsaydık bunların %0,78'inin bu kadar düşük puan alacağını biliyoruz. Buradaki katılımcımızın nörolojik hasarı olmayıp yalnızca düşük performans göstermiş olması da mümkündür. Sadece şunu söylüyoruz: Böyle bir olay o kadar düşük olasılıklıdır ki biz alternatif hipotezden yana oluyoruz.

## 4.13 Ders Değerlendirmelerine ve Kaba Araç Kullanıcılarına Dönüş

Bu bölümdeki ilk örneğimiz ders değerlendirme sonuçları ile öğrencilerin o derslerden bekledikleri puanlar arasındaki ilişkiyi inceliyordu. İkinci örneğimizde ise trafik ışıkları yeşile döndüğünde hareket etmeyen yüksek ve alt sosyoekonomik sınıftan otomobillerin arkasındaki sürücülerin korna çalma olasılıkları ele alınmıştı. Daha sonra Bölüm 9'da göreceğiniz gibi, ilk örnek, ilişkiyi göstermek için bir korelasyon katsayısını kullanmaktadır. İkinci örnek ise sadece oranları kıyaslamaktadır. Her iki örnek de, bu bölümde ele alınan tekniklerle çözümlenebilir. İlk örnek için, öğrencilerin bekledikleri notlar ile ders değerlendirme notları arasında bir ilişki yoksa, tüm öğrenciler evrenindeki gerçek korelasyon değerinin 0,00 olmasını bekleriz. Sıfır hipotezini "Evren korelasyonu 0,00'dır." biçiminde oluşturup bir örneklemdeki korelasyonun çalışmadaki örneklemden elde edilen değer kadar büyük çıkması olasılığını sorarız. İkinci örnek için, sıfır hipotezini "Evrende, sürücü-

lerin yüksek ve düşük sosyal statüden araçlara korna çalma oranları arasında fark yoktur." biçiminde oluştururuz. Daha sonra, sıfır hipotezinin doğru olması koşulu altında, bir örneklemde, elde ettiğimiz kadar büyük farkın (0,34) çıkması olasılığını hesaplarız. Bu, park yeri örneğimize çok benzer. Sadece birisi oranlarla diğeri ortalamalarla ilgili, o kadar. Şimdi bu testleri yapabiliyor olmanızı beklemiyorum ancak testleri yapmayı öğrendiğimizde, problemi nasıl kuracağımıza dair genel bir fikriniz oluşmuş olmalı.

## Anahtar Terimler

Örnekleme hatası (Giriş)
Hipotez test etme (4.1)
Örneklem dağılımları (4.2)
Standart hata (4.2)
Ortalamanın örneklem dağılımı (4.2)
Araştırma hipotezi (4.3)
Sıfır hipotezi ($H_0$) (4.3)
Alternatif hipotez ($H_1$) (4.4)
Örneklem istatistikleri (4.5)
Test istatistikleri (4.5)
Karar verme (4.6)
Ret düzeyi (manidarlık düzeyi) (4.6)
Ret bölgesi (4.6)
Kritik değer (4.7)
I. Tip hata (4.7)
$\alpha$ (alfa) (4.7)
II. Tip hata (4.7)
$\beta$ (beta) (4.7)
Güç (4.7)
Tek yönlü test (yönlü test) (4.8)
Çift yönlü test (yönsüz test) (4.8)
Koşullu olasılıklar (4.9)
Etki büyüklüğü (4.11)

## Alıştırmalar

4.1 Dün akşamki Ulusal Hokey Ligi (NHL) hokey maçının 26–13 bittiğini söyledğimi söylesem, muhtemelen haberi yanlış okuduğumu veya başka bir şeylerin maçından bahsettiğimi düşünürdünüz. Bu düşünce aslında bir hipotez test edip reddetme sürecidir.

a. Sıfır hipotezi neydi?

b. Uygulamış olduğunuz hipotez test etme sürecini ana hatlarını ifade ediniz.

4.2 Geçen yıl her bir öğlen yemeği için bir çeyrek dolar eksik veya fazla, yaklaşık olarak 4 $ civarında para harcadım.

a. Günlük harcamanın bir şeklini kabaca çiziniz.

b. Yemek için, hesaba bakmadan 5 $ verip 0,75 $ para üstü alırsam, kazıklandığıma dair endişe duymalı mıyım?

c. (b)'deki cevabınızın mantığını açıklayınız.

4.3 4.2 numaralı alıştırmada I. Tip hata ne olurdu?

4.4 4.2 numaralı alıştırmada II. Tip hata ne olurdu?

4.5 4.2 numaralı alıştırmada ret bölgesi ve kritik değer ne olurdu?

4.6 4.2 numaralı alıştırmada neden tek yönlü bir test kullanmalıyım? Dağılımın hangi ucunu seçmeliyim? Yanlış ucu seçersem ne olur?

4.7 Büyük bir devlet üniversitesine son dönemde kabul edilen lisansüstü öğrencilerin GRE puanları ortalaması 650, standart sapması ise 50'dir. (Puanlar yaklaşık normal dağılmaktadır). Annesi, mütevelli heyeti üyelerinden olan bir aday 490 puan ile kabul edilmiştir. Skandallara bayılan yerel gazete editörü adam kayırmayla ilgili bir eleştiri yazısı kaleme almalı mıdır?

4.8 4.7 numaralı alıştırmada o kadar düşük bir standart sapma neden makuldür?

4.9 4.7 numaralı alıştırmadaki kabul edilmiş öğrenciler örnekleminin normal dağılmasının (veya dağılmamasının) sebebi ne olabilir?

4.10 Bir evrenin tepedeğerinin belirli bir değere (mesela 100'e) eşit olup olmadığını test etmekte kullanılan Tepedeğer Testi adında yeni bir istatistiksel test icat ettiğinizi hayal edin. M istatistiği aşağıdaki gibi hesaplanıyor olsun:

$$M = \frac{\text{Örneklem tepedeğeri}}{\text{Örneklem açıklığı}}$$

M'nin örneklem dağılımını nasıl elde edebileceğinizi açıklayınız.

4.11 4.10 numaralı alıştırmadaki M bu bölümde hangi terim ile adlandırılır?

4.12 Rutin olarak hipotez testi yaptığımız ancak farkında olmadığımız bir günlük yaşam durumu tanımlayınız.

4.13 4.7 numaralı alıştırmada alternatif hipotez ($H_1$) ne olur?

4.14 "Örneklem hatası"nı tanımlayınız.

4.15 "Dağılım" ile "örneklem dağılımı" arasındaki farkı açıklayınız.

4.16 $\alpha$'yı küçültmek Tablo 4.1'deki değerleri nasıl etkilerdi?

4.17 İki araştırma hipotezi örneği verip bunlara karşılık gelen sıfır hipotezlerini ifade ediniz.

4.18 Şekil 4.3'teki dağılım için II. Tip hatanın $\beta = 0{,}74$ olduğunu söylemiştim. Bu değerin nasıl elde edildiğini gösteriniz.

4.19 4.18 numaralı alıştırmadaki hesaplamayı $\alpha = 0{,}01$ için tekrar ediniz.

4.20 Kesim 4.11'deki örnekte çift yönlü bir test yapsaydık test nasıl farklılık gösterirdi?

4.21 Bu bölümde ders değerlendirme örneğindeki adımları açıklayınız. Diğer bir deyişle, ders değerlendirme sonuçlarıyla puanlar arasında gerçek bir ilişki olup olmadığını nasıl belirlersiniz?

4.22 Araç sürücülerinin düşük sosyal statüdeki araçlara karşı daha kaba oldukları hipotezini nasıl test edersiniz?

## Tartışma Soruları

4.23 1. Bölüm'de dördüncü sınıf öğrencilerinin harçlıklarıyla ilgili bir örnek vermiştik. 2. Bölüm'de de böyle bir çalışmada görülebilecek türden bir veri kümesi üretmiştiniz.

a. Erkeklerin daha fazla harçlık aldığına dair hipotezi nasıl test edersiniz? Sıfır hipotezi nedir?

b. Tek yönlü bir test mi, çift yönlü bir test mi kullanırsınız?

c. Hangi sonuçlar karşısında sıfır hipotezini reddedersiniz? Hangi sonuçlar karşısında reddetmezsiniz?

d. Bu çalışmayı daha ikna edici yapmak için tek bir şey yapacak olsanız, ne yaparsınız?

4.24 Simon ve Bruce (1991), yeniden örnekleme istatistiklerini örnekle açıklamak amacıyla 1961 yılında likörün devlet tarafından satıldığı (tekel) 16 eyalet ile özel sektörce satıldığı 26 eyaletteki likör fiyatlarının farklı olduğu hipotezini test etmiştir. (Ortalamalar sırasıyla 4,35 $ ve 4,84 $'dır. Enflasyonun etkilerini görüyorsunuz.) Teknik nedenlerle bu kategorilere uymayan eyaletler çalışmaya dahil edilmemiştir.

a. Esasen test ettiğimiz sıfır hipotezi nedir?

b. 4,35 $ ve 4,84 $ için hangi terimi kullanırsınız?

c. Eğer bunlar ilgilendiğimiz eyaletlerin tamamıysa, en başından neden bir hipotez test ediyoruz?

d. Burada, hipotez test etmenin mantıklı olacağı bir durum düşünebiliyor musunuz?

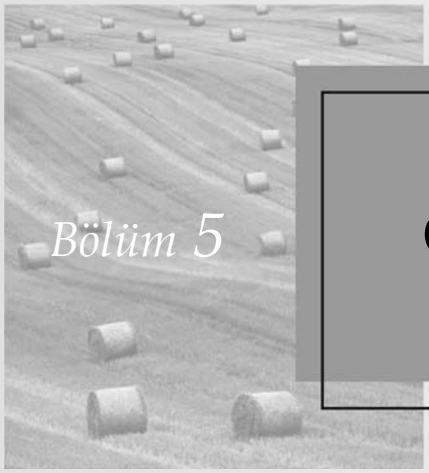
4.25 Park yeri örneğinin içerdiği soruya geleneksel yaklaşım ile Jones ve Tukey yaklaşımının farklı çözümleri nasıl olur, irdeleyiniz.

4.26 Etki büyüklüğünün rapor edilmesi önerisinin gelecekteki psikoloji çalışmalarından elde edilecek sonuçlar üzerindeki etkisi neler olabilir?

4.27 Kadınlardaki depresyon oranı, erkeklerdekinin yaklaşık iki katıdır. Lisansüstü bir öğrenci, bölgedeki psikologlardan 100 kişilik bir depresyon vakaları örneklemi elde etmiş ve bu kişilerin 61 tanesinin kadın olduğunu görmüştür. Tekrar edilen örneklem alma işlemlerinde verinin nasıl görünebileceğini modelleyebilirsiniz. Bir depresyon vakasının kadın olması olasılığının 0,50 olduğu 100'er kişilik 1000 tane örneklem oluşturabilirsiniz. Bunu yapmak için SPSS'te 1000 satırlık boş bir veri üreterek başlamak gerekli. (Bunu yapmak biraz zor olabilir. O yüzden http://www.uvm.edu/~dhowell/methods8/DataFiles/Ex4-27.sav adresinden bu şekilde hazırlanmış boş dosyayı indirebilirsiniz.) Ardından Transform/Compute menüsünü kullanarak kadın sayısı = RV.BINOM(100,.5) değişkenini oluşturun. Her bir denemede kadin sayısı değişkeninin değeri, 100 kişilik her bir örneklemdeki kadın vakaların sayısını verecektir.

a. Eğer $p$(kadın) = 0,50 olsaydı 61 olası bir sonuç olarak kabul edilir miydi?

b. "Depresyon vakalarının %75'i kadındır." hipotezini nasıl test edersiniz? Bu hesaplamaları yapmakta kullanılabilecek bir R programı kitabın Web sayfasında mevcuttur.



*Bölüm 5*

# Olasılığın Temel Kavramları

## Hedefler

Olasılık kavramına ilişkin bir bakış açısı geliştirilmesi, olasılık ile ilgili işlemlerde kullanılan temel kuralların sunulması, Bayes teoreminin arkasındaki temel fikirlerin ana hatlarıyla ortaya konulması, binom dağılımının ve bu dağılımın hipotez testindeki rolünün tanıtılması.

## Konular


**5.1** Olasılık
**5.2** Temel Terimler ve Kurallar
**5.3** Sürekli ve Kesikli Değişkenler
**5.4** Kesikli Değişkenler İçin Olasılık Dağılımı
**5.5** Sürekli Değişkenler İçin Olasılık Dağılımı
**5.6** Permütasyonlar ve Kombinasyonlar
**5.7** Bayes Teoremi
**5.8** Binom Dağılımı
**5.9** Hipotez Test Etmede Binom Dağılımının Kullanılması
**5.10** Çok Terimli Dağılım

BÖLÜM 3'te olasılık kavramını kullanmaya başladık. Örneğin; çocukların yaklaşık %19'unda davranış problemi puanlarının 52 ile 56 arasında değiştiği ve bundan dolayı tesadüfî olarak bir çocuk seçildiğinde, bu çocuğun puanının 52 ile 56 arasında olması olasılığının ,19 olduğunu görmüştük. Bölüm 6'da vardamsal istatistik kavramına odaklanacağımızdan, olasılık ifadeleri üzerinde oldukça yoğun bir biçimde durulacaktır. Orada, "Eğer hipotez doğruysa, gerçekte elde edilen değer kadar uç bir değer elde edilmesi olasılığı ,015'tir." şeklindeki ifade formlarını kullanacağız. Eğer olasılık ifadelerine güvenilecekse, olasılık kavramıyla ne anlatılmak istendiğinin, olasılıkla ilgili işlemler ve hesaplamalara ilişkin birkaç temel kuralın anlaşılması oldukça önemlidir. Bu bölümün amacı budur.

Bu bölümde sunulan içeriğin seçilmesinin iki nedeni vardır. İlk olarak, bu bölümde sunulan materyal, kitabın geri kalanında yer verilen konulara doğrudan uygulanabilen kavramlar olduğundan, kitabın geri kalanının anlaşılabilmesi için gereklidir. İkinci olarak; bu bölüm sayesinde okuyucu için yararlı olabilecek basit olasılık hesaplamalarının nasıl yapılacağının gösterilebileceği düşünülmüştür. Bölümde, bu iki amaçtan herhangi birine hizmet etmeyecek bir içeriğe yer verilmemiştir. Örneğin; dört tane kupa içeren 14 iskambil kâğıdı içerisinden seçilen bir kartın kupa olma olasılığının hesaplanması veya masa lâmbasının kullanılma süresi 250 saat olan bir öğrencinin masasındaki lâmbanın gelecek 25 saat içerisinde yanma olasılığının belirlenmesi ile ilgilenilmeyecektir. Bu tür konular ile ilgilenen öğrencilerin, söz konusu konulara yer verilecek bir olasılık teorisi dersi alması önerilebilir.

## 5.1 Olasılık

analitik bakış

Olasılık kavramına farklı bakış açılarıyla yaklaşılabilir. Olasılık kelimesi ile ne anlatılmak istendiğine dair genel bir uzlaşı yoktur. Olasılığa ilişkin en eski ve belki de en yaygın tanım **analitik bakış**tır. Olasılık tartışmalarına ilişkin en yaygın örneklerden biri bu kitapta da tercih edilen M&M şekerleridir. Neredeyse herkes M&M'ye aşina olduğundan, M&M'nin iyi bir örnek olduğu söylenebilir. Bu şekerler yapışkan bir halde olmadığından, sınıfta gösterilmesi de kolaydır. Ayrıca, örnek olarak gösterilen şekerlerden biri yenilebilir. Mars Şeker Şirketi, bu şekerlerin örnek olarak kullanılmasına gönüllü olup her bir şeker çantasına çantanın içindeki farklı renk şekerlerin yüzdesini yazmaktadır. Burada sütlü çikolata sürümüne ilişkin veriye yer verilmiş olup, bu veri Tablo 5.1'de gösterilmiştir.

Önümüzde bir M&M çantası olduğunu ve bu çantanın içindeki çikolatalardan birini çektiğimizi varsayalım. Bu tür bir kabul gerekli olmasa da; izlenen işlemleri kolaylaştırmak için M&M çantasında 100 tane çikolata olduğunu kabul edelim. Mavi bir M&M çekilmiş olma olasılığı nedir? Olasılık hakkında daha fazla hiçbir şey bilmeden bu soru cevaplanabilir. Çantadaki çikolataların %24'ü mavi olduğundan ve çikolatalardan biri tesadüfî olarak çekildiğinden seçilen çikolatanın mavi olma olasılığı ,24'tür. Bu örnek, olasılığa ilişkin tanımlardan birini gösterir.

**Tablo 5.1** Ortalama bir M&M çantasındaki renk dağılımı

| Renk | Yüzde |
|---|---|
| Kahverengi | 13 |
| Kırmızı | 13 |
| Sarı | 14 |
| Yeşil | 16 |
| Turuncu | 20 |
| Mavi | 24 |
| Toplam | 100 |

Eğer bir olay *A* durumunda gerçekleşiyor ancak *B* durumunda gerçekleşmiyorsa ve muhtemel tüm durumların gerçekleşme olasılığı eşit ise (yani, M&M çantasındaki her bir çikolatanın çekilme olasılığı eşitse); olayın gerçekleşme olasılığı *A*/(*A*+*B*) değerine; gerçekleşmeme olasılığı ise *B*/(*A*+*B*)'dir.

Çünkü 24 tane mavi (çantadaki 100 çikolatadan 24 mavi renkli çikolatanın her biri için bir) ve 76 tane başka bir renkten çikolata çekme yolu vardır; A = 24 , B = 76, $p$(A) = 24/(24 + 76) = 24.

Olasılığa alternatif bir bakış da, **frekansçı yaklaşım**dır. Torbadan her bir çekilişte seçilen renk not edilerek çantadan M&M şekerleri çekildiğini varsayalım. Bu örneklem çalışması yürütülürken; **yerine koyarak örnekleme** yapılmaktadır. Yerine koyarak örnekleme, seçilen bir M&M'nin bir sonraki seçimden önce tekrar çantaya konulması anlamına gelir. Bu şekilde çok sayıda çekme işlemi yapıldığında, seçilen M&M'lerin %24 gibi bir oranda mavi renkli olduğu görülecektir. Buna göre olasılığı, gerçekleşmesi istenen olayın bağıl frekansının, seçme sayısı arttıkça limiti[1] şeklinde tanımlayabiliriz. (Örneklem büyüklüğü artıkça olasılığın $p$'ye nasıl yaklaştığını gösteren R'de yazılmış bir programa bu kitabın internet sayfasından erişilebilir.)

**öznel olasılık (subjective probability)**

Olasılıkla ilgili üçüncü kavram birçok teorisyen tarafından savunulmaktadır. Bu **öznel olasılık (subjective probability)** kavramıdır. Buna göre olasılık tanımı, bir olayın gerçekleşmesinde, bir bireyin öznel inançlarını temsil eder. Örneğin; "Yarın güzel bir gün olacak." ifadesi, inancın subjektif bir değeridir. Uzun sürede iyi günlerin gerçekleşmesinde bağıl frekans düşük olduğundan, böyle bir olasılık değeri düşüktür ve bu tür ifadenin matematiksel temeli yoktur. Ancak bu açıklama, olasılığa ilişkin söz konusu bakış açısının mantıklı olmadığı şeklinde anlaşılmamalıdır. Öznel olasılık, bireylerin subjektif karar alma süreçlerinde son derece önemli bir rol oynamakta ve davranışlarımızın bütün yönlerini etkilemektedir. Sadece, öznel inançlara dayalı olarak alınan kararların sayısını belirli sonuçların olasılığı içinde düşünelim. Akşam yemeği için hazırlanan özel bir et yemeğinin tadından emin olmadığımız için pasta sipariş edebiliriz veya hava tahmincilerinin yarın havanın %80 olasılıkla kar yağışlı olacağı yönündeki açıklamasını dikkate alarak kayağa gitme planı yapabiliriz. Benzer şekilde, at yarışı oynarken bir atın yarışı kazanmasının yarıştaki diğer atların kazanmasına göre altı kat daha fazla olduğunu düşünerek paramızı bu ata yatırabiliriz. Bu noktada öznel olasılığın kullanımı için gerekli olan Bayes teoremini kısaca ele alacağız. Bölümün ilerleyen kesimlerinde yapıldığı gibi, istatistiksel kararlar genellikle frekansçı ya da analitik yaklaşımla başlar. Bununla birlikte, bu olasılıkların yorumlanması güçlü bir öznel bileşen içerir. Son zamanlarda olasılıklara daha sübjektif bakmaya doğru bir yönelim olduğu görülmektedir. Sonraki bölümde güven aralığından söz edilirken bu konuya değineceğiz. Bayes istatistiği ayrı bir alan olmakla birlikte, bu istatistiğin bileşenleri davranış bilimlerindeki analizlere dâhil edilmektedir. İyi bir şey olduğu düşüncesindeyim.

Tercih edilecek tanım her birimiz için önemli olmasına rağmen, hangi tanım seçilmiş olursa olsun hipotez testi açısından sonuç aynı olmaktadır. Bu kitabın geri kalan kısmı

---

[1] Limit kelimesi, çekilen M&M sayısı arttıkça mavi renkli M&M oranının giderek bir değere yaklaşacağı anlamına gelmektedir. Örneğin; mavi renkli M&M oranı 100 çekimden sonra .23, 1000 çekimden sonra ,242 ve 10.000 çekimden sonra ,2398… olabilir. Çekilen M&M sayısı arttıkça elde edilen değerlerin giderek $p$ = ,240000'e yaklaştığına dikkat edilmelidir. Bu şekilde yaklaşan değer limit olarak adlandırılmaktadır.

boyunca, hangi tanımın işe koşulacağına ilişkin tartışmalar bulunmaktadır. (Öznel olasılık tanımını tercih eden kişilerin çoğunlukla genel hipotez testi yönelimine katılmadıklarının söylenmesi gerekir.) Aslında birçok kişi, farklı yaklaşımları birbirinin yerine kullanmaktadır. Rus ruletini kaybetme olasılığının 1/6 olduğunu söylediğimizde, tabancanın altı silindirinden yalnızca birinin içinde mermi bulunduğunu kastetmiş oluruz. Tüketici raporlarında belirtilen arıza kayıtlarını göz önünde bulundurarak bir otomobili satın alırken, satın almaya karar verdiğimiz modelin büyük bir kısmının nispeten sorunsuz olduğunu düşünerek hareket ederiz. Şampiyonluğu Colorado Rockies'in kazanma olasılığından söz ederken, bu olaya ilişkin öznel inancımızı dile getirmiş oluruz. Elde edilen veri kümesindeki olasılık çok düşük olduğundan hipotezin reddedilmesine rağmen, hipotezin doğru olması halinde olasılığa ilişkin hangi bakış açısının kullanıldığının bir önemi yoktur.

## 5.2 Temel Terimler ve Kurallar

**olay**

Olasılık teorisinde verinin temel parçası **olay** olarak adlandırılır. **Olay** kelimesi, istatistikçilerin hemen her şeyi kapsayacak şekilde kullandıkları bir terimdir. Bir iskambil destesinden bir kart çekildiğinde, çekilen kartın papaz olması, yetenek testinden 36 puan alınması, Yüksek Mahkeme'ye atanacak bir sonraki kişi için "bayan" sınıflaması veya bir örneklemin ortalaması birer olay olabilir. "Bir şeyin" olasılığından söz edildiğinde, burada bir şey ile *olay* kastedilmektedir. Bir paranın atılması gibi bir süreçle ilgilenilmesi durumunda olay, para atma işleminin sonucu olup yazı ya da turadır. Torbadan M&M çekilmesi halinde, muhtemel olaylar 6 olası renktir. Bir dersten alınabilecek notlar hakkında konuşurken olası olaylar; A, B, C, D ve F'dir.

**bağımsız olay**

İki olaydan birinin gerçekleşmesi ya da gerçekleşmemesi diğer olayı etkilemiyorsa, bu iki olay **bağımsız olaylar** ifadesiyle adlandırılır. Örneğin gizli oylamalarda, tesadüfî olarak seçilen iki bireyin oy kullanma davranışı bağımsız olaylardır. Çünkü bir kişinin kullandığı oy bir başkasının kullandığı oy üzerinde etkili olmamaktadır. Bununla birlikte, aynı ailenin iki farklı üyesinin oy kullanma davranışları birbirinden bağımsız olmayabilir. Çünkü aynı ailenin fertleri benzer tutum ve inançlar paylaşabilmektedir. Buna rağmen, aynı aileden iki bireyin birbirlerinin kullandıkları oyları görmemeleri için dikkatli davranılması durumunda, olayların bağımsızlığı bu kişiler için de geçerli olabilir.

**ayrık (mutually exclusive) olaylar**

Eğer bir olayın olması diğer olayın olması ile eş zamanlı olmuyorasa, bu iki olaya **ayrık (mutually exclusive) olaylar** denir. Örneğin, bir öğrenci aynı anda birden fazla sınıfa devam edemeyeceği için bir öğrencinin lisans programının birinci, ikinci, üçüncü veya dördüncü sınıfında olması ayrık olaydır. Bir olaylar kümesi muhtemel tüm olayları kapsıyorsa bu olaylara evrensel denir. Önceki örnekteki dört kolej sınıfı bir lisans öncesi öğrencisi için **evrensel** olaylardır; çünkü herhangi bir öğrenci yukarıda belirtilen sınıfların (kategorilerin) birine düşecektir. Aynı zamanda, üniversiteye kayıtlı tüm öğrenciler düşünüldüğünde; öğrenciler arasında lisansüstü öğrenciler, tıp fakültesi öğrencileri veya özel statülü öğrenciler bulunabileceğinden, bu dört sınıf düzeyi evrensel olay olmayacaktır.

**evrensel**

Bilindiği veya olasılık tanımlarından anlaşılacağı gibi; olasılık 0,00 ile 1,00 arasında değişen değerler alır. Bir olayın olasılığı 1,00 ise bu olayın gerçekleşmesi kesindir. (Bu paragrafın sonuna ulaşana kadar yazımına devam edilecekler de dahil olmak üzere, çok az olaya ilişkin olasılık değeri 1,00'dır.) Bir olaya ilişkin olasılık değerinin 0,00 olması ise, bu olayın gerçekleşmeyeceğinin kesin, diğer bir ifadeyle gerçekleşmesinin imkânsız olduğunu gösterir. Buna göre, 1,00'a yakın olasılık değerlerine sahip olayların gerçekleşmesi daha çok muhtemel iken 0,00'a yakın olasılık değerine sahip olayların gerçekleşmemesi daha muhtemel bir durumdur.

### Temel Olasılık Kanunları

Olasılıkla ilgili tartışmaların merkezinde iki önemli teorem vardır. (Eğer burada kullandığım *teorem* ifadesi rahatsız edici olursa, teorem yerine *kural* kelimesi kullanılabilir.) Bunlar toplama kuralı ve çarpım kuralıdır.

## Toplama Kuralı

**olasılığın toplama kuralı (additive law of probability)**

Toplama kuralını açıklamak için M&M örneğinden yararlanacağız. Bu örnekteki altı rengin tümünü düşünelim. Tablo 5.1'deki veri kümesi olasılığın analitik tanımı ile ele alındığında, $p$(mavi) = 24/100 = ,24, $p$(yeşil) = 16/100 = ,16 vb. olur. Burada, çekilen M&M'nin mavi ya da yeşil renkli olma olasılığı nedir? Bu noktada **olasılığın toplama kuralına (additive law of probability)** ihtiyaç duyarız.

> Bir dizi ayrık olay dikkate alındığında, bu olaylardan birinin ya da diğerinin gerçekleşme olasılığı; bu iki olay için ayrı ayrı elde edilen olasılıkların toplamına eşittir.

Buna göre, $p$(mavi ya da yeşil) = $p$(mavi) + $p$(yeşil) = ,24 + ,16 = ,40 olur. Burada iki olayın ayrık olaylar olması gerektiği şeklinde bir sınırlama getirildiğine dikkat edilmelidir. Yani toplam kuralının uygulanabilmesi için bir olayın gerçekleşmesi diğer olayın gerçekleşmemesi anlamına gelmelidir. Eğer çekilen bir M&M mavi renkli ise, yeşil renkli olamaz. Bu gereklilik son derece önemlidir. Bu ülkedeki nüfusun yaklaşık yarısı bayandır ve ülke nüfusunun yaklaşık yarısının boyu 5'6¾"ten daha uzundur.[2] Ancak tesadüfî olarak seçilen bir kişinin bayan ya da 5'6¾"ten uzun olması olasılığının ,50 + ,50 = 1,00 olmayacağı yeterince açıktır. Buradaki iki olay ayrık değildir. Bununla birlikte, 1987 yılında Vermont'ta doğan bir kız çocuğunun adının o yıl doğan çocuklara verilen en yaygın iki isim arasında yer alan Ashley ya da Sarah olması olasılığı toplam kuralı ile hesaplanabilir. $p$(Ashley) + $p$(Sarah) = ,010 + ,009 = ,019 bulunur. Bu örnekte, bir kız çocuğunun isminin Ashley ya da Sarah olması ayrık olaylardır. Çünkü ailelerden biri iki ismi birleştirmek gibi bir fikre kapılmadığı sürece, bir kız çocuğuna aynı anda ilk isim olarak Ashley ve Sarah adı verilemeyecektir.

## Çarpma Kuralı

**olasılığın çarpım kuralı (multiplicative law of probability)**

$p$(mavi) =.24, $p$(yeşil) =.16 ve $p$(diğer) =.60 olan M&M örneği üzerinden devam edelim. Birinciyi çektikten sonra çekileni yerine koyarak ikinciyi çektiğimi varsayalım. İlk çekilenin ve ikinci çekilenin mavi olması olasılığı nedir? Bu olasılığın hesaplanabilmesi için **olasılığın çarpım kuralı**na **(multiplicative law of probability)** ihtiyaç duyarız.

> İki ya da daha fazla bağımsız olayın birlikte gerçekleşmesi olasılığı, bu iki olayın olasılıklarının çarpımına eşittir.

Buna göre, $p$(mavi, mavi) = $p$(mavi) × $p$(mavi) = ,24 × ,24 = ,0576 bulunur. Benzer şekilde; ilk çekilen M&M'nin mavi ve ikinci çekilen M&M'nin yeşil olması olasılığı $p$(mavi, yeşil) = $p$(mavi) × $p$(yeşil) = ,24 × ,16 = ,0384 elde edilir. Burada çarpım kuralını bağımsız olaylar ile sınırlandırdığımıza dikkat edilmelidir. Yani olaylardan birinin gerçekleşmesi diğer olayın gerçekleşme olasılığı üzerinde herhangi bir etkiye sahip olmamalıdır. Cinsiyet ve boy uzunluğu birbirinden bağımsız olaylar olmadığından; $p$(kadın ≤ 5' 6¾") = ,50 × ,50 = ,25 gibi bir ifade yanlış olacaktır. Bununla birlikte, $p$(kadın, ocak ayında doğmuş olma) = ,50 × 1/12 = ,50 × ,083 = ,042 olasılığı büyük ihtimalle doğru olacaktır. Çünkü bireyin cinsiyeti ile doğduğu ayın birbirine bağımlı olduğuna ilişkin herhangi bir veriye rastlanmamıştır. (Eğer bireyin cinsiyeti ve doğduğu ay birbiri ile ilişkili olsaydı, buradaki hesaplama da yanlış olacaktı.)

Bölüm 6'da iki değişkenin bağımsızlığı hakkındaki soruları cevaplandırırken, çarpım kuralını kullanacağız. Bölüm 6'dan bir örnek bu kuralın kullanılmasını göstermeye yardımcı olacaktır. Bölüm 6'da ele alınacak bir çalışmada, Unah ve Borger (2001) bir sanığın mahkemede idam cezası alıp almamasının sanığının rengi ile ilişkili olup olmadığı hipotezini test etmek istemektedir. Araştırma kapsamında sanıklar beyazlar ya da siyahiler şeklinde iki gruba ayrılmışlardır. Bu hipotezin testi, bir sanığın siyahi olması ve ölüm cezası alması olasılığının hesaplanmasını içerir. *İki olayın (renk ve idam cezası) birbirinden bağımsız olması halinde* bu olasılığın hangi değere eşit olduğunun hesaplanması gerekir. İki olayın birbirinden bağımsız olduğunun ortaya konulması, kararı alan hâkimlerin, sanığın renginden etkilenmediklerini yansıtır. Eğer bu iki olayın bağımsız olduğu varsayılırsa,

2 http://en.wikipedia.org/wiki/Human_height

olasılığa ilişkin çarpım kuralından, $p$(siyahi, idam) = $p$(siyahi) × $p$(idam) olur. Belirtilen çalışmada, sanıkların %34,4'ü siyahtır. Dolayısıyla tesadüfi olarak seçilen bir sanığın siyah olması olasılığı ,344'tür. İdam cezası alan sanıkların oranı ise %8'dir. Buna göre bir sanığın idam cezası alması olasılığı ,08'e eşittir. Dolayısıyla eğer iki olay birbirinden bağımsız ise, $p$(siyahi, idam) = $p$(siyahi) × $p$(idam) = ,344 × ,08 = ,028 = %2,8 bulunur. (Gerçekte, siyahi olan ve idam cezası alan sanıkların oranı %4'tür. Bu değer cezalara ilişkin kararın sanığın renginden bağımsız olması durumunda beklenen değerden yüksektir. Eğer bu fark güvenilir ise; böylesi bir fark, cezası ölüm olan suçlarla ilgili ne gibi bir bilgi verir?)

Son olarak; hem toplama hem de çarpma kurallarını içeren basit bir örnek ele alabiliriz. Eğer torbadan ilk çekileni yerine koyarak iki M&M çekilirse; birinin mavi, diğerinin yeşil renkli olması olasılığı ne olur? (Burada sıra önemli değildir. İlk M&M mavi ve ikincisi yeşil veya ilk M&M yeşil ve ikinci M&M mavi olabilir.) İlk olarak çarpım kuralı kullanılır.

$p$(mavi, yeşil) = $p$(mavi) x $p$(yeşil) = ,24 x ,16 = ,0384
$p$(yeşil, mavi) = $p$(yeşil) x $p$(mavi) = ,16 x ,24 = ,0384

Her iki sonuç da sorudaki gerekliliği karşıladığından, bu iki durumundan herhangi birinin gerçekleşmesi olasılığının hesaplanması gerekir. Bunun için toplama kuralına başvurulur:

$p$(mavi, yeşil) + $p$(yeşil, mavi) = ,0384 + ,0384 = ,0768

Buna göre, birinci çekilen yerine konularak iki tane M&M çekilmesi durumunda, bunlardan birinin mavi diğerinin yeşil olması olasılığı yaklaşık ,08'dir - yani, 1/10'dan daha düşük bir oranda meydana gelecektir.

Öğrenciler hızlı bir şekilde duyunca, her iki ifade neredeyse aynı şeyi anlatıyor gibi gelebildiğinden, bazen toplama ve çarpma kuralı ile ilgili karışıklık yaşayabilmektedir. Bu nedenle, iki kuralın uygulandığı durumlar arasındaki farkın anlaşılması önem arz eder. Toplama kuralının uygulandığı durumlarda tek bir sonucun elde edileceği bilinmelidir. Örneğin, çekilen M&M mavi ya da yeşil renkli olabilir. Ancak bu iki durumdan yalnızca biri söz konusu olacaktır. Çarpım kuralının uygulandığı durumlarda, en az iki sonuçtan söz edilmektedir (örneğin; elde edilen olasılıkta bir mavi ve bir de yeşil renkli M&M olacaktır). Tek bir sonuç için toplama kuralı kullanırken; çok sayıda bağımsız sonuç olduğunda çarpım kuralına başvururuz.

## Yerine Koyarak Örnekleme

"Yerine Koyarak Örnekleme" kullanmayı neden sürdürüyorum? Bu soruya verilecek cevap olayların bağımsızlığı konusuna dayanır. Mavi ve yeşil renkli M&M örneğini düşünelim. 100 M&M'nin yer aldığı çantada, 24 tane mavi ve 16 tane de yeşil renkli M&M'miz var. İlk denemede çekilen M&M'nin mavi renkli olması olasılığı 24/100 = ,24'tür. İkinci çekimden önce, ilk seçilen M&M'yi, yerine koyarsam, mavi renkli M&M'lerin diğer M&M'lere oranı yine 24/76 olacak ve ikinci çekimde de mavi renkli bir M&M seçilmesi olasılığı 24/100 = ,24 olacaktır. Ancak ilk seçilen M&M ikinci çekimden önce torbaya konulmazsa, ikinci çekimde mavi renkli bir M&M çekilmesi olasılığı ilk çekimin sonucuna bağlı olacaktır. Eğer ilk denemede çektiğim M&M mavi renkli olursa, geriye 23 mavi ve 76 tane de diğer renklerde M&M kalacaktır. Dolayısıyla ikinci çekimde $p$(mavi) = 23/99 = ,2323 olacaktır. Eğer ilk denemede çektiğim M&M yeşil olursa ikinci denemede $p$(mavi) = 24/99 = ,2424 olacaktır. Yerine koyarak örnekleme yapıldığında, $p$(mavi) olasılığı bir denemeden diğerine değişmezken, yerine koymadan örnekleme yapılması durumunda $p$(-mavi) olasılığı bir denemeden diğerine farklılık gösterecektir. Uç bir örnek olarak, eğer yerine koymadan çekim yaparsam, 60 çekimden 25'inin mavi olma olasılığı nedir? Çantada 24 tane mavi renkli M&M bulunduğundan yerine koymadan yapılan örneklemede 25 tane mavi renkte M&M çekilmesi olasılığı ,00 olup imkânsız olaydır. Bununla birlikte; yerine koyarak örnekleme yapıldığında, gerçekleşmesi mümkün olan bir durum söz konusu olur ve çekilen 25 M&M'nin mavi renkli olması olasılığı ,0011 bulunur.

## Birleşik Koşullu Olasılıklar

Olasılıkla ilgili tartışmalarda ortak ve koşullu olasılık olmak üzere iki olasılık türü önemli rol oynar.

**birleşik olasılık**

**Birleşik olasılık,** iki veya daha çok olayın birlikte gerçekleşmesine ait olasılıktır. Örneğin; Unah ve Borger'ın (2001) idam cezasında renk (siyahi ve beyaz) yanlılığını incelediği çalışmada, bir sanığın hem siyahi olması hem de idam cezası alması birleşik olasılıktır. Aynı şekilde, bir sanığın beyaz olması ve idam cezası alması da birleşik olasılıktır. Yukarıda *p*(mavi, yeşil) şeklinde gösterildiği gibi, herhangi iki olayın birleşik olasılığı *p*(*A*, *B*) şeklinde gösterilir. Yukarıda verilen örneklerde görüldüğü gibi, eğer herhangi iki olay birbirinden bağımsızsa, söz konusu iki olayın birleşik olasılığı çarpma kuralı ile hesaplanır. Eğer olaylar bağımsız değilse, ortak olasılığın hesaplanması daha karmaşık olmakta ve olayların bağımsız olması durumunda hesaplanan olasılıktan farklılık gösterir. Olayların bağımsız olmaması durumundaki birleşik olasılığını burada hesaplamayacağız.

**koşullu olasılık**

**Koşullu olasılık**, bir olayın gerçekleştiği biliniyorken bir başka olayın gerçekleşmesi olasılığıdır. Örneğin, uyuşturucu kullanıcısı olduğu bilinen birinin damardan AIDS virüsü almış olması olasılığı koşullu olasılıktır. Okuyucunun çöpe atmaması mesajını içerdiği biliniyorken; bir mağazanın reklam için hazırlattığı el ilanının çöpe atılması olasılığının ne olacağı koşullu olasılık ile ilgili bir başka örnektir. Bu kitapta sıkça tekrar edilen bir cümle olan "Eğer yokluk hipotezi doğru ise, ……. gibi bir sonuç elde edilmesi olasılığı nedir?" ifadesi koşullu olasılığa ilişkin bir başka örnek olabilir. Burada "biliniyorken" kelimesi eğer kelimesi ile yer değiştirmiş olsa da; anlamda bir farklılaşma yoktur. Bu kitapta "koşullu" kelimesinin kullanımı ile sıklıkla karşılaşılacaktır. Koşullu olasılık; "Eğer … doğru ise, …" şeklinde okunabilir. Örneğin; öğrencilerin hatırladıkları kelime sayısındaki farklılıklar, kelime sunulduğunda sözel işleme düzeyinde koşullu bir olasılıktır.

A ve B ile adlandırılan iki olay için, B olayının gerçekleştiği biliniyorken A olayının olasılığı, $p(A \mid B)$ şeklinde gösterilir ve A ile B arasına dikey bir çizgi konur. Örneğin yukarıdaki örnekler için $p(\text{AIDS} \mid \text{uyuşturucu kullanımı})$ ve $p(\text{idam} \mid \text{siyahi})$ gösterimleri kullanılır.

Çoğu zaman çocuk sahibi olan kişilerde sorumluluk duygularının daha yüksek olduğunu kabul ederiz. Dikkatsiz ve sorumsuz davranışlarla yıllarını geçiren birçok kişinin anne, baba olduktan sonra farklı insanlara dönüştükleri görülür. Anne, baba olan birçok insanın davranışlarında önemli değişiklikler gözlenir. (Bu değişikliğin gözlenebilmesi için sadece birkaç yıl beklenmelidir.) Bir radyo istasyonunun 100 kişiden oluşan bir örneklem seçtiğini ve bunlardan 20'sinin çocuk sahibi olduğunu varsayalım. Çalışmada, emniyet kemeri kullanan 30 kişiden 15'inin çocuk sahibi olduğu belirlenmiştir. Çalışmaya ilişkin veri Tablo 5.2'de gösterilmiştir.

Tablo 5.2'deki veri; basit, birleşik ve koşullu olasılık değerlerini hesaplama olanağı sunmaktadır. Tesadüfî olarak seçilen bir kişinin emniyet kemeri takıyor olması olasılığı 30/100 = ,30 olup bu hesaplama basit olasılığa bir örnektir. Bir kişinin emniyet kemeri takan ve çocuğu olan bir birey olması olasılığı 15/100 = ,15 olup bu hesaplama birleşik olasılığa örnek teşkil eder. Tesadüfî olarak seçilen bir kişinin çocuk sahibi olduğu biliniyorsa, bu kişinin emniyet kemeri takıyor olması olasılığı 15/20 = ,75'tir ve bu hesaplama bir koşullu olasılık örneğidir. Birleşik ve koşullu olasılık birbiri ile karıştırılmamalıdır. Görüldüğü gibi, bu iki olasılık türü birbirinden farklıdır. Burada birleşik olasılık hesaplanırken, neden ilgili basit olasılık değerlerinin çarpılması gibi bir hesaplamaya başvurulmadığı merak edilebilir. Çarpma kuralının kullanılabilmesi için bir kişinin çocuğunun olması ile emniyet

**Tablo 5.2** Ebeveynlik ile emniyet kemeri kullanımı arasındaki ilişki

| Ebeveyn Durumu | Emniyet Kemeri Takıyor | Emniyet Kemeri Takmıyor | Toplam |
|---|---|---|---|
| Çocuğu var | 15 | 5 | 20 |
| Çocuğu yok | 15 | 65 | 20 |
| Toplam | 30 | 70 | 100 |

kemeri takıyor olması olasılığı birbirinden bağımsız olaylar olmalıdır. Buradaki örnekte ise kişinin emniyet kemeri takması büyük ölçüde çocuk sahibi olmasına bağlı olduğundan, bu iki olay birbirinden bağımsız değildir. (Eğer iki olay birbirinden bağımsız olsaydı, birleşik olasılık ,30 x ,20 = ,06 gibi gerçek olasılık değerinin yarısından daha az bir değer olacaktı.)

Başka bir örnek vermek gerekirse, alkollü içecek içme ve kaza yapma olasılığı bir birleşik olasılıktır. Bağlı olarak az sayıda insan herhangi bir zamanda alkollü içecek kullandığından ve yine görece az sayıda kişi trafik kazası yaptığından, bu olasılık değerleri çok yüksek değildir. Buna karşın, bir kişinin alkol aldığı biliniyorken kaza yapması olasılığı ya da tam tersi kaza yaptığı biliniyorken alkol almış olması olasılığı çok daha yüksektir. Gece saatlerinde, $p$(alkollü içecek içme | kaza yapma) koşullu olasılığı ,50'ye yakın bir değerdir. Çünkü Amerika Birleşik Devletleri'nde gece kaza yapan otomobil sürücülerinin yaklaşık yarısında alkol bulunmaktadır. $p$(kaza yapma | alkollü içecek içme) koşullu olasılığının değeri bilinmemektedir. Ancak bu olasılık değerinin **koşullu olmayan bir olasılık** $p$(kaza yapma) değerine göre çok daha yüksek olduğunu tahmin etmek zor değildir.

**koşullu olmayan olasılık**

NOT

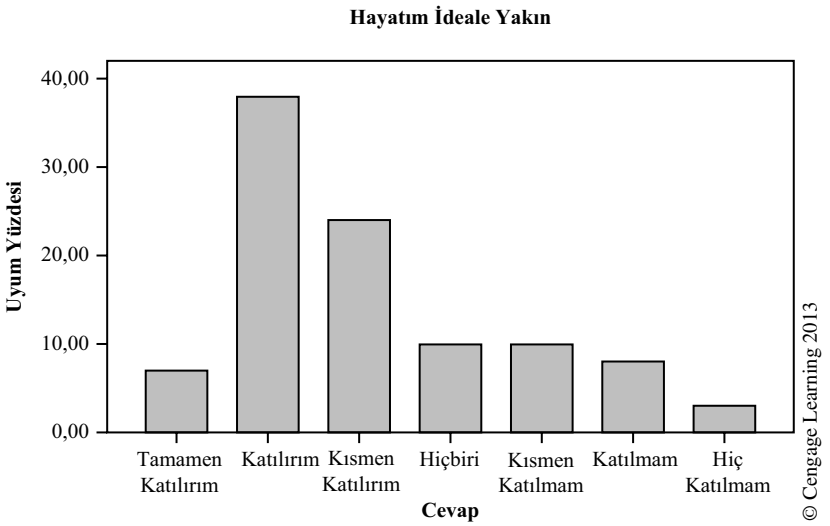
$p(X \mid Y)$ ile gösterilen koşullu olasılık, bir olayın gerçekleştiği biliniyorken, bir diğer olayın gerçekleşmesi olasılığını ifade etmek için kullanılır. Bu kitapta geçtiği her yerde "koşullu" kelimesi bu şekilde kullanılacak olup koşulun ne olduğuna bakılması gerekir.

## 5.3 Sürekli ve Kesikli Değişkenler

Bölüm 1'de kesikli ve sürekli değişken arasında bir ayrım yapılmıştır. Matematikçilerin bakış açısıyla, kesikli bir değişken sınırlı (sayılabilir) sayıda değer alabilirken; sürekli bir değişken sonsuz sayıda farklı değer alabilir. Örneğin, özel bir sinema gecesine katılan kişilerin sayısı kesikli bir değişkendir. Çünkü sinemaya giden insanlar sayılabilir çokluktadır ve kişi sayısı için kesirli bir ifade kullanılması mümkün değildir. Ancak bir kişisel alan çalışmasında iki kişi arasındaki uzaklık sürekli bir değişkendir. Çünkü iki kişi arasındaki uzaklık 20 cm gibi tam bir değer alabileceği gibi 20,8 cm veya 20,8173754814 gibi kesirli bir değer de alabilir. Burada verilen ayrım teknik olarak doğru olmakla birlikte, yaygın kullanım biraz daha farklıdır.

Pratikte kesikli değişkenler hakkında konuşurken, nispeten daha az sayıda olası değer alabilen değişkenler kastedilir (örneğin, sosyo-ekonomik düzeye ilişkin 5'li derecelendirme). Çok sayıda farklı değer alabilen ve değerleri en azından sıralama düzeyindeki ölçekte olan değişkenler ise sürekli değişken olarak kabul edilir. Dolayısıyla, IQ puanı tam sayılı değerler alsa da ve IQ puanı 105,317 gibi kesirli sayı olan bir kimseyi bulamasak da; genel olarak IQ puanı sürekli değişken olarak işlem görür. Bölüm 3'te; normal dağılıma sahip bir değişkenin $\pm\infty$ aralığındaki tüm değerleri alabildiği bilinmesine rağmen, sadece pozitif tam sayı değerlerini alabilen bir değişken olduğunu bildiğim Achenbach Toplam Davranış Puanı'nın normal dağılma sahip olduğunu kabul ettim. Yani çok sayıda farklı değer alabildiği için bu değişken sürekli bir değişken olarak işlem görür. Bu değişken normal dağılıma yakın olduğundan, normal dağılıma sahip bir değişken olarak alındığı bir çalışmanın sonuçları kabul edilebilir düzeyde doğru olacaktır.

Kesikli ve sürekli değişken ayrımı bu bölümde tekrar tanıtılmıştır. Çünkü bu iki değişkene ilişkin dağılımlar olasılık teorisinde farklı şekillerde işlem görür. Kesikli değişkenlerden söz ederken, yalnızca belirli bir olasılık değeri ile sonuçlanacak bir işlemden söz ediyoruz. Diğer taraftan sürekli değişkenlere ilişkin olasılık işlemleri, belirli bir aralığa düşen değerler ile sonuçlandırırız.




**Şekil 5.1** Yaşam kalitesi dağılımı

## 5.4 Kesikli Değişkenler İçin Olasılık Dağılımları

Kesikli olasılık dağılımına ilişkin ilginç bir örnek Şekil 5.1'de sunulmuştur. Şekilde gösterilen veri, İskoç Hükümeti tarafından 2009 yılında çevresel tutuma ilişkin bir çalışma kapsamında toplanmıştır. Çalışmada çevresel sorunların belirlenmesiyle ilgilenilmiştir. Ancak veri toplama sürecinde Illinois Üniversitesi'nden Edward Diener'ın Yaşam Doyumu Ölçeği'nden de yararlanılmıştır.[3] Şekil 5.1'de bu ölçekte yer alan "Birçok yönden hayatım ideale yakındır." maddesine verilen cevapların dağılımı sunulmuştur. $X$'in muhtemel değerleri (dereceler), apsis yani $X$ ekseninde, her bir derecenin seçilme sıklığını gösteren frekanslar (veya oranlar) ordinat yani $Y$ ekseninde gösterilmiştir. Bireylerin yaşamlarına ilişkin memnuniyet düzeyleri şaşırtıcı sonuçlar ortaya koymuştur. Örneklem için oran doğrudan olasılık değerlerine dönüştürülebilir. İlgili dönüştürme uygulanırsa, tesadüfî olarak seçilen bir kişinin "Yaşamım birçok yönden idealdir." maddesine "Katılıyorum." cevabını vermesi olasılığı ,38 bulunur. Ölçekteki diğer üç pozitif kategori de eklendiğinde, bu olasılık değeri ,69'a yükselmektedir. Diğer bir deyişle, tesadüfi olarak seçilen bir kişinin "Yaşamım ideale yakındır." maddesine en azından kısmen katılma olasılığı ,69'dur.

## 5.5 Sürekli Değişkenler İçin Olasılık Dağılımı

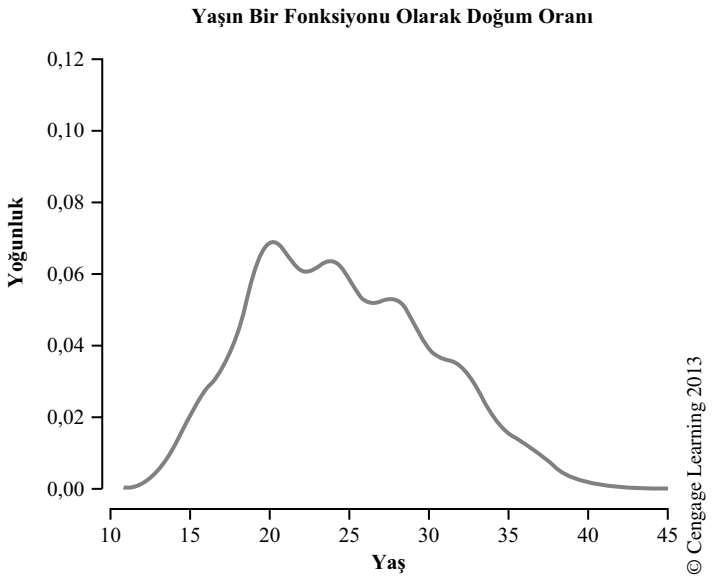
Kesikli olasılık dağılımından sürekli olasılık dağılımına geçtiğimizde, işler biraz karmaşıklaşmaktadır. Bölüm 6'daki normal dağılım düşünüldüğünde, sürekli değişkenlerle ilgilenildiği ve dağılımın ordinatının yoğunluk olarak isimlendirildiğini ve sürekli değişken için olasılık dağılımından elde edilen sonuçların tek bir değer yerine, bir aralığa karşılık geldiğini hatırlayacaksınız. Şimdi, bu nokta üzerinde biraz daha fazla durulması gerekir.

---

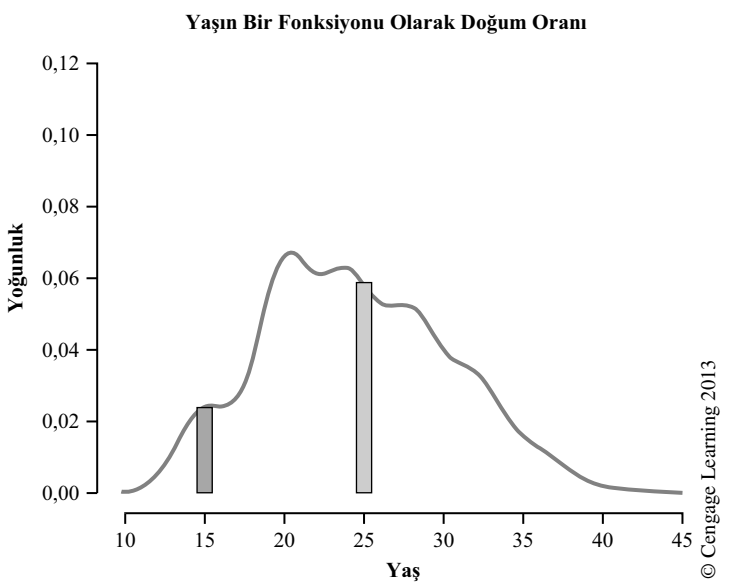
3 Bu ölçek hakkında daha fazla bilgi edinmek için http://internal.psychology.illinois.edu/~ediener/SWLS.html adresine başvurulabilir. Ayrıca, http://www.scotland.gov.uk/Publications/2009/03/25155151/0 adresinden burada kaynak olarak gösterilen çalışmaya erişilebilir.

Şekil 5.2'de ilk doğumda annenin yaşının yaklaşık dağılımı gösterilmiştir. Bu veri Hastalık Kontrol Merkezi'nden (CDC-Central of Disease Control)[4] (2003) temin edilmiştir. (Dikkat edilirse, dağılım tek modlu bir yapıdadır ve hafif pozitif çarpıktır. Yani dağılım, standart bir normal dağılım şeklinde değildir.)

Ortalama yaklaşık olarak 25, standart sapma 5 olup dağılım hafifçe sağa çarpıktır. Şekil 5.1'de ordinatın (*Y* ekseninin) "bağıl frekans" olarak adlandırıldığı, burada ise "yoğunluk" şeklinde isimlendirildiği görülmektedir.




**Şekil 5.2** İlk çocuğun doğumunda annenin yaşı




**Şekil 5.3** 15 ve 25 yaşlarında çocuk sahibi olma olasılığı

4 http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_10.pdf

yoğunluk

**Yoğunluk** olasılık ile eş anlamlı değildir. "*X*'in farklı değerlerindeki eğrinin yüksekliği" ifadesi yoğunluk kelimesine ilişkin en uygun açıklama olarak düşünülebilir. Aynı zamanda, eğrinin 20 yaşa karşılık gelen noktada 15 yaşa karşılık gelen noktadan daha yüksek olması, bir kadının 20'li yaşlarda çocuk sahibi olması olasılığının ergenlik döneminin ortasında çocuk sahibi olması olasılığına göre daha yüksek olduğu anlamına gelir. Ordinatın farklı şekilde adlandırılması, kesikli bir dağılım yerine sürekli bir dağılım ile ilgilenilmesinden kaynaklanmaktadır. Bir örnek verilmesi konunun daha iyi anlaşılmasına yardımcı olabilir. Eğrinin en yüksek olduğu nokta 20 yaşa karşılık gelmesine karşın, seçilen bir annenin tam olarak 20 yaşında (örneğin, 20,00000000) çocuk sahibi olması olasılığı son derece düşüktür. İstatistikçiler böyle bir olasılığın gerçekte 0'a eşit olduğunu iddia etmektedir. Benzer şekilde, bir kadının 20,00001 yaşında anne olması son derece zayıf bir olasılıktır. Verilen örneklerden de görüldüğü gibi, sürekli dağılımlarda, kesikli dağılımlardaki gibi, belirli bir değere karşılık gelen olasılık üzerine konuşmak mantıklı değildir. Diğer taraftan, birçok kadının yaklaşık olarak 20 yaşında anne olduğu bilinmektedir. Bundan dolayı, tek bir değere ilişkin olasılık yerine; *belirli bir aralığa düşen* olasılık hakkında konuşmak daha mantıklı görünmektedir. Örneğin, bir kadının 24,5 ile 25,5 yaşları arasında anne olması olasılığı belirlenmek istenebilir. Böyle bir aralık Şekil 5.3'te gösterilmiştir. Eğri altındaki toplam alan 1'e eşit olacak şekilde tanımlanırsa, Şekil 5.3'te 24,5 ile 25,5 yaş arasında kalan taralı alan bir bayanın 25 yaşında anne olma olasılığına karşılık gelecektir. Gerekli hesaplamaların yapılması, bu dağılımı açıklayan eşitliğin (eğrinin denkleminin) bilinmesi durumunda olasılığın hesaplanabileceğinin fark edilmesini sağlayacaktır. Eğriye ait fonksiyonun 24,5 ile 25,5 arasında integrali alınarak ilgili olasılık değeri kolaylıkla hesaplanabilir. Ancak bu problemi çözmek için hesaplama yapılmasına gerek yoktur. Çünkü burada üzerinde çalışılan dağılım, daha önce çalışılmış olan dağılımlara oldukça yakındır. Bu kitapta integral hesabı ile ilgilenmeyeceğiz. Bunun yerine, genellikle dağılım tablosuna yönlendirme yapacağız. Bölüm 3'te normal dağılım ile ilgili sunulan işlemler sayesinde bu konuda yeterli deneyime sahipsiniz.

Az önce Şekil 5.3'te 24,5 ile 25,5 arasında kalan aralık incelendi. Bununla birlikte, aynı şeyin başka herhangi bir aralık için de söylenmesi mümkündür. Şekil 5.3'te, 15 yaşın altı ay aşağısı ile altı ay yukarısına (14,5 ile 15,5) karşılık gelen alan da görülmektedir. Bu örnekte ilgili aralığa karşılık gelen gerçek olasılığı hesaplamak için yeterli bilgi bulunmamasına rağmen; Şekil 5.3'ten 25 yaş civarındaki aralığın 15 yaş yakınındaki aralığa kıyasla daha yüksek (taralı bölgenin alanı daha fazla) olduğu anlaşılabilmektedir.

Bir eğrinin altındaki alana ilişkin fikir edinmenin iyi bir yolu, şeffaf bir grafik kağıdı alıp bu kâğıdı şeklin üzerine koymaktır (veya normal bir kâğıt parçasını şeklin üzerine koyduktan sonra ışığa tutmak). Eğer belirtilen aralığa düşen dikdörtgenler sayılır ve eğri altındaki toplam dikdörtgen sayısına bölünürse, tesadüfi olarak seçilen bir puanın ilgili aralığa düşmesine ilişkin yaklaşık olasılık değeri hesaplanabilir. Bu açıklamadan açık bir şekilde anlaşılacağı gibi, kâğıt üzerine çizilen dikdörtgenlerden her birinin alanı ne kadar küçük olursa, hesaplanan olasılık değeri o kadar yaklaşık olacaktır.

## 5.6 Permütasyonlar ve Kombinasyonlar

Bölüm 7'de ve sonrasında gerekli oldukça sürekli dağılımlara başvuracağız. Şimdilik, Bölüm 6'daki Ki-kare testini geliştirmede kullanılacak iki kesikli dağılım (iki kutuplu ve çoklu) üzerinde duracağız. İlk olarak; bu dağılımlar hakkındaki tartışmalarda gerekli olacak permütasyon ve kombinasyon kavramlarını ele almalıyız.

**kombinetoriks**

Matematiğin, objelerin kaç farklı şekilde bir araya getirilebileceğiyle ilgilenen özel branşı **kombinetoriks** olarak bilinmektedir (örneğin, beş kişinin yer aldığı bir gruptan üç kişilik bir komisyon kaç farklı şekilde oluşturulabilir şeklinde bir problem kombinasyonla cevaplanabilir). Bu kitaptaki örneklerin çoğunluğu kombinasyon bilgisi gerektirmemektedir. Ancak kalan az sayıdaki örnek, kombinasyon ve permütasyon kavramlarını kısaca da olsa tanıtmayı ve bu kavramların hesaplamalarında kullanılan formülleri vermeyi gerekli kılar.

## Permütasyonlar

Konuya basit bir örnek ile başlayacağız, bu örnek, ilgili başka durumlara kolaylıkla genişletilebilir. Dört kişinin dondurma külâhı için bir çekilişe katıldığını varsayalım. Bu dört kişinin adı bir şapkanın içine konulmuş ve daha sonra adlar çekilmiştir. Adı ilk çekilen kişi iki top dondurmalı bir külâh kazanırken, adı ikinci sırada çekilen kişi bir top dondurmanın bulunduğu bir külâh kazanacaktır. Adı üçüncü sırada çekilen kişiye sadece dondurma külâhı verilecek, adı son sırada çekilen kişi ise hiç bir şey kazanamayacaktır. Adları; Abigail, Barbara, Cathy ve Donna olan bu kişilerin isimlerinin A, B, C ve D şeklinde kısaltıldığını kabul edelim. Adların aşağıdaki gibi farklı sıralarda çekilmesi mümkündür:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | B | A | C | D | C | A | B | D | D | A | B | C |
| A | B | D | C | B | A | D | C | C | A | D | B | D | A | C | B |
| A | C | B | D | B | C | A | D | C | B | A | D | D | B | A | C |
| A | C | D | B | B | C | D | A | C | B | D | A | D | B | C | A |
| A | D | B | C | B | D | A | C | C | D | A | B | D | C | A | B |
| A | D | C | B | B | D | C | A | C | D | B | A | D | C | B | A |

**permütasyon**

Bu 24 alternatifin hepsinde isimlerin farklı şekillerde sıralanması söz konusudur (bir **permütasyon** denir). Eğer N nesneden her defasında r tanesinin çekilmesine ilişkin permütasyonlar düşünülürse, bunların sayısı $P_r^N$ ile gösterilir ve aşağıdaki ifade ile hesaplanır.

$$P_r^N = \frac{N!}{(N-r)!}$$

**faktöriyel**

Bu ifadedeki N! sembolü, "*N* **faktöriyel**" şeklinde okunmakta ve 1'den N'ye kadar olan bütün doğal sayıların çarpımını temsil eder. [Diğer bir deyişle, N! = N (N-1) (N-2) (N-3)… (1). Tanım gereği 0! = 1'dir].

Dört kişilik ekipten dört ismin çekilmesine ilişkin yukarıdaki örnekte;

$$P_4^4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{1} = 24$$

bulunur. Bu sonuç yukarıda listelenen permütasyon sayısına eşittir.

Çok az kişi, içinde dondurma olmayan bir külâh kazanacağı için heyecanlanır. Bundan dolayı, içinde dondurma olmayan külâh, ödüller arasından çıkarılabilir. Böylece dört kişiden yalnızca ikisi bir ödül kazanacaktır. Bu iki kazanan ismin sırası hâlâ önemlidir. Çünkü ilk çekilen kişi iki top dondurmalı külâh alırken; adı ikinci sırada çekilen kişinin kazanacağı külâhta bir top dondurma bulunacaktır. Sadece iki kişinin ödül kazanması söz konusu olduğundan, şapkaya dört ad konulacak ve konulan dört addan ikisi çekilecektir (çünkü geriye kalan iki kişi herhangi bir ödül kazanmayacaktır). Buna bağlı olarak, bir seferde dört addan ikisinin kaç farklı şekilde çekilebileceği sorusuna cevap aranmaktadır ($P_2^4$). Bu permütasyonlar aşağıdaki gibi yazılıp daha sonra sayılarak iki adın kaç farklı şekilde çekilebileceği kolaylıkla belirlenebilir.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | B | B | A | C | A | D | A |
| A | C | B | C | C | B | D | B |
| A | D | B | D | C | D | D | C |

Veya permütasyonların sayısı aşağıdaki gibi doğrudan hesaplanabilir:

$$P_2^4 = \frac{4!}{(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2} = 12.$$

Burada kazanan kişilere ait 12 farklı sıralama vardır ve sıralamanın önemli olduğu görülmektedir. Çünkü bu 12 farklı sıralama yalnızca kazanan kişileri değil, aynı zamanda iki kişiden hangisinin iki, hangisinin tek top dondurmalı külâh kazanacağı hakkında bilgi de vermektedir.

Şimdi permütasyon içeren daha kullanışlı bir örnek ele alacağız. Fiziksel çekiciliğin slaytlardan değerlendirildiği bir deney tasarlandığını varsayalım. Slaytların sunum sırasının önemli olup olmadığının belirlenmesiyle ilgileniyoruz. Sunacağımız altı slayt var, bunların sıralamasını kaç farklı şekilde düzenleyebiliriz? Bu yine bir permütasyon sorusudur, çünkü slaytların sırası önemlidir. Daha açık deyişle, altı slaytın, her defasında altısının da alınmasıyla ilgili permütasyonla ilgileniyoruz. Veya altı slaytımızın olduğunu varsayalım, fakat bunlardan sadece üç tanesi görünecektir. Bu durumda kaç sıralama yapılabilir? Bu, altı slaytın her defasında üç tanesinin alınmasıyla ilgili permütasyon sorusudur.

Kişilerin altı slaytın tümünün alındığı ilk problem için

$$P_6^6 = \frac{6!}{(6-6)!} = \frac{6!}{0!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{1} = 720$$

Buna göre altı slaytın sıralamasına ilişkin 720 farklı alternatif vardır. Her bir katılımcıya muhtemel tüm alternatifler sunulmak istenirse, 720 denemeye ihtiyaç olacaktır. Yani çok sayıda deneme bulunmaktadır. Fiziksel çekicilik açısından değerlendirilecek kişilerin altı slaytın yalnızca üçünde sunulduğu ikinci problem için

$$P_3^6 = \frac{6!}{(6-3)!} = \frac{6!}{3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6} = 120.$$

Eğer kişiye, muhtemel bütün düzenler sunulmak istenirse, 120 denemeye ihtiyaç duyulacaktır. Böyle bir sonuç hâlâ tasarımın değiştirilmesine imkân tanımayacak kadar büyüktür. Bu durum, tüm muhtemel sıralamalar yerine tesadüfî sıralamaları kullanmamızdan kaynaklanmaktadır.[5]

## Kombinasyonlar

Dondurma çekilişine geri dönüp her iki kazanana da yalnızca bir top dondurma verileceğini varsayalım. Yine şapkaya konulan dört isimden ikisi çekilecektir. Ancak artık *adlardan hangisinin ilk sırada çekildiği önemli olmayacaktır.* Yani, AB sıralaması ile BA sıralaması arasında bir fark bulunmayacaktır. Çünkü her iki durumda da Abigail ve Barbara bir top dondurma kazanmış olacaktır. İsimlerin hangi sıra ile çekildiği artık önemli olmadığından, bu tür durumlarda permütasyon yerine **kombinasyon** kullanılır. Burada, adların hangi sıra ile çekildiği önemsenmeden kazanan adların olası kombinasyonlarının sayısıyla ilgilenilmektedir. Bu kombinasyonlar aşağıda belirtilmiştir:

**kombinasyon**

| | | | |
|---|---|---|---|
| A | B | B | C |
| A | C | B | D |
| A | D | C | D |

Altı kombinasyon vardır. Diğer bir deyişle; dört kişiden kazanan ikisi altı farklı şekilde seçilebilir. (Kazananların permütasyonunun belirlenmesine ilişkin önceki örneğe dönülürse, aynı isimleri içeren çıktıların birleştirildiği görülecektir.)

---

5 Bu kitabın ilerleyen kısımlarında permütasyon testleri olarak adlandırılan bir başlıkla karşılaşılacaktır. Pratikte bunlar tesadüfi sıralama testleridir (randomization tests). Çünkü birçok mâkul problemden dolayı, muhtemel permütasyonların sayısı (hatta kombinasyonların sayısı) son derece büyüktür. Bu sebeple yalnızca tesadüfi sıralamalar çekilmektedir.

Normal olarak, muhtemel bütün kombinasyonları belirlemek yerine bunların kaç tane olduğunu saptamak isteriz. $N$ nesneden her defasında $r$ tanesinin seçilmesi kombinasyonlarının sayısının $C_r^N$ hesaplanması aşağıdaki ifadeyle tanımlanır:

$$C_r^N = \frac{N!}{r!(N-r)!}.$$

Örneğimiz için hesaplama aşağıdaki gibi yapılır:

$$C_2^4 = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} = 6.$$

Şimdi katılımcılara slayt sunulmasına ilişkin örneğe dönelim. Permütasyon, bir dizi slaytın hangi düzende sunulacağı, yani tüm muhtemel sıralamaların sayısıdır. Kombinasyonda ise bu küme içindeki slaytların sıralaması önemli değildir. Ancak altı slayt varken, bu slaytlardan üçünün kaç farklı şekilde seçilebileceğini hesaplamak bir kombinasyon sorusudur.

Altı slayttan her deasında üçünün seçilmesi kombinasyonlarının sayısı,

$$C_3^6 = \frac{6!}{3!(6-3)!} = \frac{\overset{2}{\cancel{6}} \cdot 5 \cdot \overset{2}{\cancel{4}} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}}{\cancel{3} \cdot \cancel{2} \cdot \cancel{1} \cdot \cancel{3} \cdot \cancel{2} \cdot 1} = 20.$$

bulunur. Eğer her bir katılımcıya sırası önemli olmadan farklı üçlü slaytlardan oluşan kümeler gösterilirse, 20 katılımcıya ihtiyaç olacaktır.

Kitabın sonraki kısımlarında *permütasyon testi* olarak adlandırılan süreçler tartışılacaktır. $N$ gözlemden oluşan bir veri kümesindeki puanların iki farklı deney grubuna *tamamıyla tesadüfi* olarak atanması durumunda ($n_1$ ve $n_2$) örneklem ortalamalarının ne olacağı sorusuna cevap bulmak için gözlemlerin gruplara atanabileceği farklı yolların sayısı hesaplanır. Bu durum, $N$ gözlem içinden $n_1$ ve $n_2$ gözlemin seçilmesine ilişkin kombinasyonu yansıtır. (Kombinasyon ile ilgilenildiği halde bu işlemin neden permütasyon testleri olarak adlandırıldığı merak konusu olabilir. Ancak bu kitapta bu merakı giderebilecek bir açıklamaya yer verilmemiştir.) Veri kümesinin tesadüfî olarak oluşturulduğu farklı alternatifler bilindiğinde, ortalamalar arasında fark yaratan çıktıların -en azından saptanan kadar yüksek bir fark- oluşmasına neden olan çıktıların yüzdesi hesaplanabilir. Bu sonuç, verinin $H_0$: doğru iken verilen olasılığına karşılık gelir, bu, $p(D|H_0)$ şeklinde yazılır. Sadece, permütasyon ve kombinasyonun nasıl hesaplanacağını bilmek isteyeceğimiz durumlara örnek vermek amacıyla bundan söz edilmiştir.

## 5.7 Bayes Teoremi

**Bayes teoremi**

Olasılıkla ilgili kavramların uygulamalarına geçmeden önce, ele alınması gereken temel bir bileşen daha bulunmaktadır. **Bayes teoremi** yeni bir kavram değildir (Thomas Bayes tarafından geliştirilmiş ve 1764 yılında Londra'da gerçekleşen ilk bilimsel toplantıdan önce okunmuştur - Bayes'in ölümünden üç yıl sonra). Ancak, davranış bilimlerindeki araştırmalarda Bayes teoremine ve bu teoreme dayanan Bayes istatistiklerine her yıl biraz daha fazla atıf yapıldığı görülmektedir. Hesaplamalar ile ilgili ayrıntılar unutulsa bile; teoremin altında yatan mantığın anlaşılması önemlidir. (Teorem ile ilgili ayrıntılara istenildiği zaman bakılabilir.)

Bayes teorimi, olasılık tahminlerini gözden geçirmek için bilginin nasıl toplanacağını açıklayan bir teoremdir. Bu cümledeki "bilginin toplanması" ifadesiyle, eldeki bilgi arttıkça olasılık tahminin yeniden hesaplanmasını kastediyorum. Örneğin, bize Fred'in öldürüldüğünün söylendiğini ve Willard'ın suçlu olup olmadığı hakkındaki kişisel görüşümüzün sorulduğunu varsayalım. Willard'ın çok iyi bir insan olmadığına ve böyle bir suça karışma

potansiyelinin bulunduğuna inanarak cinayeti Willard'ın işlemiş olması olasılığını $p$ = ,15 olarak düşünebiliriz. Cinayetin işlendiği akşam Willard'ın olay yerinin yakınında görüldüğünü söylersem, bu olasılığı p =, 20'ye ve daha sonra Willard'ın kendine ait tabancasının olduğunu söyleme durumunda p =, 25'e çıkarabilirsiniz. Ardından oldukça güvenilir bir tanığın cinayetin işlendiği gece Willard'ın kendisi ile beyzbol oynadığı yönünde ifade verdiğini söylersem, olasılığı p =10'a düşürebilirsiniz ve böylece devam edebilirsiniz. Yani olay hakkında yeni bilgi edindikçe cinayeti Willard'ın işlemiş olmasına ait olasılığı gözden geçiririz. Bu, meydana gelen bir olay hakkında olasılığın elde edilmesi için bilginin toplanması sürecidir. Bayes istatistikleri ile ilgilenenler için, olasılıklar genellikle kişisel veya öznel olup, bölümün başında tanımlanan analitik ya da frekanslara dayalı bir bakış açısına değil; kişisel inançlara dayanır. Bayes teorimi tüm olasılık türlerinde oldukça iyi çalışsa da; bu teoreme en fazla subjektif olasılıklarda rastlanır.

Stefan Waner'ın http://people.hofstra.edu/Stefan_Waner/tutorialsf3/unit6_6.html bağlantılı internet sayfasından değiştirilerek alınan basit bir örneği inceleyelim (Eğer istenirse bu internet sayfasında, yararlı olabilecek başka örnekler de vardır). Psikologlar, spor hekimliği ile oldukça ilgili hale gelmişlerdir ve bu örnek de aslında bulunla ilgilidir. Ek olarak bu örnek karar verme çalışmaları ile de mükemmel uyum gösterir.

Adsız bir bisikletçinin yarışı tamamladıktan sonra yasaklı hormon kullanıp kullanmadığını saptamak için uygulanan testte başarısız olduğunu varsayalım. (Waner yarış yerine ragbi (rugby) oyununu kullanmıştır, ancak ragbi oyuncularının kurallara uyan kişiler olduğu bilinmektedir. Oysa bisikletçiler hakkında şüphelerimiz oluşmaya başlamıştır.) Bisiklet yarışçısı masum olduğunu ve kesinlikle performans arttırıcı ilâç kullanmadığını savunmaktadır. Burada bisiklet yarışçısının suçlu ya da masum olduğunun yeterli bir olasılık düzeyinde belirlenmesi istenmektedir. Üzerinde çalışılabilecek birkaç veri vardır. Öncelikle, testi pazarlayan ilâç şirketi, yasaklı hormon kullanan kişilerin %95'inde test sonuçlarının pozitif çıktığını ifade etmektedir. Diğer bir deyişle, hormon kullanan bir kişinin test sonucunun pozitif çıkma olasılığı %95'tir. Şirket verisine göre, hormon almayan kişilerin ise %10'unda test sonuçları pozitif çıkmaktadır. Ayrıca, geçmiş tecrübelerimizden bu yarışa katılan kişilerin %10'unun hormon kullandığı gibi ek bir bilgi daha elimizde bulunmaktadır. Bu veri Tablo 5.3'te bir araya getirilmiştir.

önceki olasılık

sonraki olasılık

Elimizdeki önemli bilgi parçalarından biri de, herhangi bir bilgi elde etmeden *önce* kişinin ilâç kullanıcısı olma olasılığını gösteren **önceki olasılık** değeridir. Bu olasılık tabloda $p$(kullanıcı) = ,10 olarak gösterilmiştir. Burada verinin elde edilmesi *sonrasında* hesaplanan yeni olasılığı yansıtan **sonraki olasılık** değerinin belirlenmesi istenmektedir (bu durumda yarışçının hormon testinde başarısız olduğuna ilişkin veri). Devam eden bir bilgi toplama sürecinde, bu sonraki olasılık değeri kendinden sonra gelen adım için önceki olasılık değeri olacaktır.

**NOT**

Bilgi toplandıkça olasılık tahmini yeniden hesaplandığından; bir adımın sonraki olasılık değeri, ardışık adımın önceki olasılık değeri olmaktadır. Bazen tahminin birçok kez gözden geçirilmesi söz konusu olabilir.

**Tablo 5.3** Hormon kullanımı ile ilgili olasılıklar

| Bilinenler | $p$ | Bilgi kaynağı |
|---|---|---|
| $p$(yarışçı kullanıcıdır) $p$(K) | ,10 | Yarışçıların %10'u kullanıcıdır |
| $p$(yarışçı kullanıcı değildir) $p$(KD) | ,90 | Yarışçıların %90'ı kullanıcı değildir |
| $p$(pozitif \| kullanıcı) $p$(P \| K) | ,95 | İlâç Şirketinden |
| $p$(pozitif \| kullanıcı değil) $p$(P \| KD) | ,10 | Yine İlâç Şirketinden |
| $p$(kullanıcı \| pozitif test) $p$(K \| P) | ? | Amaç |

Bayes teoremi, mevcut bilgiden hareketle sonraki olasılığın elde edilebileceğini söyler. Yarışçının hormon kullandığı hipotezini *K* (kullanıcı), hormon kullanmadığı hipotezini ise *KD* (kullanıcı değil) ile göstereceğiz. Test sonuçlarının pozitif olduğuna ilişkin yeni veriyi temsil etmek amacıyla *P* (pozitif) kullanacağız. Buna göre,

$$p(K|P) = \frac{p(P|K) * p(K)}{p(P|K) * p(K) + p(P|KD) * p(KD)}$$

Yukarıdaki tabloda yer alan bilgi kullanılarak aşağıdaki hesaplamaya ulaşılabilir:

$$p(K|P) = \frac{p(P|K) * p(K)}{p(P|K) * p(K) + p(P|KD) * p(KD)}$$

$$= \frac{(,95)(,10)}{(,95)(,10) + (,10)(,90)} = \frac{,095}{(,095 + ,09)} = ,513$$

İlâç testine ilişkin sonuçları elde etmeden önce, yarışçının hormon kullanma suçunu işlemiş olmasına ilişkin subjektif olasılık ,10'du. Çünkü yaklaşık olarak takımdaki yarışçıların yalnızca %10'unun hormon kullandığı bilinmekteydi. Pozitif ilâç testinden sonra beklendiği kadar olmasa da; subjektif olasılık değerinde artış olduğu görülmektedir. Sonraki olasılık değeri ,513 olarak hesaplanmıştır.

Yukarıda da söylediğm gibi, Bayes teoreminin avantajlarından biri, yenilemeli olarak çalışılabilmesidir. Diğer bir deyişle; başka bilgiler toplandıktan sonra ,513 değeri yeni önceki olasılık değeri olarak alınıp, enjektör iğnesi ile ilişkili diğer olasılıklar dâhil edilerek yeni bir sonraki olasılık hesaplanabilir. Yani yeni bilgi toplanarak tahmin geliştirilebilir.

Bayes teoreminin ikinci bir görünümü, elimizdeki olasılıkların bir kısmının sadece akıllıca tahminler olması durumunda yararlı olmasıdır. Örneğin, ilâç şirketi hormon kullanmayan kişilerden kaçının test sonuçlarının pozitif çıktığını açıklamayı reddederse ,20 gibi bir değer deneme amacıyla kullanılabilir. Bu durumda elde edilecek sonraki olasılık ,345 olur ki; bu değer ,513'ten küçük olup elimizdeki, orijinal tahmine göre daha iyi bir kestirimdir.

## İkinci Bir Örnek

İnsanların karar alma süreçlerinde Bayes teoreminin kullanılmasına dayalı çok sayıda çalışma vardır. Belirli bir durumda insanların ne yapmaları ya da ne söylemeleri gerektiği ile gerçekte ne yaptıkları ya da ne söylediklerinin karşılaştırılmasına odaklanan bu çalışmalar, insanların nasıl karar aldıklarını ortaya koymaya yardımcı olmaktadır. Tversky ve Kahneman (1980) tarafından karar vericilere ünlü bir problem yöneltilmiştir. Bu problem, hangi taksi şirketinin bir kazada yer aldığına ilişkin karar vermeyi içermektedir. Şehirdeki iki taksi şirketinden (Yeşil Taksi ve Mavi Taksi) birinin içerisinde yer aldığı bir kazanın gerçekleştiği biliniyor, bu iki taksi şirketinin hangisinin kazada yer aldığı bilinmemektedir. Şehirdeki taksilerin %85'i yeşil ve %15'i mavi olduğundan, yeşil ve mavi taksilerin yüzdesine dayalı ön olasılık değeri ,85 ve ,15 olacaktır. Eğer kaza ile ilgili tek bilinen bu olursa ve birinin sadece bir taksi üzerinde çalıştığı söylenirse, bunun %85 olasılıkla yeşil taksi olduğu yapılabilecek en iyi tahmin olacaktır. Sonrasında bunun mavi taksi olduğunu düşünen bir tanık gelebilir. Bunun kesin olduğu düşünülebilir. Ancak birçok insan geceleri renkleri kusursuz bir şekilde ayırt edemeyebilir. Sigorta şirketi bilgi kaynağını test edip tanığın geceleri renkleri doğru ayırt etme oranını %80 olarak belirleyebilir. Buna göre, tanığa mavi bir taksi gösterildiğinde, tanığın bunun mavi renkli olduğunu doğru bir şekilde ayırt edebilme olasılığı %80, bu taksinin yeşil olduğunu söyleme olasılığı ise %20'dir. (Benzer olasılıklar taksinin yeşil olma durumu için de geçerlidir.) Bu bilgi tanığın mavi olduğunu söylediği otomobilin mavi olmasına ilişkin sonraki olasılığı hesaplamak için yeterlidir.

Aşağıdaki formülde, otomobilin mavi olması *M* ve tanığın otomobili mavi olarak görmüş olması *m*; benzer şekilde otomobilin yeşil olması *Y* ve tanığın otomobili yeşil olarak görmüş olması *y* ile gösterilmiştir.

$$p(M \mid m) = \frac{p(m \mid M)p(M)}{p(m \mid M)p(M) + p(m \mid Y)p(Y)}$$
$$= \frac{(,80)(,15)}{(,80)(,15) + (,20)(,85)}$$
$$= \frac{,12}{,121,17} = \frac{,12}{,29} = ,414$$

Tversky ve Kahneman'ın deneyindeki katılımcıların birçoğu, otomobilin mavi bir taksi olmasını ,80 olarak tahmin etmiştir. Oysa doğru cevap bunun yaklaşık yarısı olup ,41'dir. Bundan dolayı, Kahneman ve Tversky hâkimlerin tanığın ifadesine fazla ağırlık verdiği ve önceki olasılık değerlerini yeterince dikkate almadığı sonucuna ulaşmıştır.[6] Hâkimlerin söyledikleri ile söylemeleri gerekenler arasındaki tutarsızlığı gösteren bu durum, hâkimlerin kullandıkları stratejiler ve nerde hata yaptıkları hakkında önemli ipuçları vermektedir. Yarışçılarda hormon kullanımı ile ilgili örnek birilerine sorulduğunda da, muhtemelen benzer sonuçlara ulaşılacaktır.

## Bir Genel Formül

Yukarıda verilen formül, ele alınan örneklere göre düzenlenmiştir. Araştırmacının kendi amaçlarına göre düzenleyebileceği genel bir formül sunmanın daha yararlı olacağı düşünülmektedir. Belirli bir veri ($D$=data) elimizdeyken bir hipotezin ($H$) doğru olup olmadığını araştırdığımızı varsayalım. Örneğin; $H$ "Yarışçı hormon kullanıcısıdır." ya da "Taksi Mavi Taksi Şirketi'ne aittir." yargılarını temsil edebilir. $D$ "Test sonucu pozitiftir." veya "Tanık taksinin mavi olduğunu ifade etmiştir." gibi eldeki veriye karşılık gelebilir. $\overline{H}$ sembolü, $H$ hipotezinin yanlış olduğu durumları göstermek için kullanılır. Buna göre, aşağıdaki gibi bir genel formüle ulaşılır:

$$p(H \mid D) = \frac{p(D \mid H)p(H)}{p(D \mid H)p(H) + p(D \mid \overline{H})p(\overline{H})}$$

## Hipotez Testlerine Dönüş

Bölüm 4'te hipotez testleri ve hipotez testlerine farklı yaklaşımları tartıştık. Burada bu konuya sadece değinmekle yetinilse de, Bayes teoremi hipotez testlerine ilişkin tartışmalara önemli katkılar sunmaktadır. (Tartışmanın doğasının anlaşılması önemlidir. Ancak ben bunun ötesine geçmeyeceğim.) Hipotez testlerinin, gerçekte cevaplamak istediğimiz soruya cevap bulmak için tasarlanmadığı birçok defa ifade edilmiştir. Bazı veriyi toplar ve sonrasında eldeki veriden hareketle yokluk hipotezinin doğru olma olasılığını sorgularız. Ancak bunun yerine, uyguladığımız istatistiksel işlemler yokluk hipotezinin ($H_0$) doğru olması durumunda elde ettiğimiz verinin toplanma olasılığının ne olduğu hakkında bilgi verir. Diğer bir deyişle, gerçekte sahip olduğumuz olasılık $p(D|H_0)$ iken $p(H_0|D)$ olasılığını bulmak isteriz. Birçok kişi basit bir şekilde Bayes teoremi uygulandığı takdirde aranan cevabın elde edileceğine dikkat çekmiştir.

Buradaki problem gerekli olasılıkların çoğunun bilinmemesidir. Bu olasılıklar *tahmin edilebilir,* fakat bunlar birer tahminin ötesine geçemeyecektir. Bir hormon kullanıcısının testinin pozitif olma olasılığını hesaplamak için hormon kullandığı bilinen kişiler toplanıp bu kişilerden kaç tanesinin test sonucunun pozitif olduğu saptanabilir. Ancak yokluk hipotezinin doğruluğuna ilişkin olasılığın tahmin edilmesi oldukça farklı bir işlemdir. Oto-

6 Daniel Kahneman olasılık teorisiyle 2002 yılında Nobel ödülüne layık görülmüştür. Nobel komitesinin ödülün ölümden sonra verilemeyeceği şeklinde bir politikası olmadığından, herkes bu ödülü Amos Tversky'nin de paylaştığını varsaymaktadır. Geliştirmiş oldukları teori karar alma davranışı psikolojisi üzerinde büyük bir etkiye sahip olmuştur.

parklarda bekleme süresi örneği kullanılarak, bekleyen biri olsun ya da olmasın sizin veya benim insanların aynı hızda park yeri boşalttıklarına ilişkin farklı ön olasılık tahminlerimiz olabilir. Ek olarak, istatistiksel testimiz faydalı olacak şekilde $p(D|H_0)$ değerini elde etmek için tasarlanmış olabilir. Ancak eğer aklımızda alternatif bir hipotez yoksa (yokluk hipotezinin reddi dışında), $p(D|H_1)$ olasılığını nereden elde edebiliriz? Önemli derecede zor olan, bunu elde etmek için de elimizde test sonucu pozitif olanlardan kullanıcı olmayanların yüzdesi gibi somut bir bilgi yoksa, bizi beklerken öndekinin ne kadar yavaş olduğunu bilmediğimizde önemli derecede zor olan alternatifin kestirilmesidir. 5 dakikanın veya 25 dakikanın "çok yavaş" olarak düşünülmesi durumunda elde edilecek olasılıklar önemli ölçüde farklı olacaktır. Gerçek şu ki birçok kişinin denemesine rağmen, hipotez testinde ihtiyacımız olan bu olasılıkların elde edilmesi zordur veya imkânsızdır. İleri sürülen yaklaşımlardan biri, çeşitli mantıklı tahminlerin seçilmesi ve bu farklı tahminler için ortaya çıkan sonucun belirlenmesidir. En inandırıcı tahminlerin aynı sonuca yol açması oldukça yararlı bilgi sağlar. Bayes teoremini kullananların gerçekte yaptıkları, "önceki" olarak adlandırılan nuhtemel değerlerin *dağılımını* varsaymaktır. Böylece gerçek ön olasılığın ,25 ortalama ve ,08 standart sapma değeri ile normal dağılıma sahip olduğu şeklinde bir tahmin yapabiliriz. Alternatif olarak, tüm olasılıkların aynı ön olasılıklara atandığı **bilgilendirici olmayan ön olasılığı** kullanabiliriz.

**bilgilendirici olmayan ön olasılık**

**Bayes istatistikçileri**

**Bayes istatistikçileri** olarak adlandırılan pek çok güçlü savunucu, $H_0$: $\mu = 0$ ve $H_1$: $\mu = 3$ gibi yokluk hipotezi ile alternatif hipotezin kesin bir biçimde tanımlandığı sınırlandırılmış durumları kullanırlar. Ancak davranış bilimlerinde alternatif hipotezin kesin bir şekilde tanımlandığı durumlara asla rastlamadım. Zaman geçtikçe Bayes istatistiğine ilişkin itirazlar büyümektedir. Bu kitapta söz konusu itirazlara ilişkin örnekler yer almaktadır.

## 5.8 Binom Dağılımı

**binom dağılımı**

En yaygın olasılık dağılımlarından biri olan binom dağılımını anlamak için ihtiyacımız olan kombinasyon ve olasılıkla ilgili bilgiyi edindik. Bu başlık altında, **binom dağılımı**nı kısaca tartışacak ve bu dağılımın basit hipotezleri test etmek için nasıl kullanabileceğini göreceğiz. Binom dağılımına değinmeden olasılık ile ilgili bir bölüm yazılmasının mümkün olmadığını düşünüyorum, fakat yapılabilirse daha mutlu olacak çok sayıda öğrenci ve öğretmen olduğuna inanıyorum. Binom dağılımının birçok uygulaması vardır (kısa bir şekilde ele alınacak olan işaret testi bu uygulamalara ait bir örnektir). Burada okuyucunun formülleri ezberlemek istememesi anlaşılır bir şeydir. İstenildiği zaman formüllere bakılabileceğinden formüllerin ezberlenmesine gerek de yoktur.

**Bernouli denemesi**

Binom dağılımı, bir dizi birbirinden bağımsız deneyin ayrık iki çıktıdan biriyle sonuçlandığı durumlar ile ilgilenir. Bu tür bir deney **Bernouli denemesi** (aynı adı taşıyan ünlü bir matematikçiden sonra) olarak adlandırılır. Bernouli deneyine ait en yaygın örnek, bir madenî paranın havaya atılması deneyidir. Paranın beş defa havaya atılması halinde örneğin 3 kez tura gelmesi olasılığını hesaplamak için binom dağılımı kullanılabilir. Birçok kişi için madenî bir paranın havaya atılması ilgi çekici bir örnek olmayabilir. Bir myeloma (kemik civarında oluşan bir tümör) türü hastası için hayatta kalma olasılığı ,93[7] iken, 30 kanser hastasından 20'sinin üç yıldan daha uzun süre yaşaması olasılığının hesaplanması şeklindeki bir örnek madenî para örneğinin heyecanlandırmadığı kişiler için daha ilgi çekici olabilir.

Binom dağılımı sürekli bir dağılımdan çok kesikli bir dağılım örneğidir. Çünkü bir madenî paranın havaya atılması deneyi üç ya da dört kez tura ile sonuçlanabilir. Ancak 3,897 defa tura gelmesi gibi bir durum söz konusu değildir. Benzer şekilde, 30 kanser hastasından 21 ya da 22'si hayatta kalabilir. Ancak hayatta kalan hasta sayısının 21 ile 22 arasında olması mümkün değildir.

---

7 1983 yılında myeloma tanısı konulan 65 yaş altındaki hastaların ortalama yaşam süresi 2-3 yıl idi. Şimdi bu gruptaki bireyler için hayatta kalma olasılığının .93 olduğu rapor edilmektedir. Günümüzde myeloma hastalığına yakalanan kişilerin hayatta kalma olasılıklarının 30 yıl önce bu hastalığı yaşayan kişilere göre daha yüksek olduğu dikkate alındığında; bugün myeloma tanısı konulan hastaların 30 yıl önceki hastalara göre daha şanslı olduğu söylenebilir.

Matematiksel olarak binom dağılımı aşağıdaki gibi tanımlanır:

$$p(X) = C_X^N p^X q^{(N-X)} = \frac{N!}{X!(N-X)!} p^X q^{(N-X)}$$

Bu eşitlikte;

$p(X)$ = $X$ olayının gerçekleşmesi olasılığı
$N$ = Deneme sayısı
$p$ = Herhangi bir denemedeki başarı olasılığı
$q$ = (1-$p$) = Herhangi bir denemedeki başarısızlık olasılığı
$C_X^N$ = $N$'nin $X$'li kombinasyonlarının sayısı

Bu formül iki parçaya bölündüğünde daha anlamlı olabilir. Formülün ikinci parçası ($p^X q^{N-X}$) $N$ denemedeki ilk $X$ denemenin başarılı, geriye kalan $N$-$X$ denemenin başarısız olma olasılığını gösterir. Örneğin; başarı S (success) ve başarısızlık F (failure) ile gösterilmek üzere, SFSSFFF olasılığı yedi denemeden üçünün başarılı ve dördünün başarısız olmasının yollarından biridir. Formülün ilk kısmı [$N!/X!(N-X)!$] denemelerden üçünün başarılı ve dördününün başarısız olduğu durumlara ilişkin farklı sıralamaların sayısını verir - örneğin; SSSFFFF ya da FSFSFSF.

Kombinasyondaki $r$ gösterimi $X$ ile değiştirilmiştir. Çünkü $X$ sembolü veriyi ifade etmek için kullanılır. $X$ ve $r$ sembollerinden hangisinin kullanıldığı isteğe bağlıdır. Kolaylık ve anlaşılırlık dikkate alınarak herhangi biri tercih edilebilir.

**başarı**
**başarısızlık**

İki alternatif çıktı için **başarı** ve **başarısızlık** kelimeleri keyfi olarak kullanılan adlandırmalardır. Kanser hastalığı için konuşulduğunda, anlam gayet açıktır. Bir sürücünün sol tarafa mı yoksa sağ tarafa mı döneceği hakkında konuştuğumuzda gösterim isteğe bağlıdır. Denemelerin birbirinden bağımsız olması gereklidir. Yani, $i$ numaralı denemenin sonuçları $j$ numaralı denemenin sonuçlarını etkilememelidir.

Binom dağılımını göstermek için genellikle farkında olmadan algılama ya da "bilinçaltı algı" şeklinde ifade edilen klâsik örnek ele alınacaktır.[8] Ekranda bir harf veya sayının çok kısa süre parlaması ve ekran karşısındaki kişinin görünen uyarıcının harf mi yoksa sayı mı olduğunu not etmesi yaygın bir başka örnektir. İki uyaran eşit oranda yanıp sönerse ve ekran başındaki kişi ön yargılı olmadan sadece tahmin ile cevap verirse; herhangi bir denemede doğru cevap verme olasılığı ,50'dir.

Uyarıcının sunulduğunu ve yanıtlayıcının dokuz defa doğru, bir kez yanlış cevap verdiğini varsayalım. Eğer yanıtlayan kişi sunulan uyarıcıyı görmüyor ve sadece tahmin ile cevaplıyorsa, %90 oranında (10 denemeden dokuzunda) doğru cevap vermesi olasılığı nedir?

Herhangi bir denemede doğru cevap verme olasılığı $p$ ile gösterilir ve ,50'ye eşittir. Herhangi bir deneme için yanlış cevap verme olasılığı da $q$ ile gösterilir ve yine ,50'ye eşittir. Buna göre,

$$p(X) = \frac{N!}{X!(N-X)!} p^X q^{(N-X)}$$

$$p(9) = \frac{10!}{9!1!} (,50^9)(,50^1)$$

Fakat $10! = 10 \cdot 9 \cdot 8 \cdot \cdots \cdot 2 \cdot 1 = 10 \cdot 9!$ oldugundan

$$p(9) = \frac{10 \cdot \not{9}!}{\not{9}!1!} (,50^9)(,50^1)$$

$$= 10(,001953)(,50) = ,0098$$

8 Philip Merikle, Kazdin'ın "Psikoloji Ansiklopedisi" (2000) adlı eserine bilinçaltı algı konulu mükemmel bir giriş yazısı yazmış ve bilinçaltı algı ile ilgili birtakım ilginç iddiaların maskesini düşürmüştür. Sözü edilen bölüme http://watarts.uwaterloo.ca/~pmerikle/papers/SubliminalPerception.html internet adresinden erişilebilir.

Buna göre, $p = .50$ iken, 10 denemeden dokuzunun doğru cevap ile sonuçlanması oldukça düşük bir olasılıktır. Elde edilen olasılık, bu deney 100 kez tekrarlandığında, deneylerin sadece bir tanesinde 10 yanıttan dokuzunun doğru olma olasılığı bulunduğuna işaret etmektedir. Dolayısıyla 10 uyarıcıdan dokuzuna doğru cevap vermiş olması, yanıtlayıcının uyarıcıyı algılamasa bile şans ile gerçekleşmesi beklenenden daha yüksek bir düzeyde tahminde bulunacak kadar uyarıcının farkında olduğunu gösterir.

10 denemeden 6'sına doğru yanıt verilmesi olasılığı ikinci bir örnek olabilir. Bu örnek için $p^6q^4$ ve $C_6^{10}$ olur. Buna göre;

$$p(6) = \frac{N!}{X!(N-X)!}p^X q^{(N-X)}$$

$$= \frac{10!}{6!4!}(,5)^6(,5)^4$$

$$= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot \cancel{6!}}{\cancel{6!}4 \cdot 3 \cdot 2 \cdot 1}(,5)^{10}$$

$$= \frac{5040}{24}(,00098)$$

$$= ,2051$$

Burada cevaplayıcının performansı, şansla gerçekleşmesi beklenenden daha yüksek değildir.

## Binom Dağılımının Görselleştirilmesi

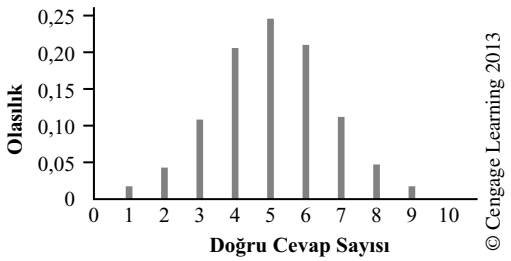
Yukarıda da görüldüğü gibi, yanıtlayıcının altı doğru cevap vermesi olasılığı, dokuz doğru cevap vermesi olasılığına göre daha yüksektir. Bu beklenen bir sonuçtur. Çünkü cevaplayıcının tesadüfî çalıştığı varsayılmakta ve doğru cevap kadar yanlış cevap vermesi de beklenmektedir. Eğer 10 denemeden 0 veya 10 doğru aralığında kalan her bir sonuç için olasılıkları hesaplarsak Tablo 5.4'te görülen sonuçları elde edederiz. Tabloya bakıldığında, bu olasılıkların toplamının 1'e eşit olduğu görülür. Bu toplam, muhtemel tüm sonuçların dikkate alındığını gösterir.

Her bir sonuca ilişkin Tablo 5.4'te yer alan olasılıklar kullanılarak, Şekil 5.4'te gösterildiği gibi, sonuçların dağılımına ait grafik çizilebilir. (Basit grafikler oluşturmaya yönelik R'de yazılmış kolay bir programa bu kitabın Web sitesinden erişilebilmektedir.)

**Tablo 5.4** $p = ,50$ ve $N = 10$ için binom dağılımı

| Doğru Sayısı | Olasılık |
|---|---|
| 0 | .001 |
| 1 | .010 |
| 2 | .044 |
| 3 | .117 |
| 4 | .205 |
| 5 | .246 |
| 6 | .205 |
| 7 | .117 |
| 8 | .044 |
| 9 | .010 |
| 10 | .001 |
| TOPLAM | 1.000 |

© Cengage Learning 2013




**Şekil 5.4** N = 10 ve p = .50 için binom dağılımı

Bu dağılım daha önce görülen dağılımlara benzese de, iki önemli yönüyle onlardan ayrılmaktadır. İlk olarak; ordinatın "frekans" yerine "olasılık" olarak adlandırıldığına dikkate edilmelidir. Çünkü Şekil 5.4 bir frekans dağılımı değil, bir olasılık dağılımıdır. Bu fark önemlidir. Frekans veya bağıl frekans dağılımlarında deney sonuçlarının grafiğini (yani gerçek veriye ait bir grafik) çizerken; burada gerçek verinin grafiğini çizmiyoruz. Bunun yerine, bir olayın veya başkasının ortaya çıkma olasılığına ait grafiği çiziyoruz.

Daha önce işaret edilen bir konuya tekrar değinilecek olursa, ordinatın (*Y*-ekseni) yoğunluk yerine olasılığı temsil etmesi (normal dağılımda olduğu gibi), binom dağılımda sürekli veri yerine kesikli veri ile ilgilenildiğini yansıtır. Normal dağılım gibi sürekli dağılımlarda, belirli bir değere karşılık gelen olasılık yaklaşık olarak 0'dır. (Örneğin, bir kişinin ağırlığının 71,19655515 olma olasılığı imkânsız denecek kadar düşüktür.) Bununla birlikte kesikli dağılımlarda, veri bağıl frekansın kategorilerinden birine düşmekte ve herhangi bir değere ilişkin olasılık kolaylıkla hesaplanabilmektedir. Diğer bir deyişle, kesikli veri kümelerinde belirli bir değere ya da olaya karşılık gelen olasılık ile ilgilenilirken; sürekli dağılımlarda bir aralığa düşen olayların olasılıklarıyla ilgilenilmektedir.

Bu dağılımı diğerlerinden ayıran ikinci husus, bir örneklem dağılımı olmakla birlikte, ampirik olarak değil matematiksel olarak elde edilmiş olmasıdır. Apsis ekseni üzerindeki değerler, her bir gözlemi ya da olayı değil; istatistikleri (verilen bir deneydeki başarılı sonuçların sayısını) temsil eder. Bölüm 4'te örneklem dağılımları ele alınmıştır ve orada ifade edilenler bu bölümde düşüneceğimiz konulara doğrudan uygulanabilecektir.

## Binom Dağılımının Ortalaması ve Varyansı

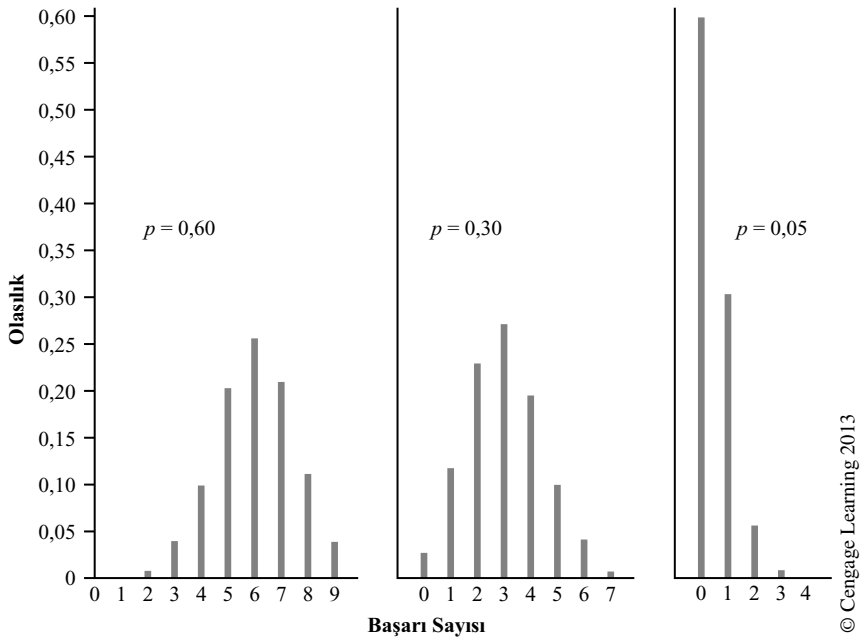
Bölüm 2'de görüldüğü gibi, bir dağılım birçok yönden tanımlanabilmektedir - dağılımı betimlemek için ortalaması, standart sapması, çarpıklığı vb. kullanılabilir. Şekil 5.4'ten yanıtlayıcının cevaplarına ilişkin çıktıların simetrik bir dağılıma sahip olduğu görülmektedir. Dağılımın simetrik yapısı, $p = q =$ ,50 olan tüm durumlar için geçerlidir. Ancak $p$ ve $q$'nun diğer değerleri için dağılım simetrik olmamaktadır. Ayrıca herhangi bir binom dağılımının ortalaması ve standart sapması kolaylıkla hesaplanabilmektedir. Bir binom dağılımında daima;

Ortalama = $Np$

Varyans = $Npq$

Standart Sapma= $\sqrt{Npq}$ değerlerine eşittir.

Örneğin, Şekil 5.4, $N = 10$ ve $p =$ ,50 olan bir binom dağılımını göstermektedir. Dolayısıyla dağılımın ortalaması, 10(,5) = 5 ve standart sapması $\sqrt{10(,5)(,5)} = \sqrt{2,5} = 1,58$ olur.

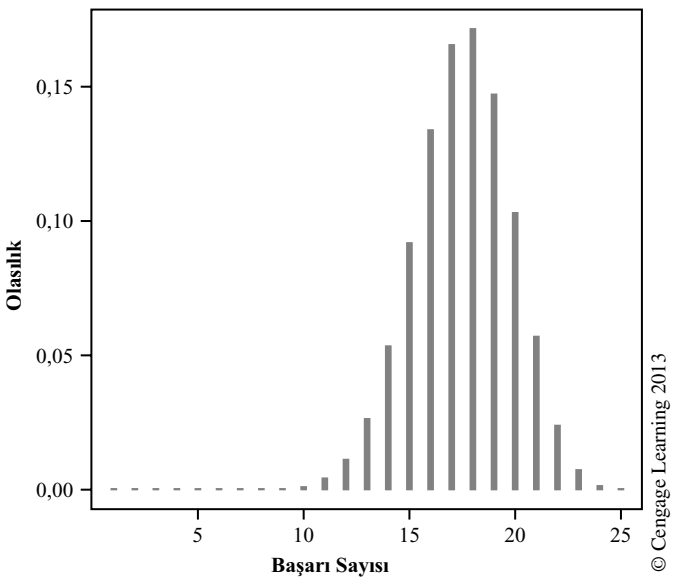



**Şekil 5.5** N =10 ve p = ,60; ,30 ve ,05 değerleri için binom dağılımı

Herhangi bir binom dağılımının ortalamasının ve standart sapmasının belirtilmesinin hipotez testleri için son derece önemli olduğunu kısa bir süre içerisinde göreceğiz. Bununla birlikte ilk olarak, iki noktaya daha dikkat çekilmesi gerekir.

Ekran başındaki kişinin sunulan uyarıcıyı farkında olmadan algılamasına ilişkin örnekte, cevaplayıcının tesadüfî olarak seçim yaptığını varsaydık ($p = q = ,50$). Uyarıcının sunumu, katılımcının doğru cevap verme olasılığını arttıracak şekilde yavaşlatıldığında, aritmetik ortalama aynı kalacak, ancak sonuçlar oldukça farklı olacaktır. Örneklendirmek amacıyla, *p*'nin farklı değerleri için elde edilen üç dağılımın grafiği Şekil 5.5'te verilmiştir.

Şekil 5.5'in sol tarafındaki dağılımda uyarıcının sunum hızı, katılımcının ancak şans düzeyinden biraz daha iyi doğru cevap verebileceği şekilde ayarlanmıştır. Bu dağılım için herhangi bir denemede doğru cevap verilmesi olasılığı ,60'tır. Ortadaki dağılım herhangi bir denemede cevaplayıcının ,30 olasılıkla doğru cevap vermesinin beklendiği durumu temsil etmektedir. Sağ taraftaki dağılım ise, cevaplayıcının sunulan uyarıcıyı *yanlış* tahmin etme konusunda şaşırtıcı bir performans gösterdiği durumu yansıtmaktadır. En sağdaki durum için cevaplayıcının verilen uyarıcıyı doğru tahmin etmesi olasılığı ,05'tir. Bu üç dağılım incelendiğinde, herhangi bir deneme sayısı için *p* ve *q* değerleri ,50'den uzaklaştıkça ortalamanın hâlâ *Np* ve standart sapmanın hâlâ $\sqrt{Npq}$ olduğu dikkat çekmekte, ancak dağılımın giderek daha çarpık hale geldiği görülmektedir. Ayrıca, deneme sayısı arttıkça *p* ve *q* değerlerinden bağımsız olarak dağılımın daha simetrik hale geldiği ve normal dağılıma yaklaştığı gözden kaçırılmamalıdır (Şekil 5.5'te *N* daima 10 olup farklı *N* değerleri için dağılımdaki simetrik yapının ve normalliğin nasıl değiştiği gösterilmemiştir). (Yukarıda belirtilen program yardımıyla değişik *N* ve *p* değerlerine karşılık gelen grafikler oluşturularak farklı *N* ve *p* değerlerinde ne tür değişikliklerin meydana geldiği kolaylıkla görülebilir.) Kural olarak; *Np* ve *Npq* değerlerinin her ikisi 5'ten daha büyük olduğu sürece dağılım normale yakın olmakta fakat dağılım normalmiş gibi işlem yapıldığında yapılan tahminler hatadan uzak olmayacaktır. Şekil 5.6, *p* = ,70 ve 25 *deney* olması durumuna ait binom dağılımını göstermektedir.




**Şekil 5.6** $n = 25$ ve $p = ,70$ için binom dağılımı

## 5.9 Hipotez Test Etmede Binom Dağılımının Kullanılması

Binom dağılımının hipotez testinde yararlı olduğu durumların çoğu, Bölüm 6'da Ki-kare testi kadar iyi bir şekilde ele alınmıştır. Bu nedenle, buradaki tartışma binom dağılımının eşsiz derecede yararlı olduğu durumlar ile sınırlandırılacaktır.

Daha önceki kesimlerde, ekran başındaki kişinin ekranda hızlı bir şekilde sunulan uyarıcıyı değerlendirmesi şeklinde bir örnek sunmuştuk. Sunulan örnek sayesinde, 10 deneme için muhtemel çıktıların dağılımının ve bu çıktılara ilişkin olasılık değerlerinin nasıl hesaplanabileceğini görmüştük. Şimdi bu soruyu kısmen değiştirip, "Ekran başındaki kişiye bir dizi uyarıcı sunulduktan sonra elde edilen veri bu kişinin sunulan karakterleri şans ile gerçekleşmesi beklenenden daha yüksek bir düzeyde tanımlayıp tanımlayamadığını yönelik bir kanıt olarak kullanılabilir mi?" sorusuna cevap arayacağız.

Örneğin, ekran başındaki kişiye sekiz uyarıcının sunulduğunu ve sekiz denemeden yedisinde sunulan uyarıcının doğru cevaplandığını varsayalım. Bu veri, uyarıcıları cevaplayan kişinin, şans ile gerçekleşmesi beklenenden daha iyi işlem yapabildiğinin göstergesi olabilir mi? Bir başka deyişle, eğer ekran başındaki kişi uyarıcılara gerçekten kör tahminle cevap verirse, yedi denemeden sekizinde uyarıcı hakkında doğru cevap alınması mümkün müdür?

Bölüm 4'te açıklanan işlemler izlenerek "ekran başındaki kişi, ekranda bir basamak gözüktüğünde, bunun rakam olduğunu bilir" şeklinde araştırma hipotezinin tanımlanması ile başlayabiliriz (en azından göstermek istediğimiz şeyin ne olduğunu). Diğer bir ifadeyle, araştırma hipotezi ($H_1$), ekran başında uyarıcılara cevap veren kişinin şans düzeyinden daha yüksek bir performans sergilediği ($p > ,50$) şeklindedir. (Örneği basitleştirmek için sadece tek yönlü test seçilmiştir; ancak genellikle iki yönlü testin kullanılması tercih edilir.) Bu durumda yokluk hipotezi, "uyarıcıların sunulduğu ekranın başında yer alan kişinin davranışları şans düzeyinden farklılık göstermemektedir" biçiminde olur ($H_0$: $p = ,50$). Yokluk hipotezinin doğru olduğu göz önüne alındığında, sekiz denemedeki doğru cevap sayısına ilişkin örneklem dağılımı $p = ,50$ için binom dağılımı kullanılarak elde edilir. Muhtemel doğru cevapların sayısının her birinin olasılığını hesaplamak yerine (örneğin, Şekil 5.5'te yapıldığı gibi), hesaplanması gereken tüm uyarıcıları değerlendiren kişinin yedi veya sekiz doğru cevap vermesi olasılığını hesaplama yoluna gideriz. Çünkü verilen

uyarıcının ne olduğu (harf mi, sayı mı) hakkında değerlendirme yapan kişinin en azından tesadüfî olarak seçim yapması durumundaki kadar bir performans sergilemesi olasılığının bilinmesi amaçlanmaktadır.

$N$ toplam ve $X$ doğru cevap ile sonuçlanan deneme sayısını göstermek üzere, sekiz denemeden yedisinin doğru olması olasılığı aşağıdaki gibi hesaplanır:

$$p(X) = C_X^N p^X q^{(N-X)}$$
$$p(7) = C_7^8 p^7 q^1$$
$$= \frac{8!}{7!1!}(,5)^7(,5)^1 = 8(,0078)(,5) = 8(,0039) = ,0312$$

Buna göre, ekran başındaki kişinin sekiz uyarıcıdan yedisini şansla doğru bilmesi olasılığı ,0312'dir. Fakat, sıfır hipotezinin "Eğer $H_0$ doğru ise, en azından bu kadar doğru cevap verilmesi olasılığı nedir?" formunda sorulmasını biliyoruz. Bir diğer ifadeyle $p(7)$ ve $p(8)$ olasılıklarını toplama ihtiyacındayız.

$$p(8) = C_8^8 p^8 q^0 = 1(,0039)(\ 1) = ,0039$$

Buradan

$$\begin{aligned} p(7) &= ,0312 \\ +\,p(8) &= \underline{,0039} \\ p(7 \text{ or } 8) &= ,0351 \end{aligned}$$

elde edilir. Burada en azından yedi doğru cevap verilmesi olasılığının yaklaşık olarak ,035 olduğunu görüyoruz. Daha önce, I. Tip hata ($\alpha$) olasılığının ,05'ten küçük veya eşit olması durumunda $H_0$ hipotezinin reddedileceğini ifade etmiştik. Eğer $H_0$ doğru ise (yani eğer $p$ = ,50 ise), sekiz denemenin en az yedisinde doğru cevap verilmesi olasılığını ,035 olarak belirlediğimizden, $H_0$ hipotezini reddederiz. Dolayısıyla, ekran başında bulunan ve verilen uyarıcının harf mi yoksa sayı mı olduğu hakkında karar veren kişinin şans düzeyinden daha yüksek bir performans gösterdiği sonucuna varırız. Daha açık bir ifadeyle, değerlendiren kişinin doğru bildiği uyarıcıların sayısı sadece kör tahminle cevap vermesi durumunda beklenen performanstan daha yüksektir.[9]

## İşaret Testi

**işaret testi**

Hipotez testinde binom dağılımının kullanılmasına ilişkin bir başka örnek, en basit testlerden biri olan **işaret testi**dir. İşaret testi çok basit olmasına rağmen, birçok durumda oldukça yararlı bir testtir. "İnsanlar birbirlerini tanıdıktan sonra bireysel farklılıklarını kabul etme eğilimlerinin yükseldiği" şeklinde bir hipotez kurduğumuzu varsayalım. Bu hipotezi test etmek için küçük bir üniversiteye kaydolan birinci sınıf erkek öğrencilerinden, fiziksel görünüm açısından değerlendirilmek üzere araştırmaya dahil edilen 12 öğrenciyi puanlamalarını isteyelim.

**Tablo 5.4** Dönem başında ve sonunda fiziksel görünüme ilişkin ortanca sıralar

| Değerlendirilen Kişiler | 0 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dönem Başlangıcı | 12 | 21 | 10 | 8 | 14 | 18 | 25 | 7 | 16 | 13 | 20 | 15 |
| Birinci Dönemin Sonu | 15 | 22 | 16 | 14 | 17 | 16 | 24 | 8 | 19 | 14 | 28 | 18 |
| Kazanç | 3 | 1 | 6 | 6 | 3 | –2 | –1 | 1 | 3 | 1 | 8 | 3 |

9 Kesikli dağılımlarla ilgili bir problem, tam olarak .05 olasılıklı çıktıların oldukça nadir olmasıdır. 7 doğru cevap verilmesine ilişkin örnekte p = .035 olduğundan yokluk hipotezi reddedilmiştir. 6 doğru cevap verilmesine ilişin olasılık hesaplanmış olsaydı .133 gibi bir olasılık elde edilir ve bu değer .05'ten büyük olduğundan yokluk hipotezi reddedilmezdi. Tam olarak .05'e eşit olası bir sonuç yoktur. Buna bağlı olarak, seçilen kritik değer ya çok sıkı ya da çok geniş olmaktadır. Bu sorunun çözümüne yönelik olarak ciddi bir biçimde kabul gören önerilerden biri, gözlenen sonuca ait olasılık ile uçlardaki sonuçların olasılıklarının toplamını alan ve "orta-p (mid-p)" olarak adlandırılan yaklaşımdır. Bu yaklaşımın tartışması için Berger'a (2005) bakınız.

Yüksek puanlar daha yüksek çekiciliği temsil etmektedir. İlk dönemin sonunda, birbirlerini tanımalarının ardından öğrenciler aynı 12 kişiyi tekrar puanlasınlar. Araştırmada, Tablo 5.5'teki veri kümesinin elde edildiğini varsayalım. Bu tablodaki değerler, dönem başında ve ilk dönem sonunda fiziksel görünümleri açısından 30 puan üzerinden değerlendirilen 12 kişinin ortalama puanlarını göstermektedir.

Bu tablodaki kazanç puanları dönem başındaki puanların dönem sonundaki puanlardan çıkarılmasıyla elde edilmiştir. Örneğin, değerlendirilen ilk bireyin fiziksel görünümüne ilişkin dönem başındaki puanı 12, dönem sonundaki puanı 15 olduğundan kazanım puanı 3 olarak hesaplanmıştır. Dikkat edilirse, 12 katılımcıdan 10'u için dönem sonundaki puanlar dönem başındaki puanlara göre daha yüksektir. Diğer bir deyişle, 12 katılımcının 10'u için kazanç puanlarının işareti pozitiftir. (Farkın büyüklüğüne değil; işaretine bakılır. İşaret testi, adını da buradan alır.)

Bu örnek için yokluk hipotezini düşünelim. Eğer insanların birbirlerini tanıması fiziksel görünüme ilişkin puanlamalarını etkilemiyorsa, puanlamalarda sistematik bir değişim olmasını beklemeyiz (başka değişkenlerin etkili olmadığı varsayılmaktadır). Birbirine bağımlı puanlar göz ardı edildiğinde (buradaki örnekte zaten bulunmamaktadır); ilk dönemin sonunda, puanların yaklaşık yarısının artması, yarısının ise azalmasını bekleriz. Buna göre; $H_0$ hipotezi altında, $p$(artış) = $p$(düşüş) = ,50'dir. Şimdi eğer $H_0$ hipotezi doğru ise, 12 kişiden en az 10'unun dönem sonundaki puanlarının dönem başındaki puanlarına göre daha yüksek olması olasılığını hesaplamak için binom dağılımını kullanabiliriz.

$$p(10) = \frac{12!}{10!2!}\,(,5)^{10}(,5)^{2} = ,0161$$

$$p(11) = \frac{12!}{11!1!}\,(,5)^{11}(,5)^{1} = ,0029$$

$$p(12) = \frac{12!}{12!0!}\,(,5)^{12}(,5)^{0} = ,0002$$

Bu hesaplamalardan, eğer yokluk hipotezi doğruysa ve fiziksel görünüme ilişkin yapılan puanlamalar insanların birbirini tanımasından etkilenmiyorsa, 12 kişiden en az 10'unun dönem sonundaki puanlarının daha yüksek olması olasılığı; ,0161 + ,0029 + ,0002 = ,0192'dir. Bu değerler kesme noktası olarak kabul edilen geleneksel ,05 değerinden daha küçük olduğundan, $H_0$ hipotezi reddedilir ve fiziksel görünüme ilişkin puanların dönem sonunda arttığı sonucuna ulaşılır. (Bu fark bireylerin birbirlerini tanıması dışındaki değişkenlerle de açıklanabilir. Ancak, en azından açıklanması gereken bir fark olduğunu göstermesi açısından analiz sonuçları önemlidir.)

## 5.10 Çok Terimli Dağılım

Yukarıda incelenen binom dağılımı, daha genel bir dağılım olan **çok terimli dağılım**ın özel bir halidir. Binom dağılımında sadece iki çıktısı olan olaylarla ilgilenilmektedir - bir madeni paranın yazı veya tura gelmesi, bir şarabın pahalı veya ucuz olarak değerlendirilmesi gibi. Olası iki çıktının söz konusu olduğu olaylarla ilgilendiğinden bu dağılım "iki terimli" olarak adlandırılır. Bununla birlikte, birçok durumda, herhangi bir olaya ilişkin muhtemel çıktıların sayısı ikiden fazla olmaktadır - bir zarın atılması deneyinde altı olası çıktı vardır, bir labirentte üç seçenek olabilir (sol, sağ ve orta) veya politik görüşler; Tamamıyla Destekleme, Destekleme, Kararsız, Karşı, Şiddetle Karşı gibi sınıflandırılabilir. Bu durumlarda daha genel bir dağılım olan çok terimli dağılıma ihtiyaç duyarız.

$k$ sayıda kategorinin her biri için olasılık değerleri $p_1, p_2, p_3,\ldots,p_k$ olarak tanımlanırsa, kategori$_1$ ile sonuçlanan denemelerin sayısı $X_1$, kategori$_2$ ile sonuçlanan denemelerin sayısı $X_2,\ldots,$ kategori$_k$ ile sonuçlanan denemelerin sayısı $X_k$ olmak üzere $p(X_1, X_2,\ldots, X_k)$ olasılığı aşağıdaki gibi hesaplanır:

$$p(X_1, X_2, \cdots, X_k) = \frac{N!}{X_1!X_2!\cdots X_k!}\,p_1^{X_1}p_2^{X_2}\cdots p_k^{X_k}$$

Burada $N$, binom dağılımındaki ile aynı anlama sahiptir. $k = 2$ iken $p_2 = 1 - p_1$ *ve* $X_2 = N - X_1$ olmakta ve binom dağılımına karşılık gelmektedir.

Kısa bir örnek olarak elimizde iki yüzü siyah, üç yüzü kırmızı ve bir yüzü beyaz olan bir zar bulunduğunu varsayalım. Bu zar atıldığında, siyah yüz gelmesi olasılığı 2/6 = ,333, kırmızı yüz gelmesi olasılığı 3/6 = ,500 ve beyaz yüz gelmesi olasılığı ,167'dir. Zar 10 kez atılırsa; 4 defa siyah, 5 defa kırmızı ve 1 defa beyaz gelmesi olasılığı nedir? Bu olasılık aşağıdaki gibi hesaplanır:

$$p(4,5,1) = \frac{10!}{4!5!1!}(,333)^4(,500)^5(,167)^1$$
$$= 1260(,333)^4(,500)^5(,167)^1 = 1260(,000064)$$
$$= ,081$$

Bu noktada, çok terimli dağılım hakkında söylenecekler tamamlanmıştır. Bölüm 6'da, Ki-kare ve ileride kullanılması muhtemel olan diğer testlerin temelini oluşturan analizler ele alınırken, çok terimli dağılım tekrar karşımıza çıkacaktır.

## Anahtar Terimler

Analitik bakış (5.1)
Frekansçı bakış (5.1)
Yerine koyarak örnekleme (5.1)
Subjektif olasılık (5.1)
Olay (5.2)
Bağımsız olaylar (2.2)
Ayrık olaylar (5.2)
Kapsayan olaylar (5.2)
Toplama kuralı (5.2)
Çarpma kuralı (5.2)
Birleşik olasılık (5.2)
Koşullu olasılık (5.2)
Koşullu olmayan olasılık (5.2)
Yoğunluk (5.5)
Kombinatoriks (5.6)
Permütasyon (5.6)
Faktöriyel (5.6)
Kombinasyon (5.6)
Bayes teoremi (5.7)
Önceki olasılık (5.7)
Sonraki olasılık (5.7)
Bilgilendirici olmayan ön olasılık
Bayes istatistiği (5.7)
Binom dağılımı (5.8)
Bernouli denemesi (5.8)
Başarı (5.8)
Başarısızlık (5.8)
İşaret testi (5.9)
Çok terimli dağılım (5.10)

## Alıştırmalar

5.1 Olasılığa ilişkin analitik, frekansçı ve öznel bakışa birer örnek veririniz.

5.2 Yerel itfaiye örgütü için düzenlenen bir organizasyon kapsamında yapılacak piyango çekilişi için A şahsının bir bilet; B şahsının iki bilet aldığını varsayalım. Toplamda 1000 bilet satıldığına göre;

a. Büyük ödülü A'nın kazanma olasılığı nedir?

b. Büyük ödülü B'nin kazanma olasılığı nedir?

c. Büyük ödülü A veya B'nin kazanma olasılığı nedir?

5.3 Alıştırma 5.2 için toplamda 10 bilet satıldığını ve iki büyük ödül olduğunu düşünelim.

a. A'nın ilk ödülü kazanmadığı bilindiğine göre ikinci ödülü kazanma olasılığı nedir?

b. İlk ödülü B'nin ve ikinci ödülü A'nın kazanma olasılığı nedir?

c. İlk ödülü A'nın ve ikinci ödülü B'nin kazanmasına ilişkin olasılık nedir?

d. İlk ve ikinci ödülü A ve B'nin kazanması olasılığı nedir?

5.4 Alıştırma 5.3'ün hangi kısmı birleşik olasılıkla ilgilidir?

5.5 Alıştırma 5.3'ün hangi kısmı koşullu olasılıkla ilgilidir?

5.6 Birleşik olasılıkla ilgili basit bir örnek yazınız.

5.7 Koşullu olasılıkla ilgili basit bir örnek yazınız.

5.8 Bazı evlerde, annenin davranışı ile bebeğinin davranışı veya tersi birbirinden bağımsız görünebilir. Bir anne çocuğuna günde iki saat ve çocuğu da annesine günde üç saat bakıyor olsun. Eğer anne ve çocuğun bu davranışları birbirinden bağımsız ise, anne ile çocuğun aynı anda birbirlerine bakma olasılıkları nedir?

5.9 Alıştırma 5.8'de hem annenin hem de çocuğun akşam 8.00 ile sabah 7.00 arasında uyudukları varsayılırsa, ilgili olasılık değeri ne olur?

5.10 Süpermarketlerdeki tanıtımlarda yer alan mesajlar hakkında bir çalışma yapıldığını varsayalım. Çalışma kapsamında, insanların üzerlerinde etkisi olmayan tanıtımlarla ilgili davranışlar araştırılıyor olsun. Eğer mesajın müşterinin davranışları üzerinde bir etkisi yoksa, tanıtımın üzerindeki "çöpe atmayınız" mesajının, tanıtımların müşteriler tarafından çöpe atılmasını azaltma olasılığı ,033'tür. Bu tanıtımların %4,5'inin çöpe atıldığını varsayalım. Bu durum mesajın etkililiği hakkında ne gibi bir bilgi verir?

5.11 Bir gözlemin belirli bir aralığa düşme olasılığı ile ilgili yaygın bir sürekli dağılım örneği veriniz.

5.12 Alışılageldik bir biçimde, kesikli değişken muamelesi yapılan sürekli bir değişken örneği veriniz.

5.13 Kesikli değişkene ait iki örnek veriniz.

5.14 Lisansüstü eğitim başvurularını değerlendiren bir komite, en yüksek başvurular arasında geçerli ayrım yapamayacağını düşünmektedir. Bu yıl lisansüstü eğitim almak isteyen 300 kişinin başvurularını puanlayan komite, en yüksek %20'lik dilim içerisinden tesadüfi olarak 10 başvuruyu seçecektir. Üst %20'lik dilim içerisinde kalan bu başvurulardan herhangi birinin kabul edilmesi olasılığı nedir (başvurulara verilen puanlar hakkında bilginizin olmadığını varsayınız)?

5.15 Alıştırma 5.14 için

a. 300 öğrenci arasında en yüksek fakülte derecesine sahip olduğu bilinen bir kişinin başvurusunun kabul edilmesine ait koşullu olasılık nedir?

b. En düşük puana sahip olduğu bilinen bir kişinin kabul edilmesine ait koşullu olasılık nedir?

5.16 Ekte sunulan ve internet sayfasında Add.dat adı ile kayıtlı olan veri dosyasını kullanınız.

a. Eğer puanlar 52,6 ortalama ve 12,4 standart sapma ile normal dağılıma sahipse, tesadüfî olarak seçim yapan bir kişinin 50'den yüksek bir ADDSC puanına sahip olma olasılığı nedir?

b. 50 puanı aşan örneklem yüzdesi nedir?

5.17 İnternet sayfasında Add.dat adı ile kayıtlı olan ve ekte yer alan veri kümesindeki dosyayı kullanınız.

a. Eğer puanlar 54,3 ortalama ve 12,9 standart sapma ile normal dağılıma sahipse, bir erkeğin 50'den yüksek bir ADDSC puanına sahip olması olasılığı nedir?

b. 50 puanı aşan erkeklerin yüzdesi nedir?

5.18 Ekte yer alan veri kümesindeki bilgiyi kullanarak ADDSC puanı en az 60 olduğu bilinen bir kişinin okulu bırakmasına ilişkin ampirik olasılığı hesaplayınız. Burada normalliğe ihtiyaç duyulmamaktadır.

5.19 Ekte yer alan veri kümesi için ADDSC'ye ilişkin 66 kesme puanının öğrencilerin okul terki için bir yordayıcı olup olmadığının belirlenmesinde koşullu olasılık nasıl kullanılabilir?

5.20 Ekte sunulan veri setindeki puanları kullanarak Alıştırma 5.18'de hesaplanan ADDSC puanı en az 60 olan bir öğrencinin okulu bırakmasına ilişkin koşullu olasılık ile ADDSC puanından bağımsız olarak kişinin okulu bırakmasına ilişkin koşulsuz olasılığı karşılaştırınız.

5.21 Beş seçenekli bir görevde, katılımcılardan araştırmacının keyfi olarak doğru saptadığı bir uyarıcı seçmeleri istenmektedir. 10 katılımcı sadece ilk denemede tahmin edebilmektedir. İlk denemede doğru cevapların sayısına ilişkin örneklem dağılımının grafiğini çiziniz.

5.22 Alıştırma 5.21 için, ikinci denemede 10 katılımcıdan 6'sı doğru ise çıkarılacak sonuç ne olur?

5.23 Alıştırma 5.21 dikkate alındığında, grup olarak katılımcıların gösterdikleri performansın şans düzeyinde olmadığı sonucuna varmak için bir denemedeki minimum doğru seçenek sayısının kaç olması gerekir?

5.24 Otomobil satan kişiler genellikle bayan ve erkek müşterilere daha farklı şekilde davranmaktadırlar. Bu davranış ile ilgili olarak; basit, birleşik ve koşullu olasılığı örneklendirecek ifadeler oluşturunuz. Bu ifadelerin doğruluğunu test etmek için nasıl bir başlangıç yapılması gerekir?

5.25 Bir yerel insan hakları örgütünün üyesi olduğunuzu varsayalım. Evdeki ayrımcılığı incelemek için olasılık hakkında bildiklerinizi nasıl kullanabilirsiniz.

5.26 İnsanın zihinsel süreçleri ile ilgili yapılan bir çalışmada, farklı kelime gruplarının (isim, sıfat, zarf, fiil) hatırlanması incelenmektedir. Araştırma kapsamında, katılımcıya bunlardan her biri gösterilmektedir. Araştırmada, bir sıra etkisi olabileceği düşünüldüğünden, farklı kelime gruplarını farklı sıralarda görecek değişik katılımcıların çalışmaya dâhil edilmesi planlanmaktadır. Kelime gruplarına ilişkin her bir sıralama için bir katılımcı gerekecekse, böyle bir araştırma için toplamda kaç katılımcıya ihtiyaç duyulur?

5.27 Alıştırma 5.26 ile ilgili olarak; zaman kısıtlamasının söz konusu olduğunu ve bundan dolayı katılımcıların her bir kelime grubundan sadece ikisini görebileceğini varsayalım. Deneyin diğer kısımları aynı kaldığına göre, bu durumda kaç katılımcıya ihtiyaç duyulur? (Uyarı: Veriyi nasıl analiz edebileceğinizden emin olmadıkça, gerçekten böyle bir deney gerçekleştirmeyi düşünmeyiniz.)

5.28 Bir öğrenme görevinde, bir katılımcıya beş farklı buton sunulmaktadır. Katılımcının önceden karar verilmiş olan bir sıra ile belirli üç butona basması gerekmektedir. Katılımcının ilk denemede butonlara doğru bir şekilde basması olasılığı nedir?

5.29 Bir dondurma dükkânında altı farklı tatta dondurma bulunmaktadır. Farklı tatlardan sadece bir top olmak üzere, bu dondurmaların herhangi bir sayıda herhangi bir kombinasyonu oluşturulabilmektedir. Kaç farklı dondurma külâhı kombinasyonu reklamı verilebilir? (Nane, ahududu ve fıstık gibi farklı tatlardan hangisinin altta hangisinin üstte bulunduğu, yani dondurma toplarının külâh üzerindeki sıralaması, dikkate alınmamaktadır. Her dondurma külâhında en azından bir top dondurma bulunmalıdır - boş dondurma külahı sayılmamaktadır).

5.30 Altı dış elektro kodun fare beynine takılacağı bir araştırma yürütülmektedir. Araştırma asistanı müzik dinlemek için kayıt cihazını eve götürdüğünde, kayıt cihazındaki altı kanallı hoparlörden ikisi patlamıştır. Buna göre, beyinden kaç farklı şekilde kayıt yapılabilir. (Sinyallerden hangisinin hangi kanala gittiği bir farka neden olmamaktadır.)

5.31 Güncel olaylar hakkındaki bilgi konulu bir çalışmada, üniversite son sınıf öğrencilerine 20 maddelik bir doğru-yanlış testi uygulanmaktadır. Çok açıkgöz olmayan öğrencilerden biri testteki sorulardan 11'ini doğru cevaplamıştır. Öğrencinin tahminden başka bir şey yaptığına inanmak için herhangi bir neden bulunmakta mıdır?

5.32 Bu bölümün önceki kesimlerinde, 60 M&M'den yerine koyarak 25 mavi M&M çekilmesi olasılığı ,0011 olarak ifade edilmişti. Bu sonucu tekrar hesaplayınız. (Uyarı: Çok basamaklı sayılar ile işlem yapabilen bir hesap makinesi kullanılmalıdır. Aksi durumda yuvarlama hataları söz konusu olabilir. Burada kullanılan hesap makinesinde ,0011 değeri rapor edilmiştir.)

5.33 Bu kolay bir soru olmayıp Bölüm 3, 4 ve 5'te yer verilen konuların bir arada ele alınmasını gerektirir. Sürücülerin %60 oranında kabul edildiğine inanmak için iyi bir nedenimizin olduğu bir sürüş testi gerçekleştirdiğimizi varsayalım. Bu test 30 sürücüye uygulanmış ve sürücülerden 22'si testi geçmiştir. Sonuç $H_0$ hipotezini ($p$ = ,60) reddetmek için yeterince büyük müdür? X = 22, 23,…, 30 değerleri için binom dağılımının kullanıldığı bir çözüm, bu problem için kullanışlı bir yol olmayacaktır. Ancak binom dağılımının ortalaması, varyansı ve grafiğinden yararlanılabilir. Dağılımın nasıl bir görünüme sahip olduğunu gösteren bir grafik yardımıyla problem çözülebilir.

5.34 İşaret testinin uygun olacağı basit bir deney hazırlayınız.

a. Mantıklı bir veri seti oluşturup işareti testini uygulayınız.

b. Sonuçlar hakkında uygun çıkarımlarda bulununuz.

## Tartışma Soruları

5.35 "Ortalamalar kanunu" ya da "kumarbaz mantıksızlığı", sık sık rastlanan bir inançtır. Bu inanca göre; tesadüfî bir olaya ilişkin bir dizi deneme, olay hakkındaki olası alternatiflerden biriyle sonuçlanmışsa, sonraki denemeler diğer alternatif/alternatifler ile sonuçlanacaktır. [Örneğin; hilesiz bir madenî paranın 20 atılışından 18'inde tura gelmişse, bir sonraki denemede yazı ve tura gelme olasılığı %50-%50 şeklinde (yanlış) olmayacaktır.] Kumarbaz mantıksızlığı bu inançtan farklı değildir. Bu inançta, önceki denemelerin sonuçları hakkında madeni paraların, insanlardan daha iyi bir belleğe sahip olduğu gibi bir yanlış söz konusudur. Ann Watkins, Şans dergisinin 1995 bahar baskısında "Ortalamalar yasasının" doğru olduğuna inanan insanlar hakkında çeşitli örnekler sunmuştur. Watkins'in verdiği örneklerden biri, Sevgili Abby'ye yazılan bir mektup hakkındadır. Bu mektubu yazan kişi, sekiz çocuğunun sekizinin de kız olmasından şikâyet etmiştir. Kaderi eleştirmiş ve ortalamalar yasasının 1'e 100 oranında onun lehine olduğu şeklinde doktoru tarafından kendisine bilgi verildiğini belirtmiştir. Watkins başka bir örneğe daha atıf yapmıştır. Bu örnekte, İngilizlerdeki şişman erkek sayısının Amerikalılardaki şişman erkek sayısından daha az olduğu belirtilmiş ve ortalamanın aynı olabilmesi için şişman İngiliz erkeklerinde fazla kilo oranının daha yüksek olması gerektiği yazılmıştır. Son olarak, bu yasanın gerçekten dikkat çekici bir uygulaması için Marlon Brando'ya atıf yapılmıştır. Brando'nun çok sayıda hayranın bulunmaktadır. Bu hayranların hepsinin tek bir cinsiyetten olmasının şaşırtıcı olacağı ifade edilmiş ve ortalamalar kanununa göre Marlon Brando'nun hem kadın hem de erkek hayranlarının bulunmasının beklendiği dile getirilmiştir (Los Angeles Times, Eylül 18, 1994, "Book Reviews," s. 13). Bu örneklerde yanlış olan nedir? Böyle bir kanunun altında ne gibi bir inanç sistemi yatmaktadır? Abby'ye mektup yazan kadının sekiz çocuğunun tümünün kız olması nasıl açıklanabilir?

5.36 40 yaşında kadınların %1'inin meme kanseri olması beklenmektedir. Meme kanseri olan kadınların %80'inin mamografi sonuçları pozitif çıkmaktadır. Ayrıca, meme kanseri olmayan kadınların %9,6'sında mamografi sonuçları pozitif çıkmaktadır. Bu yaş grubundaki bir kadının kanser testi sonuçları pozitifse, bu kadının gerçekten kanser olması olasılığı nedir? Bu problemi çözmek için Bayes teoremini kullanınız. (İpucu: MK, meme kanserini göstermek üzere; $p(MK) = .01$, $p(+|MK) = ,80$ ve $p(+|\overline{BC}) = ,096$. $p(MK|+)$ olasılığının hesaplanması istenmektedir.)

5.37 Kanser hastası olan kadınların %80'inde mamografi sonuçlarının pozitif çıktığı bilinmekle birlikte, Alıştırma 5.36'da bulunan olasılık değeri muhtemelen beklenenden daha düşüktür. Hesaplanan olasılığın beklenenden daha düşük olmasının sebebi nedir?

5.38 Alıştırma 5.36'da kanser hastası olmayan kadınların yalnızca %5'inde mamografi sonuçlarının pozitif çıktığı şeklinde bir değişiklik yapılırsa; kanser testi sonuçları pozitif olan bir kadının gerçekten kanser olması olasılığına ilişkin hesaplanan yeni değer kaç olur? (Diğer bir deyişle, yanlış olan pozitif sonuçların oranı azaltılmaktadır.)



*Bölüm 6*

# Kategorik Veri ve Ki-Kare

## Hedefler

Kategorik veri kümelerinde hipotez testi için ki-kare yönteminin açıklanması ve sonuçların ne anlama geldiğinin açıklığa kavuşturulması, sonuçlarınızın açıklığa kavuşturulmasında diğer ölçümlerin incelenmesi.

## Konular


**6.1** Ki-Kare Dağılımı
**6.2** Ki-Kare Uyum İyiliği Testi – Tek Yönlü Sınıflama
**6.3** İki Sınıflamalı Değişken: Çapraz Tablo Analizleri
**6.4** Ek Bir Örnek – 4 x 2'lik Bir Desen
**6.5** Sıralama Düzeyindeki Veri Kümeleri İçin Ki-Kare
**6.6** Ki-Kare Varsayımlarının Özeti
**6.7** Bağımlı veya Tekrarlı Ölçümler
**6.8** Tek ve İki Yönlü Testler
**6.9** Olabilirlik Oranı Testler
**6.10** Mantel-Haenszel İstatistiği
**6.11** Etki Büyüklükleri
**6.12** Uyum Ölçüsü
**6.13** Sonuçların Yazılması

BÖLÜM 1'DE ölçme verisi (nicel veri olarak da adlandırılır) ile kategorik veri (nitel veri olarak da adlandırılır) arasındaki fark açıklanmıştı. Ölçme verisinde, her gözlem bir puanı temsil eder ve en yaygın istatistikler ortalama ve standart sapmadır. Kategorik veri ele alındığında, veri iki veya daha fazla kategoriye düşen gözlemlerin sıklığını ve her bir kategorideki gözlemlerin sayısını içerir. (Örneğin; "Kaç kişi annesini en iyi arkadaşı olarak belirtmiştir ve kaç kişi babasını en iyi arkadaşı olarak belirtmiştir?")

Bölüm 5'te basit hipotezleri test etmek için binom dağılımının kullanılmasını incelemiştik. Ele alınan durumlar, iki muhtemel çıktıdan birine sahip bireysel olaylarla ilgili durumlarla sınırlıydı ve sadece tekrarlı denemelerde bir çıktının diğerinden istatistiksel olarak daha sık (manidar) çıkma durumunu sorgulamıştı. Bu bölümde aynı sorunun ki-kare testi kullanılarak nasıl daha kısa sorulabileceğini göreceğiz.

Bu bölümde, değerlendirilebilecek durum çeşitlerini arttıracağız. Önce *iki veya daha çok* muhtemel çıktısı olan durumlarla ve daha sonra iki değişkenin bağımsızlıklarıyla ilgili yokluk hipotezlerini test etmekle ilgileneceğiz. Her iki durum için de uygun olan istatistik ki-kare ($\chi^2$) testidir.

**Ki-kare**

**Ki-kare ($\chi^2$)** terimi, istatistikte bazı karışıklıklara yol açan iki farklı anlama gelir. İlk anlamında, dünyamız dışında herhangi bir referansa başvurmadan belirli matematiksel bir dağılımı tanımlamak; ikinci anlamında, $\chi^2$ dağılımı ile yaklaşık aynı dağılıma sahip bir dağılımla ilgili istatistiksel testi belirtmek için kullanılır. Bir kimsenin ki-kareden söz ettiğini duyduğunuzda, genellikle aklında ikinci anlamı vardır. (Bu test, Karl Pearson (1900) tarafından geliştirilmiştir ve diğer testlerden ayırmak için -örneğin Bölüm 18'de söz edilecek olan Friedman Testi ve bu bölümün sonunda açıklanacak olan olabilirlik oranı testi- **Pearson'un ki-karesi** olarak anılır.) Biz terimi her iki şekilde de kullanacağız, ancak anlamı her zaman açık olmalıdır.

**Pearson'ın ki-karesi**

## 6.1 Ki-Kare Dağılımı

**ki-kare ($x^2$) dağılımı**

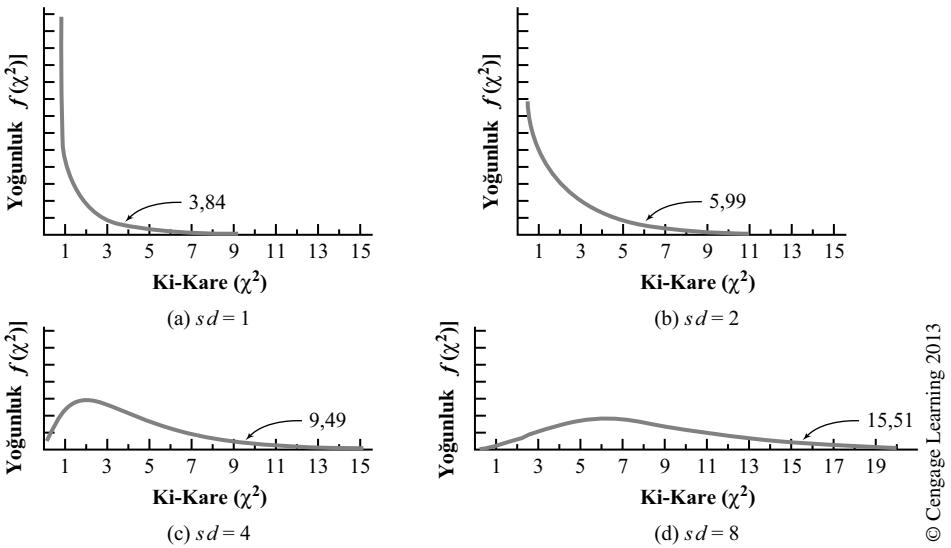
**Ki-kare ($\chi^2$) dağılımı**, aşağıdaki gibi tanımlanır:

$$f(\chi^2) = \frac{1}{2^{\frac{k}{2}}\Gamma(k/2)}\chi^{2[(k/2)-1]}e^{\frac{-(\chi^2)}{2}}$$

Bu karmaşık görünen bir fonksiyondur ve pek çok okuyucu hiçbir cebirsel durumda bu fonksiyon ile çalışmak zorunda olmadığından mutludur. Ancak, $\chi^2$ dağılımının ne anlama geldiğini anlamak için bu fonksiyonun bazı özelliklerini bilmemiz gerekir. İlk olarak ele alınması gereken **gama fonksiyonu** olarak adlandırılan ve normalde *faktöriyel* olarak tanımlanan paydadaki $\Gamma(k/2)$ terimidir. Hatta gama argümanı (k/2) bir tam sayı olduğunda, $\Gamma(k/2) = [(k/2) - 1]$ olacaktır!. Argümanlar her zaman tam sayı olmadığı için gama fonksiyonuna ihtiyaç duyulur. Matematik istatistikçilerinin gama ile ilgili söyleyebilecekleri çok şey olmakla birlikte, bu kitapta daha fazla açıklama yapılmayacaktır.

**gama fonksiyonu**

Ki-kare fonksiyonunun ikinci ve daha önemli özelliği, dağılımın sadece bir parametresinin ($k$) olmasıdır. Geri kalanlar ya sabittir ($e$) ya da ordinatı [$f(\chi^2)$] bulabilmek için istediğimiz $\chi^2$ değeridir. Normal dağılım $\mu$ ve $\sigma$ olmak üzere iki parametreli bir fonksiyonken, $\chi^2$ tek parametresi $k$ olan bir fonksiyondur. Matematikten istatistik âlemine geçtiğimiz zaman, $k$ değeri serbestlik derecesine dönüşmektedir. (Serbestlik derecesi çoğunlukla $\chi^2$'nin altına gelecek şekilde yazılır. Bu sebeple $\chi^2_3$ sembolü "3 serbestlik derecesine sahip ki-kare" şeklinde okunur. Bazı yazarlar ise $\chi^2(3)$ şeklinde gösterirler.)




**Şekil 6.1** Serbestlik derecesi *(sd)* 1, 2, 4 ve 8 olan ki-kare dağılımları. (Oklar, alfa = ,05 için kritik değeri gösterir.)

Şekil 6.1 her biri farklı $k$ değerini temsil eden $\chi^2$ dağılımlarını göstermektedir. Bu şekilden, $k$'deki değişimle birlikte dağılımın dikkate değer bir şekilde değiştiği ve $k$ büyüdükçe dağılımın daha simetrik bir hal aldığı açıkça görülmektedir. Ayrıca her $\chi^2$ dağılımının ortalama ve varyansının da $k$ değerindeki artış ile birlikte arttığı ve $k$ ile doğrudan ilişkili olduğu açıktır. Bu bütün durumlar için aşağıdaki şekilde gösterilebilir:

Ortalama = $k$
Varyans = $2k$

## 6.2 Ki-Kare Uyum İyiliği Testi – Tek Yönlü Sınıflama

**ki-kare testi**

Şimdi sıra, $\chi^2$ dağılımına dayanan ve sıklıkla **ki-kare testi** olarak adlandırılan testi açıklamaya geldi. Öncelikle tek boyutlu ve ardından iki boyutlu tablolara (çapraz tablolar) uygulanan testleri inceleyeceğiz.

Sadece iki kategorili basit, ancak ilginç bir örnekle başlayıp ardından ikiden fazla kategorili bir örnekle devam edeceğiz. İlk örneğimiz *Journal of American Medical Association*'da yayımlanan tedavi edici dokunuşlar üzerine yazılmış bir makaleden alınmıştır (Rosa, Rosa, Sarner ve Barrett, 1998). (İlk örneği kullanmam ile ilgili yasal soruların ortaya atılabileceğinin ve bunların kısaca tartışılacağını başlangıçta belirtmek gerekir. Ancak sınırlılıkları dahî bilgilendirici olduğu için örneği ele almayı tercih ettim.) Bu makaleyi ilginç yapan öğelerden biri, makalenin ikinci yazarı olan Emily Rosa'dır.[1]

[1] Bu çalışmanın ilginç bir özelliği, Annals of Irreproducible Results'ın sponsorluğunda MIT'de gerçekleştirilen "Ig Noble Prize" törenine konuşmacı olarak davet edilen Emily Rosa'dır. Bu, ilginç araştırma çalışmalarını arayan ve önem veren bilim insanlarından oluşan bir gruptur. Ig Noble Prizes, "tekrarlanamayacak olan çalışmaları" ödüllendirmektedir. Emily'nin davet edilmesi bir onur olarak adlandırılır ve Emily'ye tedavi edici dokunuşlara gerçekten inananlardan daha nazik bir yaklaşımdır. Topluluğa http://www.improb.com/ web adresinden ulaşılabilir ve bu bölüme bir ara verme ihtiyacı duyduğunuzda ziyaret etmenizi tavsiye ederim.

**Tablo 6.1** Tedavi edici dokunuşlar üzerine yürütülen deney sonuçları

| | Doğru | Yanlış | Toplam |
|---|---|---|---|
| Gözlenen | 123 | 157 | 280 |
| Beklenen | 140 | 140 | 280 |

© Cengage Learning 2013

Kendisi o zamanlar sadece 11 yaşında ve baş deneycidir. Çalışmanın öz bölümünden bir alıntı yapılırsa, "Tedavi Edici Dokunuşlar (TED) köklerini mistisizmden alan ancak bilimsel bir temeli olduğu iddia edilen ve sıklıkla kullanılan bir bakım uygulamasıdır. TED uygulayıcıları pek çok tıbbî durumu hastanın cildinin üstünde hissedilebilen "insan enerji alanını" ellerini kullanarak istenilen şekilde yönetebilmektedirler." Emily, tedavi edici dokunuş uygulayıcılarını alarak öncelikle gözlerini bağlamakta ve ardından bir elini onların bir elinin üstüne koymaktadır. Eğer tedavi edici dokunuş gerçek bir olgu ise katılımcılar hangi ellerinin Emily'nin elleri altında olduğunu anlamalarına ilişkin kurallara dayanmaktadır. 280 deneme içinde yaklaşık %44'lük bir orana denk gelen 123 tanesi doğru bilmişlerdir. Katılımcıların %50'sinin veya 140 tanesinin şans eseri doğru bilmesi beklenmektedir.

Katılımcıların beklenenden daha düşük bir performans sergilemelerine rağmen, bir testin istatistiksel olarak manidarlığıyla ilgili ilginç bir soru ortaya attığı için bu örnek kullanılmıştır. Bu konuya kısaca değinilecektir. Cevaplamak istediğimiz ilk soru, verinin çıkışındaki şans beklentisinden yola çıkılarak insanların yargılarının tesadüfî tahminden manidar şekilde büyük olup olmadığıdır. Veri Tablo 6.1'de görülmektedir.

Katılımcılar şans düzeyinde cevaplasalar bile cevabın doğruluğunun bir kategorisi diğerine göre daha yüksek sayıda gerçekleşecektir. İstediğimiz, verilen cevabın tesadüfi olmadığı sonucuna varmamızı sağlayacak şans beklentisi varyansının yeterince büyük olup olmadığını sorgulayabileceğimiz bir **uyum iyiliği testi**dir.

**uyum iyiliği testi**

Çok yaygın ve önemli bir formül olan $\chi^2$'nin gözlenen ve beklenen frekansların karşılaştırılmasını içerir. **Gözlenen frekanslar**, adından da anlaşılacağı gibi, veride gerçekten gözlenen frekanslardır – yukarıdaki tablodaki ikinci satırda yer alan değerler. **Beklenen frekanslar**, yokluk hipotezi doğru olduğunda beklediğimiz frekanslardır. Tablo 6.1'in üçüncü satırında beklenen frekanslar görülmektedir. Katılımcıların cevaplarının birbirinden bağımsız olduğunu varsayacağız. ("Bağımsız" teriminin bu kullanımında, katılımcının $k$ olayında verdiği cevabın, $k$–1 olayında verdiği cevaptan bağımsız olduğu vurgulanmaktadır; ancak, test etmeye çalıştığımız şey olan iki farklı kategorideki seçimin eş olasılıklı olduğu anlamına gelmemektedir.)

**gözlenen frekanslar**

**beklenen frekanslar**

280 denemede iki olasılık olduğundan, yokluk hipotezi altında beklediğimiz, 140 doğru ve 140 yanlış seçimdir. Gözlenen seçimlerin sayısını "$G$" ve beklenen seçimlerin sayısını "$B$" harfi ile göstereceğiz. Bu durumda ki-kare formülü, her iki cevap kategorisinin de toplamının alındığı aşağıdaki eşitlikle ifade edilir:

$$\chi^2 = \sum \frac{(G - B)^2}{B}$$

Bu formül sezgisel bir his uyandırmaktadır. Formülün payı ile başlayalım. Eğer yokluk hipotezi doğruysa, gözlenen ve beklenen frekanslar ($G$ ve $B$) birbirine yakın çıkacaktır ve sonrasında karesi alındığında bile pay küçük bir değer alacaktır. Bunun yanı sıra, $G$ ve $B$ arasındaki farkın büyük çıkması bizim beklentimize bağlı olacaktır. Eğer 140 doğru seçimden bahsediyorsak, 5 seçimlik bir fark küçük bir fark olacaktır. Ancak eğer 10 doğru seçim bekliyorsak, 5 büyüklüğünde bir fark önemli olacaktır. Son olarak bu bağlı farklılıkları birleştirebilmek için bütün bu bağıl farkları toplayacağız.

Tablo 6.1'de verilen gözlenen ve beklenen frekanslar kullanılarak elde edilen $\chi^2$ istatistiği aşağıdaki gibi hesaplanır:

$$x^2 = \sum \frac{(G-B)^2}{B} = \frac{(123-140)^2}{140} + \frac{(157-140)^2}{140}$$
$$= \frac{-17^2}{140} + \frac{17^2}{140} = 2(2,064) = 4,129$$

## Ki-Kare Dağılımı Tablosu

$\chi^2$ değerini elde ettikten sonra, bu $\chi^2$ dağılımındaki buna ait olasılığı saptayıp yokluk hipotezinin doğru olup olmadığını belirlememiz gerekir. Bu, tablolaştırılmış standart $\chi^2$ dağılımı kullanılarak yapılabilir.

$\chi^2$ dağılımı tablosu

serbestlik derecesi

**$\chi^2$ dağılımı tablosu**, diğer pek çok istatistik gibi, Bölüm 3'te gördüğümüz **serbestlik derecesi**ne (*sd*) dayanan *tablolaştırılmış* standart normal dağılımdan çok önemli bir şekilde farklılaşmaktadır. Burada olduğu gibi tek boyutlu tablolarda, serbestlik derecesi kategori sayısının bir eksiğine (*k-1*) eşittir. Eğer ,05 düzeyinde yokluk hipotezini ($H_0$) reddetmek istiyorsak, asıl ilgilendiğimiz $\chi^2$ değerimizin, dağılımın üst %5'lik $\chi^2$ değerinden büyük veya küçük olup olmadığıdır. Bu sebeple, belirli amaçlar için ihtiyaç duyduğumuz, sadece her *sd* için %5'lik kesim noktasını bilmektir. Başkaları %2,5'luk veya %1'lik kesme noktasını vd. bilmek isteyebilir; ancak, örneğin, %17'lik kesme noktasını istemeyi hayal etmek zor olabilir. Bu sebeple, Ek $\chi^2$ ve Tablo 6.2'de genel olarak ilgilenilen değerlerin bulunduğu $\chi^2$ tabloları verilmektedir.

Tablo 6.2'ye göz atıldığında, en sol sütunda serbestlik derecelerini bulacaksınız. Diğer sütunların her birinde, o sütunun üstünde belirtilen dağılımın yüzdelik $\chi^2$ kesmesindeki kritik $\chi^2$ değerlerini göreceksiniz. Bu sebeple örneğin, 1 serbestlik derecesindeki dağılımın %5'lik üst kesmesindeki $\chi^2$ değerinin 3,84 olduğu görülmektedir. (Tablo 6.2'de koyu yazılmış olan değerdir.)

Örneğimize geri döndüğümüzde, 1 *sd* için $\chi^2 = 4,129$ değerini elde etmiştik. Dağılımın üst %5'lik dilimindeki $\chi^2$ değerinin 1 *sd* için 3,84 olduğunu da görmüştük. Elde ettiğimiz değer ($\chi^2_{\text{gözlenen}}$) = 4,129, $\chi^2_{1.(,05)}$ = 3,84 değerinden büyük olduğu için yokluk hipotezi reddedilir ve gözlenen frekansların beklenen frekanslardan manidar şekilde farklılaştığı sonucuna varılır. Bu durumda, katılımcılar şansa göre daha az doğrulukla performans göstermişlerdir.

**Tablo 6.2** $\chi^2$ dağılımının üst yüzdelik noktaları

| df | ,995 | ,990 | ,975 | ,950 | ,900 | ,750 | ,500 | ,250 | ,100 | ,050 | ,025 | ,010 | ,005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,00 | 0,00 | 0,00 | 0,00 | 0,02 | 0,10 | 0,45 | 1,32 | 2,71 | **3,84** | 5,02 | 6,63 | 7,88 |
| 2 | 0,01 | 0,02 | 0,05 | 0,10 | 0,21 | 0,58 | 1,39 | 2,77 | 4,61 | 5,99 | 7,38 | 9,21 | 10,60 |
| 3 | 0,07 | 0,11 | 0,22 | 0,35 | 0,58 | 1,21 | 2,37 | 4,11 | 6,25 | 7,82 | 9,35 | 11,35 | 12,84 |
| 4 | 0,21 | 0,30 | 0,48 | 0,71 | 1,06 | 1,92 | 3,36 | 5,39 | 7,78 | 9,49 | 11,14 | 13,28 | 14,86 |
| 5 | 0,41 | 0,55 | 0,83 | 1,15 | 1,61 | 2,67 | 4,35 | 6,63 | 9,24 | 11,07 | 12,83 | 15,09 | 16,75 |
| 6 | 0,68 | 0,87 | 1,24 | 1,64 | 2,20 | 3,45 | 5,35 | 7,84 | 10,64 | 12,59 | 14,45 | 16,81 | 18,55 |
| 7 | 0,99 | 1,24 | 1,69 | 2,17 | 2,83 | 4,25 | 6,35 | 9,04 | 12,02 | 14,07 | 16,01 | 18,48 | 20,28 |
| 8 | 1,34 | 1,65 | 2,18 | 2,73 | 3,49 | 5,07 | 7,34 | 10,22 | 13,36 | 15,51 | 17,54 | 20,09 | 21,96 |
| 9 | 1,73 | 2,09 | 2,70 | 3,33 | 4,17 | 5,90 | 8,34 | 11,39 | 14,68 | 16,92 | 19,02 | 21,66 | 23,59 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

© Cengage Learning 2013

Daha önce de tavsiye ettiğim gibi, bu sonuç bir yokluk hipotezinin nasıl yorumlanacağı ile ilgili bir soruyu ortaya çıkarır. Geleneksel hipotez testi veya Jones ve Tukey'nin (2000) ortaya attığı görüşlerden hangisini ele alacağımız konusunda, farkın şansla elde edilenden daha büyük olduğu sonucuna varabiliriz. Eğer cevap örüntüleri tedavi edici dokunuşların etkililiği lehine ise, tedavi edici dokunuşların gerçek bir olgu olduğu sonucuna varabiliriz. Ancak sonuçlar tam karşıt yönde manidardır ve cevap verenlerin beklenenden çok daha fazla yanlışı olduğundan dokunma etkililiğinin desteklenmesini iddia etmek güçtür. Kişisel olarak, tedavi edici dokunuşların etkililiğini reddedebileceğimiz sonucuna varabiliriz. Ancak burada bir tutarsızlık söz konusudur: Eğer 157 doğru cevabımız varsa, "Bakınız, fark manidardır!" diyebilirim. Ama 157 doğru olmayan cevap olduğunda "Güzel, bu sadece kötü şans ve fark gerçekten manidar değil." diyebilirim. Bu durum, tutarsız davrandığımız için kendimizi suçlu hissetmemize yol açar. Diğer taraftan, katılımcıların manidar şekilde yanlışı olduğunu tahmin eden bir teori de bulunmamaktadır. Bu sebeple destekleyecek gerçek bir açıklama alternatifi yoktur. Eğer terapatik dokunma çalışırsa, insanlar yapmaları gerekeni yapmazlar. (Hayat bazen böyledir!)

Ki-kare dağılımı ile ilgili bir diğer problem, yapabilecek başka bir şey olmadığında mâkul sayıda gözleme dayanmamız ve eksikliği görmezden gelme eğiliminde olmamızdır. Problem $\chi^2$ dağılımının sürekli olmasıdır; ancak burada sözü edilen ki-kare testini kullanırken elde edeceğimiz muhtemel ki-kare değerleri kesikli ve özellikle küçük örneklem büyüklüğünde olacaktır. Örneğin, eğer Emily Rosa deneyini 10 gözlem üzerinden yürütseydi, muhtemel çıktılar 10:0, 9:1, 8:2 vb. olacaktı. Ayrıca bu değerlerin yarısı da diğer yarının yansıması olacaktı. Sonuç olarak, oldukça kesikli bir dağılımı sürekli bir dağılıma uydurmaya çalışmaktayız ve bu uyum çok iyi değildir. Bazı durumlarda, ileride göreceğimiz gibi, bununla ilgili bazı çözüm yolları vardır; ancak bu yaklaşımlar bile problemi tamamen çözmemektedir. Bu probleme ileride tekrar döneceğiz.

Ele aldığım örneğin temel bir probleminin olduğunu kabul ederek başlamıştım. Ki-kare testinin varsayımlarından biri gözlemlerin bağımsızlığıdır. İki okuyucunun bana belirttiği gibi, 280 gözlem 21 puanlayıcıdan elde edilmiştir ve bir puanlayıcının 15'nci puanı 14'ncü denemedeki puanından bağımsız olmayabilir.[2] İnsanların tesadüfî olan bir dizi cevabı üretmede oldukça kötü olduğu bilinen bir gerçektir. Dürüstçe söylemek gerekirse, bunun bir probleme yol açıp açmayacağı konusunda karar vermek zordur. Eğer cevaplayıcılar sol ve sağı değiştirirlerse, cevapların yüzdelerini de 50:50 değiştirmiş olurlar ki bu *deney* hipotezine karşı çalışacaktır. Dahası, cevap deseninin *yokluk* hipotezine karşı nasıl çalışacağını görmek de zor olacaktır. Ancak, "Bağımsızlık problemini görmezden gelebiliriz." diyebilmek de oldukça güçtür. Bu sebeple de örnek, kitapta kullanılmıştır – bu bazı araştırma problemlerinin kolayca ele alınamayacağını da göstermektedir. Bölüm 7'de bu gibi veri kümelerinin analizi için alternatif bir yol üzerinde durulacaktır.

## İkiden Fazla Kategorinin Olduğu Bir Örnek

Pek çok psikolog insanların nasıl karar verdiğiyle özel olarak ilgilenirler ve danışanlarına sıklıkla basit oyunlar oynatırlar. Bu oyunlardan biri, Mahkûmların Çelişkisi adındadır ve ayrı ayrı sorgulanan iki mahkûmdan (oyuncu) oluşmaktadır. Bu durumdaki makul strateji her bir oyuncu için suçu itiraf etmektir. Ancak insanlar sıklıkla en uygun (optimal) davranışı sergilememektedirler. Psikologlar insan davranışlarını en uygun davranışlar ile karşılaştırmak için böyle bir oyunu kullanmaktadırlar. Burada da farklı türde uluslararası

[2] Bu durumu ortaya atan kişilerden biri University of Essex, İngiltere'den Tim Rakow'du. Ne yazık ki diğer okuyucunun ismini artık hatırlayamadığım için kendisine teşekkür edemiyorum.

**Tablo 6.3** Taş/Kağıt/Makas oyununda her bir sembolün gelme sayısı

| Sembol | Taş | Kâğıt | Makas |
|---|---|---|---|
| Gözlenen | 30 | 21 | 24 |
| Beklenen | (25) | (25) | (25) |

© Cengage Learning 2013

bir çocuk oyunu olan ve sıklıkla TKM şeklinde kısaca ifade edilen "Taş/Kâğıt/Makas" ile en uygun davranışların varlığı veya yokluğuna bakılacaktır. Bu oyunda iki oyuncundan her biri bir işaret yapmaktadır. Yumruk bir taşı, düz bir el kâğıdı ve iki parmak makası temsil etmektedir. Taş makası kırar, makas kâğıdı keser ve kâğıt taşı sarar. Yani eğer siz makas yaparsanız ve ben bir taş yaparsam, benim taşım makasınızı kıracağı için ben kazanırım. Ama eğer siz makas yaptığınızda ben kâğıt yaparsam, makas kâğıdı keseceği için siz kazanırsınız. Çocuklar bu oyunu çok uzun süre oynayabilmektedirler. (Bazı yetişkinler de bu oyunu çok ciddîye almaktadır. Bununla ilgili bir makaleye http://www.danieldrezner.com/archives/002022.html adresinden ulaşılabilir. Hatta kendi web sayfaları olan bir Dünya TKM Topluluğu bulunmaktadır.)

Taş/kâğıt/makas oyununda, bu kitabın yazarı gibi tecrübesiz bir insan için optimal stratejinin tamamen tahmin edilemez olduğu açıktır ve her sembol eşit sıklıkla yapılacaktır. Dahası, rakibinizin sizin bir sonraki sembolünüzü tahmin etmemesi için her bir sembolün diğerlerinden bağımsız olması gerekmektedir. Ancak, başka stratejiler de bulunmaktadır. TKM şampiyonalarına katılan yetişkinler dışında, sahadaki en yaygın oyuncular çocuklardır. 75 kişiden oluşan TKM oynayan bir grup çocuğu gözlediğimizi ve her sembolün yapılma sıklığını kaydettiğimizi varsayalım. Bu hipotetik çalışmanın sonuçları Tablo 6.3'te verilmiştir.

Oyuncularımızın her bir sembolü eşit sıklıkta göstermeleri gerekirken, elimizdeki veride, taş sembolünü beklenenin aksine daha sık göstermişlerdir. Ama bu durum sadece şans eseri ortaya çıkmış olabilir. Hatta özellikle tesadüfî olarak sembolleri gösterseniz bile, bir tanesi diğerlerine göre daha sık gösterilmekte olabilir. (Dahası, insanlar tesadüfî hareketlerde oldukça kötüdür.) Bizim istediğimiz, çocukların gösterimlerinin tesadüfî olmadığını ve gerçekten taşı göstermelerinin şans düzeyinden daha büyük olduğuna karar verebilmek için, sapmaların, beklenen şans sapmalarından yeterince büyük olup olmadığını yoklayacak uyum iyiliği testidir.

Tablo 6.3'te, $\chi^2$ istatistiği için bu veride gözlenen ve beklenen frekanslar verilmektedir. Üç kategorimiz olduğundan, daha önce yapılanın biraz genişletilmiş hali olduğuna dikkat ediniz.

$$x^2 = \sum \frac{(G-B)^2}{B}$$
$$= \frac{(30-25)^2}{25} + \frac{(21-25)^2}{25} + \frac{(24-25)^2}{25} = \frac{5^2+4^2+1^2}{25^2}$$
$$= 1,68$$

Bu örnekte üç kategorimiz olduğu için serbestlik derecesi 2'dir. 2 *sd* için $\chi^2$'nin kritik değeri 5,99'dur ve oyuncularımızın her sembolü eşit gösterdiklerine ilişkin şüphe etmemize bir sebep yoktur.

# 6.3 İki Kategorili Değişken: Çapraz Tablo Analizleri

**çapraz tablo**

Önceki örneklerde, tek boyutta sınıflanan (sınıflama değişkeni) veri kümesi üzerinde durmuştuk. Ancak veri kümeleri sıklıkla iki (veya daha fazla) değişkenle sınıflandırılır ve bizim ilgilendiğimiz bu değişkenlerin birbirinden bağımsız olup olmadığıdır. Hatta sıklıkla bir değişkenin dağılımının ikinci bir değişkene bağlı olup olmadığıyla ilgileniriz. Bu durumda bir değişkenin dağılımını diğer değişkenin her düzeyinde göstermek için **çapraz tablo** oluşturulur. Böyle bir duruma iyi bir örnek olarak, ölüm cezasının verilişinde ırksal bir yanlılığın olup olmama durumu verilebilir.

Ölüm cezasının verilişinin, ölünün (veya kurbanın) ırkından etkilenip etkilenmediği ile ilgili yıllar boyunca yapılmış pek çok çalışma vardır. Alıştırma 6.41'de (Ç.N.: 6.34 olabilir.) böyle bir verinin genişletilmiş örneğini görebilirsiniz. Peterson (2001), Unah ve Borger (2001) tarafından 1993-1997 yıllarında Kuzey Carolina'da ölüm cezalarının incelendiği bir araştırmayla ilgili veriyi raporlamaktadır. Tablo 6.4'te kurbanın beyaz olduğu durumlarda sanıkların beyaz ve beyaz olmadığı (çoğunlukla siyahî ve İspanyol kökenli Amerikalılar) durumlar için elde edilen veri görülmektedir. Beklenen frekanslar parantez içinde yazılmıştır.

## Çapraz Tablolar İçin Beklenen Frekanslar

**hücre**

**marjinal toplamlar**
**satır toplamları**
**sütun toplamları**

Çapraz tablodaki beklenen frekanslar, eğer iki değişken birbirinden bağımsızsa bunların bir tablo (örneğimizdeki ırk ve hüküm) oluşturmasını bekleriz. İki boyutlu tabloda verilen satır ve sütun kombinasyonu (**hücre** olarak adlandırılır) için beklenen frekans, hücrenin bulunduğu satır ve sütun toplamlarının çarpılıp toplam veri sayısına ($N$) bölünmesi ile elde edilir. (Bu toplamlar, tablonun marjında bulunduğu için **marjinal toplamlar** olarak bilinir.) $E_{ij}$, $i$ satırında ve $j$ sütunundaki hücre için beklenen frekansı, $R_i$ ve $C_j$ **satır** ve **sütun toplamları**nı ve $N$'de toplam gözlem sayısını göstermek üzere aşağıdaki formül yardımıyla elde edilir:

$$E_{ij} = \frac{R_i C_j}{N}$$

örneğimiz için

$$E_{11} = \frac{284 \times 66}{825} = 22,72$$

$$E_{12} = \frac{284 \times 759}{825} = 261,28$$

**Tablo 6.4** Kurbanın beyaz olduğu durumda sanığın ırkına bağlı hüküm verme

| | Ölüm Kararı | | |
|---|---|---|---|
| **Sanığın Irkı** | **Evet** | **Hayır** | **Toplam** |
| Beyaz olmayan | 33 (22,72) | 251 (261,28) | 284 |
| Beyaz | 33 (43,28) | 508 (497,72) | 541 |
| Toplam | 66 | 759 | 825 |

© Cengage Learning 2013

$$E_{21} = \frac{541 \times 66}{285} = 43,28$$

$$E_{22} = \frac{541 \times 759}{825} = 497,72$$

Bunlar Tablo 6.4'teki parantez içinde gösterilen değerlerdir.

NOT

Beklenen değerler için bu formül, Bölüm 5'te verilen iki bağımsız olayın birlikte gerçekleşme olasılığı formülünden doğrudan elde edilmiştir. Satır$_i$'de yer alan bir gözlemin olasılığı, basitçe o satır toplamının bütün hücrelerin toplamına bölünmesi ile bulunur. Aynı şey Sütun$_j$'de yer alan bir gözlemin olasılığı için de geçerlidir. Eğer gözlemler bağımsız ise, bu iki olasılık çarpılabilir ve ardından beklenen frekansı bulmak için elde edilen sonuç *N* ile çarpılır.

## Ki-karenin Hesaplanması

Her bir hücredeki gözlenen ve beklenen frekansları bulduktan sonra $\chi^2$'nin hesaplanması basittir. Tablodaki bütün hücreleri toplamamıza rağmen, bu zamana kadar kullanılan formülü kullanacağız.

$$x^2 = \sum \frac{(G-B)^2}{B}$$

$$= \frac{(33-22,72)^2}{22,72} + \frac{(251-261,28)^2}{261,28} + \frac{(33-43,28)^2}{43,28} + \frac{(508-497,82)^2}{497,72}$$

$$= 7,71$$

## Serbestlik Derecesi

Elde ettiğimiz $\chi^2$ değerini Ek $\chi^2$'deki değer ile karşılaştırmadan önce serbestlik derecesini bilmemiz gereklidir. Çapraz tablo analizi için serbestlik derecesi, *R* satır ve *C* sütun sayılarını temsil etmek üzere aşağıdaki gibi bulunur:

$$df = (R-1)(C-1)$$

Örneğimizde $R = 2$ ve $C = 2$ olduğundan (2 - 1)(2 - 1) işlemi sonucunda *sd* değerini 1 olarak buluruz. Dört hücremiz olduğu halde *sd* değerinin 1 çıkması garip görünebilir; ancak, satır ve sütun toplamlarını bir kere bildikten sonra gerisine karar vermek için ihtiyaç duyulan sadece bir hücrenin frekansıdır.[3]

## $x^2$'nin Değerlendirilmesi

Ek $\chi^2$'de görüleceği gibi, 1 *sd* için kritik $\chi^2$ değeri 3,84'tür. Elde ettiğimiz değer olan 7,71, kritik değerden büyük olduğu için değişkenlerin birbirlerinden bağımsız olduğunu belirten yokluk hipotezini reddederiz. Bu durumda, ölüm cezasının, sanığın ırkı ile ilişkili olduğu sonuca varabiliriz.

[3] Pearson'ın yanıldığı nokta burasıdır. Serbestlik derecesinin *(R - 1)(C - 1)* yerine *RC - 1* olduğunu düşünmüştür. Fisher, bu hatayı ortaya çıkardığında ise biraz sinirlenmiştir.

Kurban beyaz olduğunda, beyaz olmayan sanıkların ölüm cezasına çarptırılmaları, beyaz olan sanıklara göre daha fazla gözlenmektedir.[4]

## Başka Bir Örnek

İkinci örnekte, Walsh ve arkadaşlarının (2006) anoreksiya (iştahsızlık) tedavisinde antidepresanların kullanımı ile ilgili yürüttükleri çalışma sunulmaktadır.

Uzun zamandır anoreksiya tedavisini başarı ile tamamlayan kızların hastalıklarının nüksetme sebeplerinden biri olarak depresyon öne sürülmektedir. (Normal kilolarına dönmelerine rağmen, bir yıl içinde hastaların %30 ile %50 oranındaki kısmı hastaneye geri dönmüşlerdir.) Çok yaygın bir yaklaşım olarak hem depresyona hem de depresyonun nüksetmemesine iyi gelen Prozac veya benzer bir antidepresan ilaç yazılmaktadır.

**double blind study**

Walsh ve arkadaşları, vücut kütlesinin önemli bir bölümünü koruyan 93 hasta üzerinde çalışmışlardır. Bu hastaların 49'u, bir yıl boyunca Prozac kullanmış, 44'üne ise yalancı bir ilâç (placebo) verilmiştir. Bu çalışmada, hem hastanın hem de çalışma koordinatörlerinin ilâç veya yalancı ilâç verilip verilmediğini bilmediği bir çalışma türü (**double blind study**) uygulanmıştır. Bağımlı değişken, her bir grupta bir yıl boyunca kilolarını başarılı bir şekilde koruyan hasta sayısıdır. Tablo 6.5'te, çapraz tablo şeklinde düzenlenen veri görülmektedir. (Beklenen frekanslar parantez içinde yazılmıştır.)

Bu tablo, ümitlendirici değildir. Öncelikle, hastaların %70'inden fazlasında hastalık nüksetmiştir. Tabloda ayrıca, ilâç grubunun yalancı ilâç grubunu aştığı görülmektedir. Yalancı ilâç grubunda %26,5'e karşı %31,5 gibi bir oranla daha az nüksetme görülmüştür. Bilmek istediğimiz ise bu az nüksetme durumunun manidar olup olmadığı veya sadece şansla böyle bir sonuç çıkıp çıkmadığıdır. (Prozac'ın bir kızın kilosunu koruma yeteneğini düşürdüğü ve bu sebeple bu kişilere yazılmasının gerçekte zarar verici olduğu daha mantıklı görünmektedir.)

Öncelikle marjinal toplamların çarpımlarının veri toplamına bölünmesi ile beklenen frekanslar hesaplanmaktadır. Örneğimiz için hesaplamalar aşağıdaki gibidir:

$$E_{11} = \frac{49 \times 27}{93} = 14,226$$

$$E_{12} = \frac{49 \times 66}{93} = 34,774$$

$$E_{21} \frac{44 \times 27}{93} = 12,774$$

$$E_{22} = \frac{44 \times 66}{93} = 31,226$$

**Tablo 6.5** Prozac ve anoreksiya arasındaki ilişki

| | Çıktı | | |
|---|---|---|---|
| **Tedavi** | **Başarı** | **Tekrarlama (Nüksetme)** | **Toplam** |
| İlaç | 13 (14,226) | 36 (34,774) | 49 |
| Yalancı ilaç (placebo) | 14 (12,774) | 30 (31,226) | 44 |
| Toplam | 27 | 66 | 93 |

© Cengage Learning 2013

[4] Başka veri kümelerinde bulunmasına rağmen, eğer kurban beyaz değilse, ırk ve hüküm arasında manidar bir ilişki yoktur. Kurban beyaz olmadığında ise davaya bakan kişinin cezanın başka bir yolla tahsil edilmesi yoluna sıklıkla gittiği ve ölüm cezası oranının da çok daha az olduğu, araştırmacılar tarafından belirtilmektedir.

Ardından,

$$x^2 = \sum \frac{(G-B)^2}{B}$$
$$= \frac{(13-14,226)^2}{14,226} + \frac{(36-34,774)^2}{34,774} + \frac{(14-12,774)^2}{12,774} + \frac{(30-31,226)^2}{31,226}$$
$$= ,315$$

Bu örnek için de, 2x2'lik bir tablo olduğundan, *sd* aşağıdaki gibidir:
$sd = (R-1)(C-1) = 1$

## $\chi^2$'nin Değerlendirilmesi

Ek $\chi^2$'de görüleceği gibi, 1 *sd* için kritik $\chi^2$ değeri 3,84'tür. Hesaplanan değer olan 0,315, kritik değerden küçük olduğundan değişkenlerin birbirinden bağımsız olduğunu savunan yokluk hipotezi reddedilemez. Bu durumda, bir genç kızın hastalığının nüksetmesinin, kendisine Prozac veya yalancı ilâç verilmesi ile ilişkili olduğuna dair bir kanıt olmadığı kararına varırız. Dikkat edilmesi gereken nokta, "iki değişkenin birbirinden bağımsız olduğu ispatlanmıştır" demiyorum; sadece, "ilişkili olduklarını göstermedik" diyorum. Ancak, farkın gerçekte yalancı ilâç lehine olduğu ve yokluğun altındaki olasılığın oldukça büyük olduğu (ki-kare olasılığı $\geq$ 0,315 = ,57) bulunmuştur. Bu sebeple Prozac'ın kesinlikle istenen etkiyi göstermediği sonucuna varılabilir.

## 2x2'lik Tablolar Özel Durumlardır

2x2'lik tabloların bazı eşsiz özellikleri vardır ve üzerinde çalıştığımız örnek bu özellikleri keşfetmek için bize fırsatlar sunmaktadır.

## Süreklilik İçin Düzeltme

**süreklilik için Yates düzeltmesi**

Özelikle beklenen frekanslar küçük olduğunda, pek çok kitap Tablo 6.4'teki gibi 2x2'lik tablolarda, **süreklilik için Yates düzeltmesi** uygulanmasını tavsiye eder. (Düzeltme, paydaki her değerin karesini almadan önce mutlak değerce 0,5 birim azaltılmasını içerir.) Literatürde Yates düzeltmesi ile ilgili olumlu ve olumsuz pek çok görüş vardır. Ancak, bir sonraki kesimde açıklanacak olan Fisher'ın Koşullu Testinin (Fisher's Exact Test) yaygın kullanılabilirliği Yates düzeltmesini, pek çok bilgisayar çıktısında görünmesi ve standart seçenek olmasına rağmen gereksiz hale getirmektedir.

## Fisher'ın Koşullu Testi (Fisher's Exact Test)

Fisher, Fisher'ın Koşullu Testi olarak adlandırılan testini 1934 yılında Royal Statistics Society toplandısında tanıtmıştır. Good (1999) tarafından rapor edildiğine göre, Fisher'ın sunumunu, kendisini izleyen konuşmacılardan biri "Altın eşek anırması" olarak nitelendirmiştir. O zamanki tartışmalar can sıkıcı olmaktan uzaktı ve hiç şüphe yok ki Fisher'ın bu konuşmacıya aynı nezakette söyleyecek sözü vardı.)

Ayrıntıya girmeden, Fisher'ın bildirisi, marjinal toplamları sabit olarak oluşturulabilecek bütün 2x2'lik tabloları kapsamıştır. Ayrıca, verimizde, tablolara göre, toplamları uçlarda olan olasılıkları belirlemiştir. Eğer bu toplam $\alpha$'dan daha küçük olursa, iki değişkenin bağımsız olduğunu savunan yokluk hipotezi reddedilir ve 2x2'lik tabloyu oluşturan iki değişken arasında istatistiksel olarak manidar bir ilişki olduğu sonucuna varılır. (Bu, **koşullu test** olarak sınıflandırılır; çünkü toplam örneklem sayısından oluşan bütün marjinal

**koşullu test**

toplamlar yerine, gerçekten elde edilen marjinal toplamlardan koşullanmıştır.) Neredeyse bütün istatistiksel yazılımlar verdiği için Fisher'ın Koşullu Testi formülünü burada vermeyeceğim. (SPSS, daha büyük tablolar için de uygun olmakla birlikte, bu istatistiği, bütün 2x2'lik tablolar için vermektedir.) Fisher'ın Koşullu Testinin gerçekte bir ki-kare testi olmadığına dikkat edilmelidir. Genellikle onun yerine kullanılmaktadır. Bu testi kullanmanın avantajı ki-kare dağılımına dayanmadığı için ki-kare varsayımlarına da bağlı olmamasıdır.

Fisher'ın Koşullu Testi, önerildiği günden bu yana tartışmalı bir konu olmuştur. Problemlerden ilki, koşullu bir test olmasıdır (sabit marjinallere dayalıdır). Bazı araştırmacılar eğer deney aynen tekrarlanırsa, farklı marjinal toplamlar elde etmek mümkün olduğunu savunmakta ve bu tabloların hesaplamaya niye dahil edilmediğini sorgulamaktadırlar. Modern bilgisayarların hızı yüksek olmasına rağmen, testi koşulsuz yapmak hesaplamaları karmaşıklaştırmaktadır. Bu durum çözülmesi kolay bir tartışma gibi görünmektedir. Ancak, **sabit ve tesadüfi marjinaller**le ilgili geniş literatür okunduğunda, bunun çözülmesi zor bir tartışma olmasının yanında, karışık olduğu da ortaya çıkacaktır. (Bu konu ile ilgili mükemmel bir tartışma Agresti'de (2002, 95-96) bulunabilir.)

**sabit ve tesadüfi marjinaller**

Fisher'ın Koşullu Testi, tek ve iki-yönlü testlerden dolayı da tartışmalara yol açmaktadır ve karşı kuyrukta çok uç sonuçlar elde edilmektedir. Daha uç bir çıktının ne olduğuna karar vermek yerine, okuyucuların tek- veya iki-yönlü testten hangisini istediklerine karar vermeleri gerekir (İki yönlü testleri şiddetle tavsiye edereim.) ve ardından standart istatistik yazılım sonuçlarını vermelidirler. Bütün yaygın istatistik yazılımları Fisher'ın Koşullu Testinin yanı sıra Pearson'ın ki-karesi ve ilgili test istatistiklerini de vermektedir. Test ki-kare istatistiği üretmemesine rağmen, $p$ değerini vermektedir. Bizim örneğimizdeki $p$ değeri, standart ki-kare testinde olduğu gibi oldukça küçüktür (0,007).

### Fisher'ın Koşullu Testi'ne karşı Pearson'ın Ki-Karesi

Şu an 2x2'lik frekans tabloları için en az iki istatistiksel test bulunmakta olup bir üçüncüsüne de değinilecektir – hangisini kullanmalıyız? Muhtemelen en yaygın çözüm Pearson'ın ki-karesini kullanmaktır; bunun sebebi, her zaman bu şekilde yapılmasından kaynaklanıyor olabilir. Hatta bu kitabın önceki baskılarında yazar olarak ben de koşullu doğasından dolayı Fisher'ın Koşullu Testi'ne karşı olmuştum. Ancak, son yıllarda Fisher'ın Koşullu Testi gibi permütasyon ve tesadüfi testlerde önemli gelişmeler olmuştu. (Bu yaklaşım Bölüm 18'de daha ayrıntılı olarak tartışılacaktır.) Bu tür testlerin mantığı ve basitliği karşısında oldukça etkilendim ve Fisher'ın Koşullu Testi'ni desteklemeye başladım. Her zaman olmamakla birlikte pek çok durumda her iki yaklaşım için varılacak sonuç aynı olacaktır. 2x2'den daha büyük tablolarımız olduğunda, Fisher'ın yaklaşımı modifikasyon olmadan uygulanamamaktadır ve bu gibi durumlarda neredeyse her zaman Pearson'ın ki-karesini kullanmaktayız. (Ancak Howell ve Gordon, 1976'ya ve kitabın web sayfasındaki R programına bakınız.)

## 6.4 Ek Bir Örnek – 4 x 2'lik Bir Desen

Cinsel istismar, toplumumuzdaki ciddî problemlerden biridir ve arkasındaki faktörleri anlamak oldukça önemlidir. Jankowski, Leitenberg, Henning ve Coffey (2002) çocuklukta cinsel istismar ve bir yetişkin olarak sonrasında cinsel istismar arasındaki ilişkiyi incelemişlerdir. 934 lisans düzeyindeki kadının çocukluktaki istismar kategorisi sayılarını (artan şiddet sırasıyla) ve yetişkin cinsel istismar bildirimlerini çapraz tablo ile göstermişlerdir. Sonuçlar, Tablo 6.6'da verilmiştir. Parantez içindeki değerler beklenen frekanslardır.

**Tablo 6.6** Yetişkin cinsel istismarı ile ilişkili çocukluk cinsel istismarı

| İşaretlenen Çocukluk İstismarı Kategori Sayısı | Yetişkin Olarak İstismar | | Toplam |
|---|---|---|---|
| | Hayır | Evet | |
| 0 | 512 (494,49) | 54 (71,51) | 566 |
| 1 | 227 (230,65) | 37 (33,35) | 264 |
| 2 | 59 (64,65) | 15 (9,35) | 74 |
| 3-4 | 18 (26,21) | 12 (3,79) | 30 |
| Toplam | 816 | 118 | 934 |

© Cengage Learning 2013

Cinsel istismar verisi için ki-karenin hesaplaması aşağıdaki gibidir:

$$x^2 = \sum \frac{(G-B)^2}{B}$$

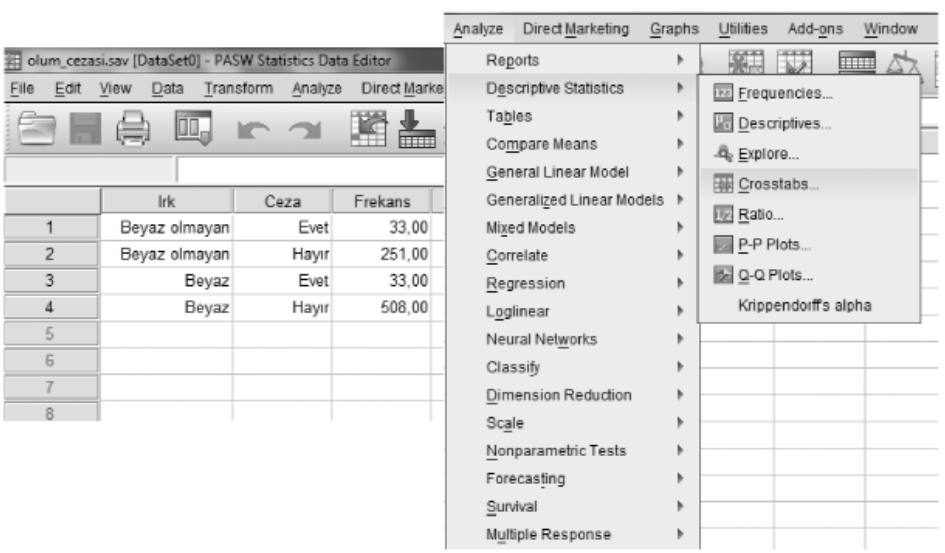
$$= \frac{(512-494,19)^2}{494,19)} + \frac{(54-71,51)^2}{71,51} + \frac{(18-26,21)^2}{26,21} + \frac{(12-3,79)^2}{3,79}$$

$$= 29,63$$

Örnekteki tablo 4x2 biçiminde olduğundan $sd = (4-1)\ x\ (2-1) = 3$'tür. 3 *sd* için kritik ki-kare değeri 7,82 olduğundan yokluk hipotezi reddedilir ve yetişkin cinsel istismarı düzeyinin çocukluktaki cinsel istismar ile ilişkili olduğu sonucuna varılır. Hatta çocukluktaki istismarın şiddeti arttıkça yetişkin istismarının da arttığı görülmektedir. Bu görüşe kısaca tekrar değinilecektir.[5]

## Bilgisayar analizleri

Bu örnek için diğer testlerin yanında Fisher'ın Koşullu Testini de gösterdiği için Unah ve Boger'ın ceza örneği verisi kullanılacaktır. Verinin ilk sütunu ("ırk" olarak adlandırılan), sanıkların ırkına bağlı olarak beyaz (B) ve beyaz olmayan (BO) sembolleri kullanılacaktır.




SPSS Inc.

**Gösterim 6.1a** SPSS Veri dosyası ve analiz için iletişim kutusu

5 Bu veri ile ilgili en rahatsız edici şey, çalışmadaki kadınların yaklaşık %40'ının istismara uğramış olmalarıdır.

İkinci sütun ("ceza" olarak adlandırılan), ölüm cezasının verilip verilmeme durumuna göre "Evet" ve "Hayır" ifadelerini içerir. Son olarak, üçüncü sütun her bir hücre ile ilgili frekansları gösterir. (İlk iki sütun için tutarlı olmak şartıyla sayısal kodlar da kullanılabilir.) Bunun yanı sıra Frekans sütunu olarak adlandırılan ve hücrelerin frekanslarını içeren sütunun belirtilmesi gerekir. Bu, SPSS'te **Data/Weight cases** sekmesinde **Frq** değişkeni için "Weight cases by" komutu ile yapılır. Veri dosyası ve hangi testin kullanılacağına ilişkin iletişim kutusu görseli Gösterim 6.1a'da ve çıktı dosyası Gösterim 6.1b'de görülmektedir.

Bazı sebeplerden dolayı SPSS ki-kare ve ilgili testleri standart (default) olarak göstermemektedir. Testi ve ilgili istatistikleri seçmek için **Statistics** butonunu kullanmak gerekir. Aynı problem tek yönlü Fisher'ın Koşullu Testi için de ortaya çıkar. Beyaz olmayanların, beyazlara göre daha yüksek oranda ölüm cezası aldığı tek yönlü hipotezi test etmek istediğimizi varsayalım. Tek yönlü olasılık, çıktının da vereceği gibi 0,005 olacaktır. Ancak bu durumun yerine, beyazların daha yüksek oranda ölüm cezası aldığı tek yönlü olasılığı test etmek istersem, veriden de görülebileceği gibi, yokluk hipotezini reddedemiyorum. Ama çıktı olarak verilen tek yönlü olasılık bu durumda da 0,005'tir. Vurgulamaya çalıştığım nokta, SPSS'in tek yönlü hangi hipotezi test ettiğini bilmeyişimizden dolayı SPSS sonucunun dikkate alınmamasıdır. (Diğer taraftan, R, olması gerektiği gibi iki farklı olasılık girebilme şansı vermektedir. Bu kitabın web sayfasındaki R dosyalarında bu durumu uygulayabilmeniz için bir örnek bulunmaktadır.)

Gösterim 6.1b'de şu ana kadar üzerinde durmadığımız pek çok istatistik bulunmaktadır. Olabilirlik oranı testi kısaca söz etmemiz gerekenlerden biridir ve ki-kare hesaplanabi-

**Irk*Ölüm Çapraz Tablosu**
**Sayı**

| | | Ölüm | | Toplam |
|---|---|---|---|---|
| | | Evet | Hayır | |
| Irk | Beyaz Olmayan | 33 | 251 | 284 |
| | Beyaz | 33 | 508 | 541 |
| | Toplam | 66 | 759 | 825 |

**Ki-Kare Testi**

| | Değer | sd | Asimt. Man. (2-yönlü) | Man. Düzeyi (2-yönlü) | Man. Düzeyi (1-yönlü) |
|---|---|---|---|---|---|
| Pearson Chi Square | 7,710[a] | 1 | 0,005 | | |
| Continuity Correction[b] | 6,978 | 1 | 0,008 | | |
| Likelihood Ratio | 7,358 | 11 | 0,007 | | |
| Fisher's Exact Test | | | | 0,007 | 0,005 |
| Linear-by-Linear Assoclation | 7,701 | | 0,006 | | |
| McNemar Test | | | | 0,000[c] | |
| N of Valid Cases | 825 | | | | |

[a] Beklenen sayısı 5'ten az olan 0 hücre (%0) vardır. En düşük beklenen sayısı 2,72'dir.
[b] 2x2'lik tablolar için hesaplanır.
[c] Binom dağılım kullanılmıştır.

**Gösterim 6.1b** Ölüm kararı verisi için SPSS çıktısı

© Cengage Learning 2013

**Simetrik Ölçüler**

| | | Değer | Asimtotik Std. Hata[a] | Yaklaşık T[b] | Yaklaşık Manidarlık |
|---|---|---|---|---|---|
| Nominal X. Nominal | Phi | ,097 | ,898 | | ,005 |
| | Cramer's V | ,097 | | | ,005 |
| | Contingency Coefficient | ,096 | | | ,005 |
| Uyum Ölçüsü | Kappa | ,068 | ,026 | 2,777 | ,005 |
| N Geçerli Durumlar | | 825 | | | |

[a] Yokluk hipotezini varsaymadan
[b] Yokluk hipotezini varsayarak ve asimtotik standart hatayı kullanarak

SPSS Inc'dan adapte edilmiştir.

**Gösterim 6.1c** Unah ve Boger'ın verisi için ilişki ölçüleri

**Risk Kestirimi**

| | Değer | %95 Güven Aralığı | |
|---|---|---|---|
| | | Alt | Üst |
| Irk için olasılık oranı (Beyaz olmayan / Beyaz) | 2,024 | 1,221 | 3,356 |
| Ölüm = Evet | 1,905 | 1,202 | 3,019 |
| Ölüm = Hayır | 0,941 | 0,898 | 0,987 |
| Geçerli durum sayısı | 825 | | |

SPSS Inc'dan adapte edilmiştir.

**Gösterim 6.1d** Ölüm cezası verisi için risk kestirimi

lecek basit bir diğer yaklaşımdır. Süreklilik Düzeltmesi Oranı'nın yok sayılmasını tavsiye ederim. Bu basitçe Yates düzeltmesidir ve Fisher'ın testi bunun gerekliliğini ortadan kaldırır. Doğrusallık ilişkisi de görmezden gelinebilir. Bu örnek için uygulanmayacaktır, Bölüm 10'da tekrar değinilecektir. McNemar testi de daha sonra kısaca tartışılacak ve bu örnek için gerekli değildir.

Gösterim 6.1c'deki üç istatistik (phi, Cramer's V ve contingency coefficient) ve Gösterim 6.1d'deki olasılık oranı, bu bölümde daha sonra ele alınacaktır. Bu istatistiklerin her biri etki büyüklüğünü değerlendirmek için kullanılır.

## Beklenen Frekansların Küçüklüğü

**beklenen frekansların küçüklüğü**

Pearson ki-kare testini kullanmanın en önemli gerekliliklerden biri beklenen frekansların büyüklüğüdür. Bu gerekliliğe, süreklilik için düzeltme tartışmalarında özetle değinilmiştir. *Küçük* kavramını tam olarak nasıl tanımladığımızı açıklamadan önce, **beklenen frekansların küçüklüğü**nün neden bu kadar büyük sorunlara yol açtığını incelememiz gerekir.

Belli bir örneklem büyüklüğü için oluşturulabilecek farklı çapraz tablo sayısı ve bu nedenle genellikle farklı ki-kare değeri sayısı sınırlıdır. Eğer sadece birkaç farklı $\chi^2_{\text{gözlenen}}$ değeri oluyorsa, sürekli olan $\chi^2$ dağılımı, kesikli olan istatistik dağılımının mantıklı bir kestirimini sağlayamayacaktır. Ancak, sadece birkaç mümkün $\chi^2_{\text{gözlenen}}$ değeri ile sonuçlanan bu gibi durumlar, bir veya daha fazla hücrede beklenen frekansların küçük olduğu durumlardır. (Bu durum tam olarak şu örneğe benzemektedir: Eğer hilesiz bir metal parayı üç kere atarsanız, üste gelecek yazı sayısı için sadece dört muhtemel hâl vardır ve oluşan örneklem dağılımı kesinlikle yeterli derecede normal dağılıma yaklaşmayacaktır.)

Beklenen frekanslar küçük olduğunda güçlüklerin ortaya çıktığını gördük; ancak ne kadar küçük olduğu sorusu hâlâ duruyor. Var olan görüşler çelişmektedir ve sadece birkaç iddianın biri diğerinden daha üstündür. Muhtemelen en yaygın olanı bütün beklenen frekansların en az beş olması gerektiği ile ilgili olandır. Bu tutucu bir durumdur ve bunu ihlâl ettiğim zaman kendimi suçlu hissetmem. Bradley ve arkadaşları (1979), bilgisayar tabanlı bir örnekleme çalışması yapmışlar, bu çalışmada boyutları 2x2'den 4x4'e kadar değişen tabloları kullanmışlardır. Satır ve sütunların marjinal toplamları büyük oranda çarpık değilse, *toplam* örneklem sayısı 10 olsa bile 1. Tip hatanın 0,06'yı geçmesinin çok küçük oranda çıktığını görmüşlerdir. Camilli ve Hopkins (1979) oldukça küçük beklenen frekanslarla bile 2x2'lik tablolar için toplam örneklem büyüklüğünün sekize eşit veya daha büyük olduğu durumlarda testin çok az 1. Tip tip hata ürettiğini göstermişlerdir. Ancak hem Camilli ve Hopkins (1979) hem de Overall (1980) bu tür testlerin yanlış bir yokluk hipotezini reddedecek düşük güçleri olduğunu belirtmişlerdir. Küçük örneklemlerde, 1. Tip hatadan daha önemli bir problem güç olarak belirtilmektedir.

Fisher'ın Koşullu Testinin önemli avantajlarından biri $\chi^2$ dağılımına dayanmaması ve sürekliliğin olmaması durumundan etkilenmemesidir. Bu test ile ilgili en güçlü savunmalardan biri küçük beklenen frekanslarda da iyi uygulanabilmesidir.

Campbell (2007), küçük beklenen frekanslı 2x2'lik tabloların alternatif ele alınışları üzerine bir çalışma yürütmüştür. Bir hücredeki beklenen frekansın 1 olduğu durumda Egon Pearson tarafından önerilen bir düzeltilmiş ki-kare (adjusted chi-square) istatistiğinin en uygun olduğunu göstermiştir,

$$x^2_{düzeltilmiş} = \frac{x^2 \times N}{N-1}.$$

1'den büyük olan bütün beklenen frekanslar için Fisher'ın Koşullu Testini önermiştir.

## 6.5 Sıralama Düzeyindeki Veri Kümeleri İçin Ki-Kare

Ki-kare, kategorik veri analizi için önemli bir istatistiktir. Ama bazen ihtiyaçlarımız için yeterli olmamaktadır. Eğer bir çapraz tabloya ki-kare uygulanırsa ve ardından bir veya daha fazla satır veya sütunu yeniden düzenlenip ki-kare tekrar hesaplanırsa, tam olarak aynı sonuca ulaşırsınız. Bunun sebebi ki-karenin satır veya sütun sıralamasını göz önünde bulundurmamasıdır.

Satır veya sütunların sırası bir farklılığa yol açıyorsa ne yapılmalıdır? Sıra bilgisi nasıl ele alınıp analizin bir parçası haline getirilebilir? Bu duruma ilişkin ilginç bir örnek, İngiltere'deki Leicester Üniversitesi'nden Jennifer Mahon'dan aldığım çalışmada görülmektedir. .

Ms. Mahon, yeme bozukluklarının tedavisi ile ilgili bir çalışma yapmıştır. Araştırmacı, katılımcıların tedaviye devam etme ve ayrılma benzerliklerini ve bu durumun çocukluklarında yaşadıkları derinden sarsıcı olay sayısıyla ilişkisini incelemiştir. Araştırmacının iddiası, çocukluk döneminde daha çok derinden sarsıcı olay yaşayanların tedaviyi bırakmalarının daha muhtemel olduğudur. Dikkat edilmesi gereken nokta, araştırmacının hipotezinde derinden sarsıcı olayların sayısının bir sıralı değişken olduğu ve bu durumu ki-karenin görmezden geldiğidir. Bu problemin bir çözümü bulunmaktadır; ancak korelasyon açıklandıktan sonra üzerinde durulması daha uygun olacaktır. Bu probleme Bölüm 10'da tekrar dönülecek ve uygun bir yaklaşım üzerinden gösterilecektir. (İstendiği takdirde Bölüm 10, Kesim 10.4'teki açıklamalar okunabilir.) Geniş ölçüde bir korelasyonel tekniği olmasına rağmen, $\chi^2$ ile ilgili açıklamalar yaparken bu noktaya değinilmiştir. Bunun yanı sıra, böyle bir tekniğe bakan bir kişinin de öncelikle bu bölüme bakması mantıklı olacaktır. Çözüme ilişkin açıklamalar, ayrıca kitabın web sayfasındaki Ekler kısmında bulunabilir.

# 6.6 Ki-Kare Varsayımlarının Özeti

Literatürde ki-karenin yanlış kullanımının yaygın olduğu görüldüğü için $\chi^2$'nin varsayımlarını bir arada ele almak önemli olacaktır. $\chi^2$'nin yanlış kullanılmasına ilişkin titiz bir çalışma Lewis ve Burke (1949) tarafından yapılmıştır, bu çalışma ve yansımaları incelenebilir. Yazılmalarının üzerinden 60 yıl geçmesine rağmen, bu makaleler hâlâ geçerliğini korumaktadır. Daha güncel bir çalışma için Deluchi'nin (1983) çalışması incelenebilir; ancak bu çalışma da 25 yıldan daha eskidir. (İstatistikteki bazı bilgi çok hızlı değişmekle birlikte bazı konuların üzerinde çok uzun süre konuşulmaya devam edilir.)

## Bağımsızlık Varsayımı

Bu bölümün başında, *gözlemlerin* birinin diğerinden bağımsız olduğunu varsaymıştık. Bu bölümde *bağımsızlık* kelimesi iki farklı şekilde kullanılmaktadır. *Gözlemlerin* bağımsızlığı $\chi^2$'nin temel varsayımıdır ve bir bireyin seçiminin (örneğin, kahve seçiminin) bir başka bireyin seçimini etkilememesi örneğiyle açıklanabilir. Bağımsızlıktan kastettiğimiz budur. Ayrıca çapraz tabloyu açıklarken ele aldığımız *değişkenlerin* bağımsızlığı üzerinde durmuştuk. Bu durumda bağımsızlık, test edilen durum iken bir önceki örnekte varsayımdır. Yani biz *gözlemlerin* bağımsız olmasını isteriz ve test ettiğimiz hipotez *değişkenlerin* bağımsızlığıdır.

Daha önce değindiğim gibi, tedavi edici dokunuşlar örneğinin kullanılması, gözlemler aynı bireylerden olduğu için *gözlemlerin* bağımsız olmadığını düşünen okuyucular için karmaşık bir örnek olmuştu. Bu görüş, neden hâlâ bu örneği kullandığımızı sorgulamamıza ve neden doğru olduğunu düşünmemize yol açan mükemmel bir nokta olmuştur. Şimdi de Sauerland, Lefering, Bayer-Sandow ve Neugebauer'ın (2003) the *British Society for Surgery of the Hand* dergisinde yayımladıkları makalelerinden bir örnek ele alalım. Yazarlar, bazı hastalıkların öncelikle elin belli bölgelerini etkilediklerini belirtmektedirler. Fakat yaygın yaklaşımın her *n* sayıdaki hastanın elindeki (örneğin) üç noktanın ölçülmesi olduğunu vurgulamaktadırlar. Bu noktada bir problemle karşı karşıyayız. Çünkü eğer ilk ölçümünüz yüksekse muhtemelen ikinci ve üçüncü ölçümler de yüksek olacaktır. Biraz farklı bakarsak, hastanın bütün ölçümlerini yere döksek, sonra bunları gelişigüzel toplasak ciddî bir kayba uğrarız. Sylvia için üçüncü, Bob ve Walker için ikinci ölçümleri almış olabiliriz, Bu durumda hastaların gözlemlerinin eşleşmesini kaybederiz, bu ciddî bir kayıptır. Yani bir birey için karşılaştırma üçüncü ölçümde bitecekken bir başkası için ikincide bitecektir. Gözlemlerle hastaların eşleşmesini kaybedersiniz, ki bu da ciddî bir kayıptır.

Ancak, eğer tedavi edici dokunuşlar örneğindeki gözlemleri ele alır ve bir şapkanın içine atıp karıştırırsanız önemli bir şey kaybetmezsiniz; çünkü ikinci gözlemin aslında yirmi ikinci gözlem olması çok problem değildir. Bir diğer ifadeyle, Sauerland ve arkadaşlarının yaptığı çalışmanın aksine gözlemlerin sırası çalışma ile ilgisizdir. Şu ana kadar beni endişelendiren, bütün gözlemlerin tek bir bireyden geldiği durumda, gözlemlerin bağımsızlığıdır. (85 çocuğun her birinden 3 ölçümün elde edildiği durumda daha çok düşünmem gerekecektir.)

Her şeyi daha fazla karıştıracak risk, TKM oyunundaki stratejilere odaklanmak olabilirdi. Makasın kâğıttan sonra gösterilme sıklığı bir tabloda gösterilebilirdi. Ne sıklıkta sonrasında taş gösteriliyor, ne sıklıkta taştan sonra kâğıt gösteriliyor vd. Bu durumda da ki-kare testi kullanılabilirdi. Çünkü gözlemlerin sırası yokluk hipotezine göre bağımsızdır. Bu da benim test etmeye çalıştığım şeydir.

## Olumsuzluğun (Nonoccurrence) Dâhil Edilmesi

Gerçekleşmeme (Olumsuzluk) durumundan şu ana kadar özel olarak söz edilmemesine rağmen, ki-karenin formülünün çıkarımının doğasında vardır. Muhtemelen en iyisi bir örnek üzerinde açıklanması olacaktır. Kırsal kesimdeki 20 öğrencinin 17'sinin gün ışığından daha fazla yararlanma (GFY) tarafında olduğunu varsayalım. Kentteki 20 öğrenciden sadece 11'inin GFY tarafında olduğunu düşünelim. Belirlemeye çalıştığımız şey, kırsal kesimdeki öğrencilerin kentteki öğrencilere göre manidar şekilde GFY tarafında olup olmadıklarıdır. Bunu test etmenin hatalı bir yöntemi, GFY yanlısı öğrenci sayılarını aşağıdaki tabloda olduğu gibi oluşturmaktır:

| | Kırsal | Kent | Toplam |
|---|---|---|---|
| Gözlenen | 17 | 11 | 28 |
| Beklenen | 14 | 14 | 28 |

**olumsuzluk durumu**

Veri kümesi üzerinden hesaplanan $\chi^2 = 1{,}29$ çıkar ve $H_0$ kabul edilir. Ancak, bu veri tablosu, Lweis ve Burke'ün (1949) adlandırdığı şekliyle **olumsuzluk durumunu** yani olumsuz görüşleri göz önünde bulundurmamaktadır. Bir diğer ifadeyle, GFY'ye karşı olan kırsal kesim ve kentte yaşayan öğrencilerin sayıları dâhil edilmemiştir. Ancak ki-karenin temelleri hem karşıt hem de lehteki görüşlerin dâhil edildiğini varsaymaktadır. Bu sebeple aşağıdaki gibi bir tabloya ihtiyacımız vardır:

| | Kırsal | Kent | Toplam |
|---|---|---|---|
| Evet | 17 | 11 | 28 |
| Hayır | 3 | 9 | 12 |
| | 20 | 20 | 40 |

Bu durumda bir öncekiyle tamamen farklı bir sonuç olan $\chi^2 = 4{,}29$ değerine ulaşılır ve bu sonuç $\alpha = ,05$ düzeyinde manidardır.

Gerçekleşmeme durumunu dâhil etmenin neden gerekli olduğuna dair daha can alıcı bir nokta da, kırsal kesimdeki 2000 öğrenciden 17, kentteki 20 öğrenciden 11, GFY taraftarı olduğunu varsayalım. İki tablonun yorumunun ne kadar farklılaşabileceğini düşünün. Eğer olumsuzluk durumunu dâhil etmeseydik, iki analizin birbirinden ne kadar farklı sonuçlanacağını göremezdik.

Olumsuzluk durumunu ele almama, hem testin geçerliğini olumsuz etkiler hem de $\chi^2$ değeri küçük çıkacağı için $H_0$ hipotezini reddetme durumunu düşürür.

# 6.7 Bağımlı veya Tekrarlı Ölçümler

Bir önceki kesimda, bir çapraz tabloya ait standart ki-kare testinde verinin bağımsız olduğunun varsayıldığı belirtilmişti. Gözlemlerin bağımsızlığı üzerinde durulmuş ve verinin aynı bireylerden gelmesi durumunda dahi bağımsızlıkla ilgili sıklıkla bir sorun çıkmadığı açıklanmıştı. Bu noktada ise bağımsızlıkla ilgili farklı bir problem karşımıza çıkmaktadır. Bununla ilgili iyi bir örnek Denver Üniversitesi'nden Stacey Freedenthal tarafından verilmiştir. Burada kullanılacak olan verinin Dr. Freedenthal'ın verisiyle bir ilgisi yoktur ve bu sebeple onun sonuçlarını da yansıtmamaktadır. Dr. Freedenthal, çocuklardaki yardım arayışı davranışıyla ilgili bir çalışma yürütmüştür. Çalışmasında 70 çocuktan oluşan bir sınıfı ele almış ve çocukların yardım isteme davranışlarını arttıracak bir uygulamanın öncesinde ve sonrasında öğrencilerin yardım isteme davranışlarını kaydetmiştir. Güz döneminde yardım istemeyi ölçmüş, yeni yıl zamanı bir müdahalede bulunmuş ve ardından bahar döneminde aynı çocuklar üzerinde yardım isteme davranışını tekrar ölçmüştür.

Her çocuğu iki kere ölçtüğümüz için ölçümler arasındaki bağımlılığın sonuçlarımızı etkilemediğinden emin olmalıyız. Bunu gerçekleştirmenin bir yolu, her bir çocuğun yıl boyunca ne kadar değiştiğine odaklanmaktır. Bunu yapmak için her bir çocuğun davranışını ayrı ayrı tanımlamalıyız. Böylece, her bir çocuğun güz veya baharda yardım isteyip istemediğini belirlemeliyiz. Bundan sonra, her çocuğun çoklu ölçümleri yerine değişimine odaklanmalıyız. Bağımsızlığın neden önemli olduğunu görebilmek için uç bir durum düşününüz. Eğer tam olarak aynı çocuklar hem güzde hem de baharda yardım isterse ve diğer çocuklar da hiçbir şekilde istemezse, yardım isteme davranışındaki yüzde değişimi sıfır ve standart hata sıfır olacaktır. Ama bir çocuğun güzde yardım isteyip istememesi bahardakinden bağımsız olursa, iki durumun yüzdeleri arasındaki fark hâlâ sıfır olacaktır; ama standart hata bağıl olarak yüksek çıkacaktır. Bir diğer ifadeyle, puanlardaki değişimin standart hatası puanların bağımlılığının bir fonksiyonu olarak değişecektir.

Bu deneyi yürütüp Tablo 6.7'deki çok uç olmayan veriyi elde ettiğimizi varsayalım. Tablo 6.7 bir çapraz dağılım tablosuna benzemekle birlikte bir farkı vardır. Tablo basitçe uygulama sonucunda çocukların nasıl değişip değişmediklerini göstermektedir. Bazı çocukların güzden bahara davranışlarını değiştirmeleri sürpriz değildir. Ve eğer uygulamanın bir etkisi olmazsa (bir diğer ifadeyle eğer yokluk hipotezi doğruysa) "Evet"ten "Hayır"a olan değişim miktarının "Hayır"dan "Evet"e doğru olmasını da bekleriz. Ama eğer uygulama etkiliyse, "Hayır"dan "Evet"e doğru daha çok çocuğun olmasını bekleriz. Test etmeye çalıştığımız durum da tam olarak budur.

**McNemar testi**

Böyle bir durumda **McNemar testi** (McNemar, 1947) olarak adlandırılan testi kullanırız ve basit ki-kare uyum iyiliği testine dönüştürürüz. Burada veri iki köşegen dışı hücrelerin ve beklenen frekansların her birindeki değişiklik gösteren çocukların sayısıdır. Bu aşağıda gösterilmiştir.[6]

$$x^2 = \frac{\sum (G - B)^2}{B} = \frac{(4 - 8{,}0)^2}{8{,}0} + \frac{(12 - 8{,}0)^2}{8{,}0} = 4{,}0$$

Bu değer, *sd*'nin 1 olduğu kritik değer olan 3,84'ten büyük olduğu için manidardır. Uygulamanın başarılı olduğu sonucuna varmak mümkündür.

**Tablo 6.7** Güz ve baharda yardım arayışı davranışı

| Güz | Bahar | | |
|---|---|---|---|
| | Evet | Hayır | Toplam |
| Evet | 38 | 4 | 42 |
| Hayır | 12 | 18 | 30 |
| **Toplam** | **50** | **22** | **72** |

© Cengage Learning 2013

**Tablo 6.8** Çocuklardaki yardım arayışı davranışı üzerine deney sonuçları

| | Hayır → Evet | Evet → Hayır | Toplam |
|---|---|---|---|
| Gözlenen | 12 | 4 | 16 |
| Beklenen | 8,0 | 8,0 | 16 |

© Cengage Learning 2013

6 Bu tam olarak yaygın biçimde kullanılan ve bireylerin büyük bir oranının bir yönden diğer yöne değişiminin manidar olup olmadığını sorguladığımız durumda kullandığımız z testine eşdeğerdir.

## Bir Adım Ötesi

Dr. Freedenthal'ın sorduğu soru, aslında benim cevapladığımdan daha karmaşıktı. Çünkü kendisinin iki durumda da değerlendirilen ancak işlem uygulanmayan bir kontrol grubu da vardı. Deney grubundaki değişimin kontrol grubundaki değişimden daha büyük olup olmadığını test etmek istemişti. Bu aslında beklenenin aksine daha kolay bir teste dönüşmektedir. Bu test Marascuilo ve Serlin'e (1979) atfedilebilir. Her iki işlem grubunda da farklı öğrenciler olduğu için ve bir yöndeki değişim diğer yöndeki değişimden farklı olduğu için veri bağımsızdır. Bu sebeple sütunlarda İşlem Durumu ve satırlarda Artma/Azalma Durumlarının yer aldığı 2x2'lik bir çapraz tablo oluşturup güzden bahara davranışlarında değişiklik olan her gruptaki çocukların kümesini bu tabloya girmeye ihtiyacımız vardır. Böyle bir çapraz tabloya uygulanan ki-kare testi, iki gruptaki değişimin eşit derecede olduğunu savunan yokluk hipotezini test eder. (Bağımsız olmayan frekans veri kümelerinin test edilmesiyle ilgili daha ayrıntılı açıklamaya http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Chi-square/Testing Dependent Proportions.pdf bağlantısından ulaşılabilir.)

# 6.8 Tek ve İki Yönlü Testler

İnsanlar sıklıkla ki-karenin tek- veya iki-yönlü bir test olup olmadığını karıştırırlar. Bu karışıklık, tek- veya iki-yönlü testlerin tanımlanmasının farklılaşmasından dolayı ortaya çıkmaktadır. Eğer $\chi^2$ örneklem dağılımını düşünürsek, $H_0$ hipotezini, $\chi^2$ değerimiz dağılımın en sağ ucunda yer aldığında reddettiğimiz için $\chi^2$'nin tek yönlü bir test olduğunu savunabiliriz. Diğer taraftan, $\chi^2$'nin elde edildiği veri kümesine odaklanırsak iki yönlü bir test olduğunu iddia edebiliriz. Örneğin, eğer bir madenî paranın hilesizliğini test etmede ki-kareyi kullanıyor olsaydık, çok fazla yazı veya çok fazla tura gelmesi durumlarının her ikisi de çok büyük bir $\chi^2$ değerinin çıkmasına yol açacağı için, $H_0$ hipotezini ret edecektik.

Önceki tartışmalar üzerinde, anlambilim üzerine bir tartışma başlatma amacı için değil, aksine ki-kare testinin zayıf noktalarından birini vurgulamak amacıyla durulmuştur. Buradaki zayıflık, test normal bir şekilde uygulandığında, testin yönsüz olmasıdır. Basit bir örnek üzerinden açıklamak gerekirse, bir hayvanın yemeğine katılan artan kinin miktarının yiyeceği daha az çekici hale getirdiğinin gösterilmek istendiğini varsayalım. 90 tane fareye içlerinde değişen miktarlarda kinin bulunan üç kâsedeki yiyeceklerden birini seçmeleri için fırsat sunulduğunu düşünelim ve verinin aşağıdaki gibi olduğunu varsayalım.

| Kinin Miktarı | | |
|---|---|---|
| Küçük | Orta | Geniş |
| 39 | 30 | 21 |

Hesaplanan $\chi^2$ değeri 5,4 olup 2 *sd* için $p < ,05$ düzeyinde manidar değildir.

Veri ile ilgili önemli bir durum, aynı frekanslar (21, 30, 39) ile elde edilebilecek altı farklı dizilişte de aynı $\chi^2$ değerinin elde edileceğidir. Bu sebeple, *çok-kuyruklu* test yapılmalıdır. Bu probleme birkaç sayfa önce, Jennifer Mahon'ın ortaya attığı problem üzerinde durulurken değinilmişti. Bunun çözümü, iki değişken arasındaki korelasyon ölçüsünün tartışıldığı Bölüm 10'da (Kesim 10.4) verilecektir. Bu problemin burada verilmesinin sebebi, okuyucuların cevabı bu bölümde aramalarından dolayıdır.

# 6.9 Olabilirlik Oran Testleri

**olabilirlik oranı**

Kategorik veri kümelerinin analizinde kullanılabilecek bir diğer alternatif yaklaşım **olabilirlik oranı**na dayanır. (Şekil 6.1'de verilen SPSS çıktısı standart Pearson ki-karenin yanı sıra olabilirlik oranını da içermektedir.) Küçük örneklemlerde standart Pearson ki-kare olabilirlik oranı, ki-kareye göre daha iyi uyum sağlamakla birlikte, büyük örneklemler için her iki testte de eşittir (Agresti, 2002). Olabilirlik oran testleri, toplanabilir olma özelliğinden dolayı çapraz tablo için log-lineer modellerde sıklıkla kullanılır. Bu tarz modeller gittikçe daha çok kullanılmaktadır ve siz de bu tarz modellerle en az düzeyde de olsa mutlaka karşılaşırsınız.

Ayrıntıya girmeden olabilirlik oranının genel mantığı basitçe açıklanabilir. Veriyi topladığımızı ve yokluk hipotezini doğrulayacak olasılık veya olabilirliği hesaplandığımızı varsayalım. Ayrıca, başka alternatif hipotezler (veri için gerçekleşme olasılığının yüksek olduğu) için olabilirliği hesaplayalım. Eğer veri $H_0$ yerine alternatif hipotezler için daha olası ise $H_0$ hipotezini reddetme, ancak, diğer alternatiflere kıyasla $H_0$ lehine ise $H_0$ hipotezini kabul etme eğiliminde oluruz. Bu sebeple, olabilirlik oran (bu iki olabilirliğin oranı) sıfır hipotezini değerlendirecek bir imkân sağlar.

Olabilirlik oranlarını kullanarak testler geliştirmek mümkündür. Bu testler bir ve iki boyutlu tabloların analizinde çoğunlukla "en yüksek olabilirlik $\chi^2$ testleri" olarak adlandırılır. Bu testlerin gelişimi için Agresti 2002'ye bakınız.

Tek boyutlu uyum iyiliği durumu için aşağıdaki eşitlik kullanılır:

$$x^2_{(C-1)} = 2\sum G_i \ln\left(\frac{G_i}{B_i}\right)$$

Yukarıdaki eşitlikte $G_i$ ve $B_i$ her bir hücre için gözlenen ve beklenen frekansları gösterir ve "ln"e tabanına göre logaritmayı temsil eder. Buradaki $\chi^2$ değeri, $C$–1 serbestlik derecesi için standart $\chi^2$ tablosu kullanılarak değerlendirilir.

Hesaplamada çapraz tablolar için de temelde aynı olan, $G_{ij}$ ve $B_{ij}$ her bir hücreye ait gözlenen ve beklenen frekansları göstermek üzere aşağıdaki eşitlik kullanılır:

$$x^2_{(R-1)(C-1)} = 2\sum G_{ij} \ln\left(\frac{G_{ij}}{E_{ij}}\right)$$

Beklenen frekanslar, standart Pearson ki-karedeki gibi hesaplanır ve $\chi^2$ değeri $\chi^2$ tablosu kullanılarak $(R\text{-}1)(C\text{-}1)$ serbestlik derecesi için değerlendirilir.

Çapraz tablolar için olabilirlik oranı testinin kullanılmasını örneklendirmek amacıyla ölüm kararı çalışmasındaki veri kümesi kullanılacaktır. Hücreler ve marjinal frekanslar aşağıdaki gibidir:

| | Ölüm Kararı | | |
|---|---|---|---|
| Sanığın Irkı | Evet | Hayır | Toplam |
| Beyaz olmayan | 33 | 251 | 284 |
| Beyaz | 33 | 508 | 541 |
| Toplam | 66 | 759 | 825 |

$$x^2 = 2\sum G_{ij} \ln\left(\frac{G_{ij}}{B_{ij}}\right)$$
$$= 2\left[33\ln\left(\frac{33}{22,72}\right) + 251\ln\left(\frac{251}{261,28}\right) + 33\ln\left(\frac{33}{43,28}\right) + 508\ln\left(\frac{508}{497,72}\right)\right]$$

$$= 2[33(,3733) + 251(-0,0401) + 33(-0,2172) + 508(0,0204)]$$
$$= 2[3,6790] = 7,358$$

Bu sonuç, Şekil 6.1b'de bulunan olabilirlik oranı istatistiği ile benzerdir. 1 *sd* derecesinde $\chi^2$' tablo değeri $\chi^2_{,05}(1)$= 3,84 değerinden büyük olduğu için $H_0$ reddedilecektir.

## 6.10 Mantel-Haenszel İstatistiği

**Mantel-Haenszel istatistiği**

**Cochran-Mantel-Haenszel istatistiği**

**Simpson paradoksu**

Buraya kadar üzerinde durulan iki boyutlu tabloların anlaşılması görece daha kolaydır. Ancak, diğer başka değişkenler üzerinde tekrarlanan veri bulunan 2x2'lik çapraz tablolar sıklıkla karşımıza çıkar. Bu durumlarda üçüncü bir değişken için kontrol etmeye çalıştığımız pek çok değişken bulunur. Stres (*X*) ve zihinsel durum (*Y*) arasındaki ilişkinin incelendiği bir durumda farklı durumlar (*Z*) üçüncü bir değişken olabilir. Veya sanığın ırkı (*X*) ile verilen kararın şiddeti (*Y*) arasındaki ilişkide suçun şiddetine (*Z*) bağlı bir durum incelenebilir. **Mantel-Haenszel istatistiği** (öncesinde Cochran (1954) üzerinde çalıştığı için sıklıkla **Cochran-Mantel-Haenszel istatistiği** olarak bilinir), tam olarak bu gibi durumlarla baş etmek üzere tasarlanmıştır. 1970'li yılların başında Berkeley'deki lisansüstü başvurularında cinsiyet ayrımcılığı üzerine yürütülen bir örnek kullanılarak açıklanmaya çalışılacaktır. Bu örnek **Simpson paradoksu** olarak bilinen bir olayı da gösterdiği için ele alınmıştır. Bu paradoks, 1950'lerin başında Simpson tarafından tanımlanmıştır; ancak yaklaşık yarım yüzyıl öncesinde Yule olarak biliniyordu. (Yule-Simpson paradoksu olarak adlandırılması daha iyi olacaktır.) İki değişken arasındaki ilişkiyi tanımlarken üçüncü bir değişkenin düzeyinin etkisi olarak tanımlanır. Mantel-Haenszel istatistiği, üçüncü bir değişken için 2x2'lik bir çapraz tablonun analizi kontrol edilmek istendiğinde anlam kazanacaktır.

1973 yılında Berkeley'deki California Üniversitesi'nde lisansüstü başvurularındaki cinsiyet ayrımcılığı incelenmiştir (Bickel, Hammel ve O'Connell, 1975). O yıl yapılan başvurular yüzeysel bir şekilde incelendiğinde yaklaşık olarak erkeklerin %45'i kabul edilirken kadınların sadece %30'unun kabul edildiği ortaya çıkmıştır. Bu durum açık bir cinsiyet ayrımcılığının göstergesidir. Ama lisansüstü başvuruların kabulü üniversite kabul birimi tarafından değil bölümler tarafından yapılmaktadır. Bu sebeple, başvuru incelemelerinin bölüm düzeyinde yapılması daha uygun olacaktır. Tablo 6.9'da, Berkeley'deki altı büyük bölümün cinsiyete göre elde edilen başvuru bilgisi bulunmaktadır. (Bu bilgi 101 lisansüstü bölümün verisini yansıtacak niteliktedir.) En büyük bölüm olan A bölümünün verisi şimdilik bir kenara bırakılarak, daha açık bir durum ortaya konulacaktır.

A bölümünün dâhil edilmediği Tablo 6.9'un en son satırına bakıldığında beş bölüm tarafından erkeklerin %36,8'inin kadınların ise %28,8'inin kabul aldığı görülmektedir. Veri üzerinde yürütülen ki-kare testi ile $\chi^2$ = 37,98 değerine ulaşılmıştır; bu değer $H_0$ yokluk hipotezine göre 9'uncu ondalık basamağa kadar 0,00'dır. Bu da erkeklerin, kadınlara kıyasla daha yüksek bir oranda kabul edildiklerinin kanıtı olarak gösterilebilir. Ancak, veri bölümlere göre ayrıldığında, üç bölüm için kadınların daha yüksek bir oranda kabul aldıkları görülürken kalan iki bölümde durum erkekler lehine çok küçük bir oranda yüksektir.

Mantel-Haenszel istatistiği (Mantel ve Haenszel, 1959), her bir bölümdeki veriyi ayrı ayrı incelemek üzerine tasarlanmıştır. Ardından bölümlerden elde edilen sonuçlar toplanacaktır. Bu istatistik her bir bölümdeki ki-kare istatistiğinin bir toplamı olmamasına rağmen, yaklaşık olarak bu şekilde olduğu düşünülebilir. Basitçe bağımsız ki-karelerin birleştirilmesinden daha güçlüdür ve 2x2'lik bireysel tablolarda beklenen frekansların küçüklüğü problemine ise daha az yatkındır (Cochran, 1954).

**Tablo 6.9** Berkeley'deki (1973) lisansüstü bölüm başvuruları
(Yüzdeler parantez içinde verilmiştir.)

| Ana dal | Erkekler | | Kadınlar | |
|---|---|---|---|---|
| | Kabul | Ret | Kabul | Ret |
| A | 512 (%62) | 313 | 89 (%82) | 19 |
| B | 353 (%63) | 207 | 17 (%68) | 8 |
| C | 120 (%37) | 205 | 202 (%34) | 391 |
| D | 138 (%33) | 279 | 131 (%35) | 244 |
| E | 53 (%28) | 138 | 94 (%24) | 299 |
| F | 22 (%6) | 351 | 24 (%7) | 317 |
| B-F Toplamı | 686 | 1180 | 468 | 1259 |
| % B-F Toplamı | %36,8 | %63,2 | %28,8 | %71,2 |

© Cengage Learning 2013

Mantel-Haenszel istatistiğinin hesaplanması, 2x2'lik tablolara dayanmakta, marjinal toplamları veren herhangi bir hücredeki veri, diğer hücrelerdeki değerleri belirler. Bu, her bir bölüm için sadece hücre$_{11}$'deki veri kullanılarak bir istatistik oluşturulabilmesi anlamına gelir. Mantel-Haenszel istatistiğinin pek çok çeşidi olmakla birlikte en yaygın olanı aşağıdakidir:

$$M^2 = \frac{\left(\left|\sum G_{11k} - \sum B_{11k}\right| - \frac{1}{2}\right)^2}{\sum n_{1+k} n_{2+k} n_{1+k} n_{+2k} / n^2_{++k}(n_{++k} - 1)}$$

Yukarıdaki eşitlikte yer alan $G_{11k}$ ve $B_{11k}$, 2x2'lik tablolardaki üst sol hücreye ait gözlenen ve beklenen frekanslardır. Paydada 2x2'lik tabloların her birindeki marjinal toplamlar ile genel toplam bulunmaktadır. Payda, payın varyansını temsil etmektedir. Paydaki -1/2 değeri, daha önce değinilen süreklilik için Yates düzeltmesi ile aynıdır. Yapılan hesaplamalar aşağıdaki gibidir:

$$M^2 = \frac{\left(\left|\sum G_{11k} - \sum B_{11k}\right| - \frac{1}{2}\right)^2}{\sum n_{1+k} n_{2+k} n_{1+k} n_{+2k} / n^2_{++k}(n_{++k} - 1)}$$

$$= \frac{\left(\left|686 - 681{,}93\right| - \frac{1}{2}\right)^2}{132{,}603} = \frac{(4{,}07 - {,}5)^2}{132{,}603} = 0{,}096$$

Bu istatistik 1 *sd* için ki-kare olarak değerlendirilir ve $H_0$ altında olasılığı 0,76'dır. Bölümlerin ayrıldığı durumda elde edilen sonuçla tamamen zıt bir sonuç olarak, kabullerin cinsiyetten bağımsız olduğunu savunan yokluk hipotezi kesinlikle reddedilemez.

**Tablo 6.10** Berkeley verisi için gözlenen ve beklenen frekanslar

| Bölüm | $G_{11}$ | $B_{11}$ | Varyans |
|---|---|---|---|
| A | 512 | 531,43 | 21,913 |
| B | 353 | 354,19 | 5,563 |
| C | 120 | 114,00 | 47,809 |
| D | 138 | 141,63 | 44,284 |
| E | 53 | 48,08 | 24,210 |
| F | 22 | 24,03 | 10,737 |
| B-F Toplamı | 686 | 681,93 | 132,603 |

© Cengage Learning 2013

Mantel-Haenszel istatistiğinin hesaplanmasında, A bölümüne ait veri dışarıda bırakılmıştır ve bunun sebebini muhtemelen merak ediyorsunuzdur. Bu durumun açıklaması bir sonraki kısımda ele alınacak olan olasılık oranına (adds ratio) dayanmaktadır, bunu sonraki bölüme kadar tartışmayacağım. Bu durumun kısa açıklaması ise şöyle yapılabilir: A bölümünde, daha homojen olan diğer bölümlere kıyasla, cinsiyet ve kabul arasında daha farklı bir ilişki bulunmaktadır. Mantel-Haenszel istatistiği, kabul alma durumuna bağlı olarak bölümlerin homojen olması varsayımına dayanır.

Analiz sonuçlarına bakılarak elde edilen sonucun neden ortaya çıktığı, sorulması gereken bir sorudur. Birleştirilmiş veride erkekler lehine açık bir yanlılık söz konusu iken bölümlere göre ayrıldığında böyle bir sonuçla karşılaşılmamıştır. Eğer her bölüm için başvuran adayların yüzdeleri hesaplanırsa, A, B ve D bölümlerinin, başvuranların %50'sinden fazlasını kabul ettiği görülmektedir. Bu bölümler ayrıca erkeklerin daha fazla sayıda başvuru yaptığı bölümlerdir. Diğer taraftan, C ve E bölümlerine daha ağırlıklı olarak kadınlar başvurmuştur. Bu bölümler ise başvuranların üçte ikisini reddetmiştir. Diğer bir ifadeyle, hem erkek hem de kadınlar için daha seçici davranan bölümlere kadınlar daha çok başvurduğu için onların kabul alma oranları toplamda daha düşük çıkmıştır. Bölüm toplamları alındığında ise bu durum görünmemektedir.

## 6.11 Etki Büyüklükleri

Bir ilişkinin "istatistiksel manidarlığı" pratikteki manidarlığı ile ilgili çok fazla bilgi vermez. Hatta istatistiksel olarak bağımsız olmayan iki değişken, bağımsızlıktan yoksun olmanın çok önemli veya dikkate değer olmadığı anlamına gelir. Gerçekte, eğer örneklem büyüklüğünün yeterince büyümesine izin verilirse, neredeyse her iki bağımsızlıktan yoksun değişken istatistiksel olarak manidar çıkacaktır.

O zaman ihtiyaç duyulan, örneklem büyüklüğünü yansıtacak bir veya daha fazla istatistik kullanarak manidarlığı test etmektir. Bir örneklem büyüklüğü etkisini yansıtmak üzere tasarlanmış iki farklı ölçüm türü vardır. Biri Rosenthal (1994) tarafından ortaya atılan ve bağımsız değişkenin grupları veya düzeyleri arasındaki *farklılık* ölçümlerine dayanan **d-ailesi**dir. Örneğin, ileride de kısaca değinileceği gibi, bir ölüm cezası alma olasılığı, beyaz olmayan sanıklar için yüzde beş daha yüksektir. Bir diğer ölçüm türü ise **r-ailesi** olarak adlandırılan iki bağımız değişken arasındaki bir tür korelasyon katsayısını belirtir. Korelasyon kavramı Bölüm 9'da derinlemesine tartışılacaktır; ancak burada da kısaca değinilecektir. r-ailesi ölçümleri sıklıkla **ilişki ölçüleri** olarak adlandırılır.

**d-ailesi**

**r-ailesi**

**ilişki ölçüleri**

### Klâsik Bir Örnek

1988 yılında yürütülen bir çalışmada günlük küçük bir doz alınan aspirinin erkeklerdeki kalp krizini azalttığı yönünde bir bulguya ulaşılmıştır. 22000'in üzerinde doktor, aspirini veya yalancı bir ilâcı bir yıldan daha uzun bir süre boyunca hastalarına vermiş ve kalp krizinin tekrar etme durumunu gözlemiştir. İlgili veri Tablo 6.11'de görülmektedir. Bu desen bir **geleceğe dönük çalışma**dır **(prospective study)**; çünkü öncelikle uygulamalar yapılır (aspirinin verilip verilmemesi) ve ardından gelecekteki çıktılara ulaşılır. Geleceğe dönük çalışmalar sıklıkla **grup çalışması (cohort study)** olarak (çünkü katılımcıların iki veya daha fazla özelliği tanımlanır) veya özellikle tıpta katılımcılar durumlara tesadüfen atandığı için tesadüfî klinik deneme (randomized clinical trial) olarak adlandırılır. Diğer taraftan, sıklıkla durum-kontrol deseni (case-control design) olarak adlandırılan bir **geçmişe dönük araştırma (retrospective study)**, sıklıkla **olay/durum deseni (case-control design)** olarak adlandırılır; daha önce bir kalp krizi geçiren ve geçirmeyen insanları seçer ve ardından geçmişte aspirin alma alışkanlıklarının olup olmadığına bakar.

**geleceğe dönük çalışma (prospective study)**

**grup çalışması (cohort study)**

**geçmişe dönük araştırma olay/durum deseni (case-control design)**

Bu veri için elde edilen $\chi^2 = 25{,}014$'tür, 1 serbestlik derecesinde ve $\alpha = {,}05$ düzeyinde manidardır. Bu sonuç, bir kişinin aspirin alıp almama durumunun kalp krizi geçirmesiyle ilişkili olduğunu gösterir.

**Tablo 6.11** Kalp krizi anında aspirinin etkisi

| | Sonuç | | |
|---|---|---|---|
| | Kalp krizi var | Kalp krizi yok | |
| Aspirin | 104 | 10933 | 11037 |
| Sözde ilâç | 189 | 10845 | 11034 |
| | 293 | 21778 | 22071 |

© Cengage Learning 2013

## *d*-ailesi: Riskler ve Olasılıklar

Özellikle 2x2'lik çapraz tablolar için kategorik veri kümeleriyle ilgili iki önemli kavram risk ve olasılıktır. Bu kavramlar birbiriyle yakından ilişkilidir ve sıklıkla karıştırılırlar, ama temelde çok basittirler.

Aspirin verisi için aspirin grubundaki deneklerin %0,94'ü (104/11037) ve kontrol grubundakilerin %1,71'i (189/11034) çalışma sırasında bir kalp krizi atlatmışlardır. (Eğer sağlığından endişe eden orta yaşlı bir erkek değilseniz, sayılar küçük görünebilir. Ancak önemlidir.) Bu iki istatistik yaygın bir şekilde risk kestirimi olarak tanımlanır; çünkü birinin aspirin kullanmasına veya kullanmamasına bağlı olarak bir kalp krizi geçirme **riski** tahmini olarak tanımlanır. Örneğin, aspirin kullanmayan erkeklerin %1,71'inin çalışmadaki gibi aynı zaman aralığında bir kalp krizi geçireceği tahmin edilebilir. Risk ölçümü etki büyüklüğüne bakmanın kullanışlı bir yolunu sunar.

**risk**

**risk farkı**

**Risk farkı** basit olarak iki oran arasındaki farktır. Örneğimizde bu fark %1,71 - %0,94 = %0,77 olarak hesaplanmıştır. Yani iki durum arasında yaklaşık olarak yüzde dörtte üç puan fark bulunmaktadır. Bir diğer ifadeyle, aspirin alan ve almayan erkeklerin risk farkı yüzde bir puanın dörtte üçü kadardır. Çok büyük bir değer olarak görünmemekle birlikte konunun ciddî bir durum olan kalp krizi olduğu gözden kaçırılmamalıdır.

Risk farkı ile ilgili bir problem, büyüklüğünün genel risk düzeyine bağlı olmasıdır. Kalp krizleri oldukça düşük riskli olaylardır, bu sebeple iki durum arasında çok büyük bir farklılık beklenmemektedir. (Ölüm kararı verisine bakıldığında, ölüm cezası alma olasılığı 5 yüzde puanlık bir risk farkı ile %11,6 olarak ve %6,1 bulunmuştur. Bu değer aspirin çalışmasındaki %0,75'lik farktan çok daha büyük bir etkidir. Bu durum ölüm cezası çalışmasının etki büyüklüğünün fazlalığı anlamına mı gelmektedir? Aslında bu tam olarak risk farkına dayanmaktadır.)

**bağıl risk**

**risk oranı**

Riskleri karşılaştırmanın bir diğer yolu, iki riskin oranı alınarak hesaplanan **bağıl risk** (relative risk) olarak da adlandırılan bir **risk oranı** (risk ratio) oluşturmaktır. Kalp krizi riski için risk oranı aşağıda hesaplanmıştır:

$$RO = Risk_{\text{aspirin yok}}/Risk_{\text{aspirin var}} = \%1,71/\%0,94 = 1,819$$

Bu sonuca göre, aspirin alınmadığında bir kalp krizi geçirme riski aspirin alınmasına kıyasla 1,8 kat daha yüksektir. Ölüm cezası çalışması için elde edilen risk oranı %11,6 / %6,1 = 1,90 neredeyse aspirin çalışmasından elde edilen risk oranıyla aynıdır.

**olasılık oranı**

**olasılık**

Ele alınması gereken üçüncü bir etki büyüklüğü ölçüsü ise **olasılık oranı**dır (odds ratio). İlk bakışta **olasılık** ve **olasılık oranı**, risk ve risk oranı gibi görünmekte ve bu kavramları çok iyi bilen bireyler tarafından bile sıklıkla karıştırılmaktadır. Hatırlanacağı gibi, aspirin alan gruptaki bireylerin bir kalp krizi geçirme riskleri, kalp krizi geçiren kişi sayısının o gruptaki toplam kişi sayısına bölünmesiyle elde edilmiştir (örneğin 104/11037 = 0,0094 = %0,94). Aspirin alan gruptaki bir kişinin kalp krizi geçirmesine ilişkin elde edilecek olasılık, kalp krizi geçiren kişi sayısının kalp krizi geçirmeyen kişi sayısına bölünmesiyle elde edilir (örneğin 104/10933 = 0,0095). Fark (burada çok küçük olmasına rağmen) paydada

kullanılan terimden kaynaklanmaktadır. Risk, toplam örneklem büyüklüğünü kullanır ve bu sebeple o koşuldaki kalp krizi geçiren kişi sayısının oranıdır. Olasılık, kalp krizi geçirmeyen kişi sayısını payda olarak kullanmakta ve bu sebeple kalp krizi geçirenlerin geçirmeyenlere oranıdır. Bu örnek için paydalar birbirine çok benzer olduğu için sonuçlar neredeyse birbirinden farklılaşmamaktadır. Bu her zaman karşılaşılacak bir durum değildir. Jankowski'nin cinsel istismar çalışmasında, çocukken istismara uğramış bir kadının yetişkinlikte uğrama riski 0,40 iken olasılık değeri 0,67'dir. (Olasılık, değeri, sırf benzediği için bir olasılık olarak düşünülmemelidir. Cinsel istismara uğramamanın olasılık değerinin 1,50 çıktığı örnekten de görülebileceği gibi –kadınların cinsel istismara uğramamaları, uğramalarına kıyasla 1,5 kat daha fazladır– olasılık, olasılık değildir.)

İki riski birbirine bölerek bir risk oranı elde edildiği gibi iki olasılık değerini birbirine bölerek olasılık oranı oluşturulabilir. Aspirin örneği için aspirin alınmadığında kalp krizi geçirme olasılık değeri 189/10845 = 0,017'dir. Aspirin alındığında kalp krizi geçirme olasılık değeri ise 104/10933 = ,010'dur. Olasılık oranı, basitçe bu iki değerin oranıdır ve aşağıdaki gibi hesaplanır:

$$OO = \frac{Olasılık/Aspirin\ yok}{Olasılık/Aspirin} = \frac{0,0174}{0,0095} = 1,83$$

Bu durumda, aspirin kullanmadan kalp krizi geçirme olasılık değeri, aspirin kullanıldığında geçirilen kalp krizi olasılık değerinden 1,83 kat daha yüksektir.[7]

Birbirlerine bu kadar benzemelerine rağmen, neden hem olasılık oranı hem de risk oranı hesaplayarak işleri daha karmaşık hale getiririz? Bu çok iyi bir sorudur ve bazı iyi cevapları da bulunmaktadır. Eğer davranış bilimlerinde bir araştırma yapıyorsanız her iki oran da anlaşılmalıdır. Aspirin kullanılmadığında kalp krizi geçirme riskinin ,0171 olduğunu söylediğimizde, bu durumdaki bireylerin %1,7'sinin bir kalp krizi geçireceğini söylüyoruzdur ve oldukça anlaşılırdır. Pek çok insan risk oranlarını sadece bir sebepten dolayı tercih eder. Hatta Sackerr, Deeks ve Altman (1996) risk oranını sadece bu temelde savunurlar – olasılık oranının doğru olmakla birlikte yanıltıcı olduğunu düşünürler.[8] Kalp krizi geçirme durumu olasılık değeri 0,0174 olduğunda söylenen, kalp krizi geçirme olasılık değeri geçirmeme olasılık değerine göre %1,7'dir. Bu yarış atı bahislerinin oynanmasında popüler bir yoldur; ancak tatmin edici değildir. O zaman olasılık oranının ilk etapta ele alınmasının sebebi nedir?

Olasılık oranının lehine söylenebilecek en az iki özellik vardır. Öncelikle, gerçek bir risk oranı hesaplanamadığı durumlarda hesaplanabilir. Kalp krizi geçirmiş ve geçirmemiş birer grup insanın bulunduğu ve bu kişilerin önceden aspirin alıp almadıklarına bakıldığı geçmişe dönük bir araştırmada, risk gerçekten hesaplanamayabilir. Risk gelecek odaklıdır. Eğer 1000 kişiye aspirin verip diğer 1000 kişiye vermezsek, bu insanlara on yıl sonra bakıp kalp krizi geçirme riskini (ve risk oranını) hesaplayabiliriz. Ama eğer 1000 kalp krizi geçirmiş ve geçirmemiş kişiyi alıp geriye dönük bu kişilere bakarsak, gerçekten riski hesaplayamayız, çünkü örneklemi kalp krizi geçiren kimseleri evrendeki normal oranından farklı bir oranda oluşturduk. (Örneklemimizin %50'sini kalp krizi geçirenlerden oluşturduk fakat evrenin %50'si kalp krizi geçirmiş değildir.) Ancak olasılık oranını her zaman hesaplayabiliriz ve kalp krizi geçirmede olduğu gibi *düşük olasılıklı olaylarla ilgili konuşurken,* olasılık oranı,

---

7 Bir olasılık oranı hesaplanırken, hangi olasılık değerinin pay, hangisinin paydada olacağıyla ilgili bir kural yoktur. Uygunluğa bağlı olarak değişir. Mantıklı olduğunda, büyük değeri paya yazmak sonucun 1.0'dan büyük çıkması için tercih edilebilir. Eğer bu örnekte tam tersini yaparsak *OO* = ,546 çıkacaktır ve aspirin kullanıldığında kalp krizi geçirmeye yönelik elde edilen olasılık değerinin kullanılmadığındaki değerin yaklaşık yarısı kadar olduğu sonucuna varılacaktır. Bu basitçe orijinal *OO* değerinin tersidir (,546 = 1/1,83).

8 Pek çok insanın risk oranını olasılık oranına tercih etmeleriyle ilgili açıklamalara http://itre.cis.upenn.edu /~myl/languagelog/archives/004767.html adresinden ulaşılabilir.

risk oranının ne olacağı ile ilgili kestirmelerde genellikle çok iyi bir kestiricidir.[9] (Sackett, Dicks ve Altman (1996), yukarıda açıklananı belirterek olasılık oranının faydalı olduğunu belirtmişlerdir.) Olasılık oranı geleceğe dönük, geçmişe dönük ve kesitler arası desenlerde eşit düzeyde geçerlidir. Bu önemlidir. Ancak geleceğe dönük bir çalışmada risk oranı hesaplanabilir ve risk ile ilgili normalde algılanana yakın bir sonuç çıkar.

Olasılık oranının ikinci önemli avantajı, olasılık oranının "e" tabanına göre logaritmasının [ln(OO)] alınması ile pek çok durumda oldukça kullanışlı olacak bir istatistiği vermesidir. Bu durumlardan ikisi lojistik regresyon ve log-doğrusal modellerdir. Lojistik regresyon bu kitap kapsamında ilerleyen bölümlerde ele alınacaktır. Pek çok kişinin olasılık oranının logaritmik dönüşümünü ilginç bulacağı düşünülmemekle birlikte, bu oldukça önemli bir noktadır.

## 2x*K*'lık Tablolarda Olasılık Oranları

2x2'lik tablolarda olasılık oranını (veya risk oranını) hesaplamak kolaydır. Basitçe iki olasılık değerinin (veya risklerin) oranı alınır. Ama 2x*k* biçimindeki bir çapraz tabloda hesaplamalar biraz karmaşıklaşır; çünkü üç veya daha fazla olasılık değeri vardır ve oranı neyin oluşturduğu açık değildir. Bazı durumlarda olasılık oranları çok fazla anlam ifade etmezler; ancak özellikle bir değişkenin düzeyleri sıralı diziler oluşturduğunda bir anlam ifade ederler. Jankowski'nin cinsel istismar çalışmasındaki veri bu durum için iyi bir örnektir. Tablo 6.12'de bu veri görülmektedir.

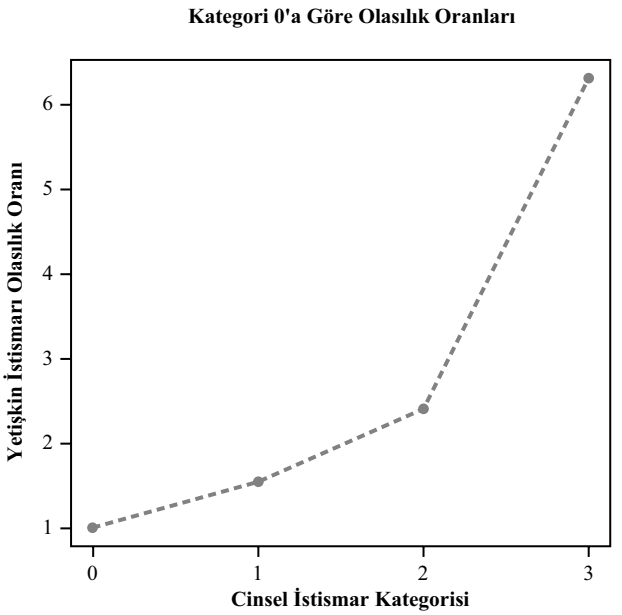
Bu çalışmada, yetişkin istismarının erken çocukluk istismarından nasıl etkilendiğine bakıldığı için çocuklukta istismara uğramamış bireylerin oluşturduğu grup referans grup olarak kullanılabilir. Bu durumda diğer iki grubun bu gruba karşı olasılık oranı alınabilir. Örneğin, çocuklukta istismara uğrayanlardan kategori 1 için olasılık oranı ,163/,106 = 1,54'tür. Bu durumda, bir yetişkin olarak istismara uğramış ve Kategori 1'de yer alan bireylerin değeri Kategori 0'daki bireylere göre 1,54 kattır. Diğer iki grup için olasılık oranı Kategori 0'a göre 2,40 ve 6,29'dur. (İlgili risk oranı değerleri 1,27; 2,13 ve 4,20'dir.) Çocukluktaki cinsel istismarın etkisi, Şekil 6.2'deki gibi bir grafikle gösterildiğinde daha açık bir hal alacaktır. İstismara uğrama olasılık değeri, çocukluktaki cinsel istismarın ciddîyeti arttıkça daha belirgin bir hal almaktadır. (Eğer olasılık yerine risk tercih edilirse, aynı şey risk oranları alınarak da yapılabilir.)

**Tablo 6.12** Yetişkin cinsel istismarının çocukluk cinsel istismarı ile ilişkisi

| | Yetişkinlerde Cinsel İstismar | | | | |
|---|---|---|---|---|---|
| Çocukluktaki Cinsel İstismar Kategorisi | Hayır | Evet | Toplam | Risk | Olasılık |
| 0 | 512 | 54 | 566 | ,095 | ,106 |
| 1 | 227 | 37 | 264 | ,140 | ,163 |
| 2 | 59 | 15 | 74 | ,203 | ,254 |
| 3-4 | 18 | 12 | 30 | ,400 | ,667 |
| **Toplam** | **816** | **118** | **934** | **,126** | **,145** |

© Cengage Learning 2013

9 Olasılık oranı $OO = RR\left(\frac{1-p_1}{1-p_2}\right)$ eşitliğiyle tanımlanır. Eşitlikteki OO = olasılık oranı, BR = bağıl risk, $p_1$ bir gruptaki kalp krizi nüfus oranı ve $p_2$ diğer gruptaki kalp krizi nüfus oranıdır. Bu iki olasılık 0'a yakın olduğunda, birbirini götürecek ve $OO \cong BR$ çıkacaktır.




© Cengage Learning 2013

**Şekil 6.2** Cinsel istismara uğramaya göre olasılık oranları

## 2x2x*K*'lık Tablolarda Olasılık Oranı

2x2 boyutlarında bir tablo için olasılık oranı (veya risk oranı) hesaplandığı gibi bölüm, sosyal sınıf veya gelir düzeyi gibi pek çok durumda da olasılık oranı hesaplanabilir. Bütün kısımlar için olasılık oranı aşağıdaki eşitlik ile elde edilir:

$$OO = \frac{\sum(n_{11k}n_{22k}/n_{..k})}{\sum(n_{12k}n_{21k}/n_{..k})}$$

Berkeley verisi için elimizde aşağıdaki bilgi bulunmaktadır:

| Bölüm | Veri | | $n_{11k}n_{22k}/n_{..k}$ | $n_{12k}n_{21k}/n_{..k}$ |
|---|---|---|---|---|
| B | 353 | 207 | 4,827 | 6,015 |
| | 17 | 8 | | |
| C | 120 | 205 | 57,712 | 50,935 |
| | 202 | 391 | | |
| D | 138 | 279 | 42,515 | 46,148 |
| | 131 | 244 | | |
| E | 53 | 138 | 27,135 | 22,212 |
| | 94 | 299 | | |
| F | 22 | 351 | 9,768 | 11,798 |
| | 24 | 317 | | |
| Toplam | | | 141,957 | 137,108 |

© Cengage Learning 2013

Bölüm B'nin sağındaki iki değer için 353 x 8/585 = 4,827 ve 207 x 17/585 = 6,015 değerleri elde edilir. Diğer satırlar için hesaplamalar benzer şekilde yapılır. Toplam olasılık oranı, bu iki sütunun oranıdır ve aşağıdaki gibi elde edilir:

OO = 141,957/137,108 = 1,03

Olasılık oranı, bir erkek olarak kabul alma olasılık değeri bir kadın olarak kabul alma olasılık değerinin 1,03 katıdır bilgisini verir; yani olasılık değerleri neredeyse eşittir.

Mantel Haenszel istatistiğinin altında yatan; bölüm, sosyal sınıf veya gelir düzeyi gibi değişkenlerde olduğu gibi, olasılık oranlarının bütün kategorilerde karşılaştırılabilir olduğu varsayımıdır. Ancak örneğimizdeki Bölüm A açık bir şekilde uç değerdir. Bu bölümde, kadınların erkeklere olasılık oranı ,35 iken diğer bütün bölümlerde olasılık oranı ,80 ile 1,22 aralığında yani 1,0 civarındadır. Bu bölümün dâhil edilmesi testin varsayımını ihlâl edecektir. Bu durumda, kadınlara karşı ayrımcılığı kontrol ettiğimizde bu bölümü dışarıda bırakmak nihaî sonucu değiştirmez. Bölüm A, erkeklere kıyasla daha fazla kadını kabul etmiştir.

## *r*-ailesi Ölçüleri

Yukarıda değinilen ölçüler, farkın doğrudan hesaplanması ya da risk veya olasılık oranlarının kullanılması yoluyla durumlar arasındaki farkın karşılaştırılmasına odaklanıldığı için *d*-ailesi olarak adlandırılır. Daha eski ve geleneksel bir ölçü birimi olan ve "ilişki ölçüleri" olarak adlandırılan ölçüler ise iki değişken arasındaki korelasyona bakar. Bölüm 9'a kadar korelasyon konusu üzerinde durmayacağız, fakat korelasyon konusunu aşağıdakileri anlayacak kadar bildiğimizi beklemekteyim. Buna rağmen burada kısaca ilgili konulara değineceğim.

Oldukça fazla ilgi ölçüsü bulunmaktadır ve bunların pek çoğu üzerinde durulmayacaktır. Bu konu ile ilgili bilgiye, önemli bir klâsik eser olan Nie, Hull, Jenkins, Steinbrenner ve Bent'in (1970) çalışmasından ulaşılabilir.

## Fi (Phi, $\phi$) ve Cramér *V*

**fi ($\phi$)**

2x2'lik tablolarda, Bölüm 10'da değineceğimiz korelasyon katsayısı iyi bir ilişki ölçüsü olacaktır. Bu katsayı **fi** ($\phi$) olarak adlandırılmakta ve her biri iki kategorili olan iki değişken arasındaki korelasyonu temsil etmektedir (iki kategorili değişken, alabileceği değer sayısı iki olan değişkendir). Eğer Aspirin örneğinde Evet için 1 ve Hayır için 2 veya kalp krizi örneğinde Evet için 1 ve Hayır için 2 şeklinde kodlanır ve iki değişken arasındaki korelasyona bakılırsa (bakınız Bölüm 9 ve 10) fi (phi) değeri elde edilir. (Kodlamada hangi iki değerin kullanıldığının bir önemi yoktur. Önemli olan her iki koşulun da birbirinden farklı ama tutarlı değerler almasıdır.)

$\phi$ değerinin hesaplanmasının kolay bir yolu aşağıdaki gibidir:

$$\phi = \sqrt{\frac{X^2}{N}}$$

Tablo 6.7'deki Aspirin verisi için $\chi^2$=25,014 ve $\phi = \sqrt{25,014/22,071} = ,034$'tür. Bu sonuç çok büyük bir korelasyon değeri değildir; ancak diğer taraftan söz konusu olan hayatî bir olay olduğundan çok küçük bir korelasyon değeri bile anlamlı olabilir.

$\phi$ katsayısı sadece 2x2'lik tablolarda uygulanabilir, Cramér (1946) bunu aşağıdaki formülde olduğu gibi tanımlayarak genişletmiştir.

$$V = \sqrt{\frac{X^2}{N(k-1)}}$$

**Cramér'in *V*'si**

Yukarıdaki eşitlikte $N$ örneklem büyüklüğünü ve $k$ de $R$ (satır) ve $C$ *(sütun)* sayısının küçük olanını belirtir. Bu formül **Cramér'in V'si** olarak bilinir. $k$ = 2 olduğunda iki istatistik eşdeğerdir. Geniş tablolar için yorumu $\phi$ ile benzerdir. $V$ ile ilgili problem, ikiden fazla kategori olduğunda ve bunlar sıralı boyutlarda bulunmadığında yorum yapılmasının oldukça zor olmasıdır.

Pek çok durumda anlamlı bir yorum olmadığını düşündüğüm için kitabın yazarı olarak *r*-ailesi ölçülerinden memnun değilim. *d*-ailesi ölçülerinden biri olan risk oranını kul-

lanmak ve sonucunda eğer aspirin alınmazsa alanlara göre 1,82 kat daha fazla kalp krizi geçirme riskinin olacağını açıkladığı için daha kullanılabilir buluyorum. Pek çok kişi bu cümlenin ne anlama geldiğini anlayabilir. Ancak, fi gibi bir *r*-ailesi ölçüsünü kullanmak ve aspirin alımı ile kalp krizi arasında 0,034 değerinde korelasyonun olduğunu söylemek, kullanışlı bir bilgi sunmamaktadır. (Bu değerin karesini alıp kalp krizi varyansının %0,1'ini aspirin kullanımının karşıladığını söylemek de yetersizdir.) Literatürde bu katsayılarla karşılaşılmasına rağmen çok iyi bir gerekçeniz olmadığı müddetçe *r*-ailesi ölçülerinden uzak durulmasını tavsiye ederim.

# 6.12 Uyum Ölçüsü

Tartışacağımız bir ölçü daha vardır. Aslında daha önceki ölçüler gibi tam olarak bir etki büyüklüğü ölçüsü değildir; ancak özellikle puanlayıcılar arası uyuma bakıldığı durumlarda kategorik veri kümeleri için önemli bir istatistiktir.

## Kappa ($\kappa$) – Bir Uyum Ölçüsü

**kappa ($\kappa$)**

Yaygın olarak Cohen'ın kappası (Cohen, 1960) olarak bilinen ki-kareye dayanmayan ancak çapraz tablonun kullanıldığı önemli bir istatistik **kappa**'dır **($\kappa$)**. Bu istatistik, puanlayıcılar arası uyumu ölçer ve sıklıkla puanlamaların güvenirliğinin incelendiği durumlarda kullanılır.

Dikkate değer düzeyde klinik tecrübesi olan bir puanlayıcıdan 30 yetişkin ile görüşme yapmasının istendiğini ve bu bireyleri (1) problemli davranış yok, (2) içselleştirilmiş problemli davranışa sahip ve (3) dışsallaştırılmış problemli davranışa sahip olacak şekilde sınıflandırmasının beklendiğini varsayalım. Çalışmayı gözden geçiren herhangi bir kimsenin yapılan ölçümlerin güvenirliği ile ilgili kaygıları olabilir – bu puanlamanın yazı-tura atmaktan daha iyi olduğunu nasıl bilebiliriz? Kontrol etmek amacıyla aynı süreci ikinci bir puanlayıcının yürütmesini ve aynı yetişkin bireyleri puanlamasını ister, daha sonra, bu iki puanlayıcı arasındaki uyum ve uyumsuzlukları, bir çapraz tabloda gösterebiliriz. Elde edilen verinin Tablo 6.13'teki gibi olduğunu varsayalım.

Parantez içindeki değerleri şimdilik görmezden gelelim. Bu tabloda, Puanlayıcı I, sütun 1'de toplam olarak görülebileceği gibi 16 yetişkini problem yok olarak sınıflandırmıştır. Bu 16 bireyin 15'i için Puanlayıcı II de aynı görüşü bildirmiştir; ancak aynı zamanda 1 tanesinin içselleştirilmiş problemli davranış ve 0 tanesinin dışsallaştırılmış problemli davranış gösterdiği şeklinde sınıflandırmıştır. Tablonun köşegeninde yer alan değerler (15, 3, 3) iki puanlayıcı arasındaki uyumu gösterirken köşegen dışında yer alan değerler uyumsuzluğu göstermektedir.

**uyum yüzdesi**

Bu veri için basit bir yaklaşım **uyum yüzdesi** hesaplanmasıdır. Bu istatistik için ihtiyaç duyulan tek şey toplam 30 durumun 21'inde (15+3+3) puanlayıcıların aynı kararı vermiş ol-

**Tablo 6.13** İki puanlayıcı arasındaki uyum

| | Puanlayıcı I | | | |
|---|---|---|---|---|
| Puanlayıcı II | Problem Yok | İçselleştirmiş | Dışsallaştırmış | Toplam |
| Problem Yok | 15 (10,67) | 2 | 3 | 20 |
| İçselleştirilmiş | 1 | 3 (1,20) | 2 | 6 |
| Dışsallaştırmış | 0 | 1 | 3 (1,07) | 4 |
| Toplam | 16 | 6 | 8 | 30 |

© Cengage Learning 2013

maları bilgisidir. Ardından 21/30 = ,70 = %70 uyum elde edilir. Ancak, bu ölçü ile ilgili bazı problemler vardır. Örnekteki yetişkinlerin çoğunluğu problemli davranış göstermemektedir ve her iki puanlayıcı da (doğru bir şekilde) Problem Yok sınıflaması yapıp diğer sınıflamalardan uzak durmuştur. Problem Yok olasılığı Puanlayıcı I için 16/30 = ,53 ve Puanlayıcı II için 20/30 = 0,67 olarak kestirilebilir. İki puanlayıcının Problem Yok şeklinde aynı sınıflamayı yapma olasılıkları ,53 x 0,67 = 0,36 olacak ve bu durumda tamamen şans eseri Problem Yok şeklinde 30 puanlama için bu değer 0,36 x 30 = 10,67 bulunacaktır.

Cohen (1960), kappa olarak bilinen şans faktöründen arındırılmış bir uyum ölçüsü öne sürmüştür. Kappayı hesaplayabilmek için öncelikle puanlamaların birbirinden bağımsız olduğunun varsayılması ve köşegendeki her hücre için beklenen frekansların hesaplanması gerekir. Bu hesaplamalar, standart ki-kare testindeki beklenen değerlerin hesaplanmasıyla aynı şekilde yapılır. Örneğin, her iki puanlayıcının Problem Yok şeklinde yaptıkları sınıflamaların beklenen frekansı (20 x 16)/30 = 10,67'dir. İçselleştirilmiş kategorisi için bu değer (6 x 6)/30 = 1,2 ve dışsallaştırılmış kategorisi için bu değer (4 x 8)/30 = 1,07'dir. Bu değerler tabloda parantez içinde gösterilmiştir.

Kappa, $f_G$ köşegendeki gözlenen frekansları ve $f_B$ köşegendeki beklenen frekansları göstermek üzere aşağıdaki eşitlikle tanımlanır:

$$\kappa = \frac{\sum f_G - \sum f_B}{N - \sum f_B}$$

Burada,

$$\sum f_G = 15 + 3 + 3 = 21$$

ve

$$\sum f_B = 10,67 + 1,20 + 1,07 = 12,94$$

Buradan

$$\kappa = \frac{21 - 12,94}{30 - 12,94} = \frac{8,06}{17,06} = ,47$$

bulunur. Burada dikkat edilmesi gereken, daha önce %70 düzeyinde hesaplanan uyumdan daha düşük bir katsayının hesaplanmış olmasıdır. Şans faktörü düzeltildikten sonra %70 uyum yerine %47 uyum elde edilmiştir. Kappa sıklıkla şans faktörü düzeltilmiş uyum ölçüsü olarak adlandırılır.

Kappanın formülünü incelediğinizde, düzeltmenin uygulandığını görebilirsiniz. Formülün payında, uyum sayısından şansla ortaya çıkacak uyum sayısı çıkarılmıştır. Paydada ise toplam puanlama aynı ölçüde azaltılmıştır. Ardından, şans faktörü düzeltilmiş değerlerin oranı hesaplanır.

Cohen ve arkadaşları kappanın manidarlığı için istatistiksel bir test geliştirmişlerdir (Bakınız Fleiss, Nee ve Landis, 1979). Fakat kappanın manidarlığı nadiren incelenir. Eğer manidarlığı sorgulanacak kadar düşük bir kappa değeri elde edilirse, puanlayıcılar arasındaki uyumum eksikliği ciddî bir problemdir.

## 6.13 Sonuçların Yazılması

Jankowski'nın cinsel istismar çalışmasını örnek olarak alacağız. Eğer bu sonuçlar yazılsaydı, muhtemelen aşağıdakilere benzer cümlelerin yazılmasını isterdiniz:

> Çocukluktaki cinsel istismarın yetişkinlikteki cinsel istismara etkisinin incelendiği bir araştırmada, lisans düzeyinde öğrenim gören 934 kadına eğer varsa çocukluktaki cinsel istismarın düzeyi ve yetişkin olarak cinsel istismara uğrayıp uğramama durumları

sorulmuştur. İstismarın şiddeti, katılımcıların cevaplarına bağlı olarak istismar kategori sayısı olarak alınmıştır. Toplanan veri ve çocukluktaki cinsel istismarın şiddeti ile yetişkin cinsel istismarın tekrar ettiği ortaya çıkmıştır. Yetişkin ve çocukluk istismarı arasındaki ki-kare testi sonucu $\chi^2_3 = 29,63$ olup $p < ,05$ düzeyinde manidardır. Bir yetişkin olarak sadece bir kategoride çocuklukta istismara uğramış bir birey için bağıl risk, çocuklukta istismara uğramayan gruba göre 1,27'dir. Risk oranı, çocukluktaki istismarın şiddeti arttıkça 2,13 ve 4,2'ye yükselmektedir. Bir çocuk olarak cinsel istismara uğramak sonraki dönemde bir yetişkin olarak cinsel istismarın güçlü bir göstergesidir.

## Anahtar Terimler

Ki-kare ($\chi^2$) (Giriş)
Pearson'ın ki-karesi (Giriş)
Ki-kare dağılımı ($\chi^2$) (6.1)
Gama fonksiyonu (6.1)
Ki-kare testi (6.2)
Uyum iyiliği testi (6.2)
Gözlenen frekanslar (6.2)
Beklenen frekanslar (6.2)
$\chi^2$ tablosu (6.2)
Serbestlik derecesi (*sd*) (6.2)
Çapraz tablo (6.3)
Hücre (6.3)
Marjinal toplamlar (6.3)
Satır toplamı (6.3)
Sütun toplamı (6.3)
Double blind çalışması (6.3)
Süreklilik için Yates düzeltmesi (6.3)
Koşullu test (6.3)
Sabit ve tesadüfi marjinler (6.3)
Küçük beklenen frekans (6.4)
Olumsuzluk (6.6)
McNemar Testi (6.7)
Olabilirlik oranları (6.9)
Mantel-Haenszel testi (6.10)
Cochran- Mantel-Haenszel (CMH) (6.10)
Simpson Paradoksu (6.10)
*d*-ailesi (6.11)
*r*-ailesi (6.11)
İlişki ölçüleri (6.11)
Geleceğe dönük çalışma (6.11)
Tesadüfi klinik deney (6.11)
Geçmişe dönük çalışma (6.11)
Durum-kontrol deseni (6.11)
Risk (6.11)
Risk farkı (6.11)
Risk oranı (6.11)
Bağıl risk (6.11)
Olasılık oranı (6.11)
Olasılık (6.11)
Fi ($\phi$) (6.11)
Cramer'in *V*'si (6.11)
Kappa ($\kappa$) (6.12)
Uyum yüzdesi (6.12)

## Alıştırmalar

6.1 Bir psikoloji bölüm başkanı, bölümdeki öğretim üyelerinin bazılarının daha popüler olduğundan şüphelenmektedir. Psikolojiye giriş dersi için 10:00, 11:00 ve 12:00 saatlerinde sırasıyla Profesör Anderson, Klatsky ve Kamm tarafından verilmek üzere üç farklı şube açılmıştır. Derslere kayıt yaptıran öğrenci sayıları aşağıdaki gibidir:

| Profesör Anderson | Profesör Klatsky | Profesör Kamm |
|---|---|---|
| 32 | 25 | 10 |

Yokluk hipotezini kurunuz, uygun ki-kare testini yapınız ve sonuçları yorumlayınız.

6.2 Geçerli bir deney tasarımı yönünden (aritmetiksel hesaplamanın aksine), Alıştırma 6.1 ve bu bölümde kullanılan örnekler arasında önemli bir farklılık bulunmaktadır. Alıştırma 6.1'deki veri, bölüm başkanının cevaplanmasını istediği soruyu gerçek anlamda cevaplayamaz. Buradaki sorun nedir ve deney nasıl iyileştirilebilir?

6.3 Clark ve Clark (1939) tarafından yapılan klâsik bir çalışmada, Siyahi Amerikalı çocuklara beyaz ve siyah oyuncak bebekler gösterilerek oynamak için tercih edecekleri oyuncağı

seçmeleri istenmiştir. 252 çocuktan 169'u beyaz oyuncak bebeği seçerken 83'ü siyah oyuncak bebeği tercih etmiştir. Buradan bu çocukların davranışları ile ilgili hangi sonuca varabiliriz?

6.4 Clark ve Clark'ın çalışmasından 30 yıl sonra Hraba ve Grant (1970), Alıştırma 6.3'teki çalışmayı tekrarlamıştır. Çalışmalar benzer olmalarına rağmen tam olarak eşdeğer değillerdir ama ilginç sonuçlara ulaşmışlardır. Hraba ve Grant'in sonuçlarına göre 89 Siyahi Amerikalı çocuktan 28 tanesi beyaz bebeği tercih ederken 61 tanesi siyah bebeği tercih etmiştir. Veri üzerinde uygun ki-kare testini çalıştırarak, sonuçları yorumlayınız.

6.5 Alıştırma 6.3 ve Alıştırma 6.4'teki veri kümelerini iki çapraz tablo üzerinde birleştirerek uygun testi çalıştırınız. İki yönlü çapraz tablonun belirttiği soru ile Alıştırma 6.3 ve Alıştırma 6.4'ün belirttiği sorular nasıl farklılıklar gösterir?

6.6 Sigaranın insanlar üzerinde çok çeşitli hastalıklara yol açtığı bilinmektedir; diğer bütün etkilerinin arasında doğurganlığa etkisi ile ilgili kanıtlar vardır. Weinberg ve Gladen (1986) sigara içmenin etkilerini ve kadınların kolay hamile kalmasını incelemişlerdir. 586 hamile kalmayı planlamış olan kadın alınarak bu kişilere korunmayı bıraktıktan kaç adet döngüsü sonrasında hamile kaldıkları sorulmuştur. Weinberg ve Gladen ayrıca bu kadınları sigara kullanan ve kullanmayan şeklinde de sınıflandırmıştır. Bu durumla ilgili veri aşağıdaki gibidir:

| | 1 Döngü | 2 Döngü | 3+ Döngü | Toplam |
|---|---|---|---|---|
| Sigara içenler | 29 | 16 | 55 | 100 |
| Sigara içmeyenler | 198 | 107 | 181 | 486 |
| Toplam | 227 | 123 | 236 | 586 |

© Cengage Learning 2013

Sigara içmek, kolay hamile kalmayı etkiler mi? (Alıştırmaya cevabınız ne olursa olsun, sigara içmek doğum kontrol yöntemi olarak önerilmez.)

6.7 Süreklilik için düzeltme tartışılırken, çalışmanın kopyasının toplamda aynı satır veya sütun üretmesi olarak açıklanan sabit marjinler konusu üzerinde durulmuştu. Aşağıda istenenlere uygun birer çalışma örneği veriniz:

a. hiçbir marjin toplamı sabit değildir .

b. bir küme marjin toplamı sabittir.

c. satır ve sütun olmak üzere iki kümede makul derecede sabittir. (Zor soru.)

6.8 Howell ve Huessy (1981), ikinci sınıftaki çocukların yaygın olarak Dikkat Eksikliği Hastalığı (DEH) ile ilişkili davranışları gösterip göstermediklerini belirlemek amacıyla bir puanlama ölçeği kullanmışlardır. Aynı çocuklar dördüncü ve beşinci sınıftayken de aynı şekilde gözlenmişlerdir. Bu çocuklar dokuzuncu sınıfın sonuna geldiklerinde, araştırmacılar okul kayıtlarını incelemiş ve hangi çocukların yardımcı İngilizce aldıklarına bakmışlardır. Aşağıdaki veri kümesinde, herhangi bir zamanda DEH davranışı gösteren çocuklar bir grupta birleştirilmişlerdir (DEH olarak adlandırılan grup):

| | Yardımcı İngilizce Alan | Yardımcı İngilizce Almayan | |
|---|---|---|---|
| Normal | 22 | 187 | 209 |
| DEH | 19 | 74 | 93 |
| | 41 | 261 | 302 |

© Cengage Learning 2013

İlköğretim sırasındaki davranışlar lisedeki sınıflamayı etkilemekte midir?

6.9 Alıştırma 6.8'deki veriyi ki-karenin örneklem büyüklüğünün bir fonksiyonu olarak nasıl değiştiğini göstermek için kullanınız.

a. Her bir hücredeki değerleri iki katına çıkarıp ki-kareyi tekrar hesaplayınız.

b. Hipotez testlerindeki örneklem büyüklüğünün rolü ile ilgili (a) seçeneğinde elde ettiğiniz değer ne gibi bir bilgi sunmaktadır?

6.10 Alıştırma 6.8'de çocuklar ikinci, dördüncü ve beşinci sınıfların herhangi birinde DEH davranışı gösterip göstermediklerine göre sınıflandırılmışlardır. Eğer sınıflar birbirinden ayrılırsa, aşağıdaki veri elde edilir:

| | Asla | 2. | 4. | 2. ve 4. | 5. | 2 ve 5. | 4 ve 5. | 2., 4. ve 5. |
|---|---|---|---|---|---|---|---|---|
| Yardımcı İng. Alan | 22 | 2 | 1 | 3 | 2 | 4 | 3 | 4 |
| Yardımcı İng. Almayan | 187 | 17 | 11 | 9 | 16 | 7 | 8 | 6 |

© Cengage Learning 2013

a. Ki-kare testini uygulayınız.

b. Beklenen frekansların küçüklüğü problemini yok sayarsanız, ne gibi bir sonuca ulaşırsınız?

c. Bu kadar küçük beklenen frekanslar ile kendinizi ne kadar rahat hissediyorsunuz? Eğer rahat değilseniz, bu sorunun üstesinden nasıl gelirsiniz?

6.11 Vermont Eyaleti parlamentosu 2000 yılında eşcinsel veya lezbiyen çiftler arasında evliliğe izin veren bir yasayı onaylamıştır. Bu her iki taraf için oldukça ciddî konuları doğuran çekişmeli bir tartışma olmuştur. Bu konuya kadın ve erkeklerin nasıl baktıklarına ilişkin farklılıklar cinsiyete göre oylara bakıldığında görülebilir. İlgili veri aşağıdaki gibidir. Bu veriden yola çıkarak nasıl bir sonuca varılabilir?

| | Oy | | |
|---|---|---|---|
| | Evet | Hayır | Toplam |
| Kadın | 35 | 9 | 44 |
| Erkek | 60 | 41 | 101 |
| Toplam | 95 | 50 | 145 |

© Cengage Learning 2013

6.12 Stresin fiziksel sağlık durumunu etkilediği uzun süredir bilinmektedir. Visintainer, Volpicelli ve Seligman (1982) yürüttükleri çalışmada, 60 denemede kaçınılmaz şoka maruz kalan farelerin, 60 deneme kaçınılabilir şok alan veya 60 denemede hiç şok almayan farelere göre, yerleştirilen bir tümörü ret etmelerinin daha az düzeyde olduğu hipotezini incelemişlerdir. Elde edilen veri aşağıdaki gibidir:

| | Kaçınılmaz Şok | Kaçınılabilir Şok | Şok Yok | Toplam |
|---|---|---|---|---|
| Ret | 8 | 19 | 18 | 45 |
| Ret Yok | 22 | 11 | 15 | 48 |
| | 30 | 30 | 33 | 93 |

© Cengage Learning 2013

Bu sonuçlardan Visintainer ve arkadaşları nasıl bir sonuca ulaşabilirler?

6.13 Yetişkinlerdeki yeme bozukluklarının incelendiği bir çalışmada, Gross (1985) katılımcılara kilo almayı mı, kilo kaybetmeyi mi yoksa kilolarını korumayı mı tercih ettiklerini sormuştur. (Dikkat: Gross'un örneklemindeki kızların sadece %12'si normal ağırlıklarının %15 üstündedir – "kilolu" tanımlaması için kullanılan kesme noktasıdır.) Elde edilen veri kızlar için ırka göre (Siyahî Amerikalılar ve beyazlar) bölündüğünde, aşağıdaki değerler elde edilmiştir (diğer ırklar örneklem büyüklükleri çok küçük olduğu için ele alınmamıştır):

| | Kaybedenler | Koruyanlar | Kazanalar | Toplam |
|---|---|---|---|---|
| Beyaz | 352 | 152 | 31 | 535 |
| Siyahi Amerikalı | 47 | 28 | 24 | 99 |
| | 399 | 180 | 55 | 634 |

© Cengage Learning 2013

a. Bu veriden yola çıkarak nasıl bir sonuca ulaşabilirsiniz?

b. Irk değişkeni görmezden gelindiğinde, kızların kendi ağırlıklarına karşı tutumlarıyla ilgili nasıl bir sonuca varılabilir?

6.14 Alıştırma 6.8'deki verinin analizini olabilirlik oranı yaklaşımını kullanarak yapınız.

6.15 Alıştırma 6.10'daki verinin analizini olabilirlik oranı yaklaşımını kullanarak yapınız.

6.16 Alıştırma 6.10'daki tablonun 2. satırındaki veriye tek yönlü ki-kare testinin uygulanması mümkündür. Bu analizi yapsaydınız hangi hipotezi test ediyor olurdunuz? Bu hipotezin Alıştırma 6.10'da test edilenden farkı nedir?

6.17 Bir grup katılımcıya Pazartesi Gecesi Futbolundan hoşlanıp hoşlanmadıkları sorulmuş, ardından bir maç izlettirilmiş ve tekrar sorulmuştur. Burada incelenen, bir maç izlemenin, insanların düşüncelerini değiştirip değiştirmediğini görmektir. 80 katılımcıdan 20 tanesi olumludan olumsuza, 5 katılımcı ise olumsuzdan olumluya doğru değişen görüş bildirmiştir. (Diğer katılımcılar fikirlerini değiştirmemişlerdir.) Maç izlemenin düşüncelerin değişmesinde sistematik bir etkisi var mıdır? [Bu McNemar (1969) tarafından tavsiye edilen bir testtir.]

a. Testi çalıştırınız.

b. Bu testin yokluk hipotezini nasıl test ettiğini açıklayınız.

c. Bu durumda test, futbol maçı seyretmenin düşünce değişimi üzerinde ciddî bir etkisi olup olmadığına ilişkin soruyu cevaplamamıştır. Bunun sebebi nedir?

6.18 Pugh (1983) jürinin tecavüz vakalarında nasıl karar verdiğine ilişkin bir çalışma yürütmüştür. 358 kişinin tecavüz davası sunulmuştur. Bu davaların yaklaşık olarak yarısında maktul de zanlı gibi suçlu görülürken diğer yarısında suçsuz bulunmuştur. Jüri kararları aşağıdaki tabloda verilmektedir. Nasıl bir sonuca ulaşırsınız?

| Hata | Suçlu | Suçlu Değil | Toplam |
|---|---|---|---|
| Az | 153 | 24 | 177 |
| Çok | 105 | 76 | 181 |
| Toplam | 258 | 100 | 358 |

© Cengage Learning 2013

6.19 Aşağıda verilen Görünüm 6.2'deki SPSS çıktısında, Alıştırma 6.13'teki verinin analiz sonuçları yer almaktadır.

a. Alıştırma 6.13a'daki cevabın doğruluğunu kanıtlayınız.

b. Satır ve sütun yüzdelerini yorumlayınız.

c. "Asymp. Sig." (asimptotik manidarlık) olarak adlandırılan değerler nedir?

d. Katsayıları yorumlayınız.

**IRK*HEDEF Çapraz Tablosu**

| | | Hedef | | | Toplam |
|---|---|---|---|---|---|
| | | Kazanç | Kayıp | Aynı | |
| IRK Afrikalı-Amer | Sayı | 24 | 47 | 28 | 99 |
| | Beklenen Sayı | 8,6 | 62,3 | 28,1 | 99,0 |
| | % IRK içi | 24,2% | 47,5% | 28,3% | 100,0% |
| | % HEDEF içi | 43,6% | 11,8% | 15,6% | 15,6 : |
| | % Toplam | 3,8% | 7,4% | 4,4% | 15,6% |
| Beyaz | Sayı | 31 | 352 | 152 | 535 |
| | Beklenen Sayı | 46,43 | 336,7 | 151,9 | 535,0 |
| | % IRK içi | 5,8% | 65,8% | 28,4% | 100,0% |
| | % HEDEF içi | 56,4% | 88,2% | 84,4% | 84,4% |
| | % Toplam | 4,9% | 55,5% | 24,0% | 84,4% |
| Toplam | Sayı | 55 | 399 | 180 | 634 |
| | Beklenen Sayı | 55,03 | 99,0 | 180,0 | 634,0 |
| | % IRK içi | 8,7% | 62,9% | 28,4% | 100,0% |
| | % HEDEF içi | 100,0% | 100,0% | 100,0% | 100,0% |
| | % Toplam | 8,7% | 62,9% | 28,4% | 100,0% |

SPSS Inc. çıktısından adapte edilmiştir.

**Gösterim 6.2**

*(devam edecek)*

| Ki-Kare Testleri | | | |
|---|---|---|---|
| | Değer | *sd* | Asimtotik Mnd. (2-sided) |
| Pearson Ki-Kare | 37,229* | 2 | ,000 |
| Olabilirlik Oranı | 29,104 | 2 | ,000 |
| N Geçerli Durumlar Sayısı | 634 | | |

[a]0 Hücre (,0%) beklenen frekans 5'ten küçük. Minimum beklenen frekans 8,59.

| Simetrik Ölçüler | | | |
|---|---|---|---|
| | | Değer | Yaklaşık Manidarlık |
| Nominal | Fi | ,242 | ,000 |
| Nominal | Cramer'in V | ,242 | ,000 |
| | Contingency Katsayısı | ,236 | ,000 |
| N geçerli durumlar sayısı | | 634 | |

a. Sıfır hipotezini varsaymadan.

b. Standart hata kullanarak sıfır hipotezi varsayarak.

**Gösterim 2** *Devamı*

6.20 Aşağıda Tablo 6.7'den alınan kalp krizi ve aspirin ile ilgili daha büyük bir veri kümesi bulunmaktadır. Burada kalp krizi geçirip geçirmeme arasındaki farkın yanı sıra ölümcül olan ve olmayan krizler arasındaki farka da bakılmıştır.

| | Miyokardiyal Enfarktüs | | | |
|---|---|---|---|---|
| | Ölümcül Kriz | Ölümcül Olmayan Kriz | Kriz yok | Toplam |
| Yalancı İlaç | 18 | 171 | 10,845 | 11,034 |
| Aspirin | 5 | 99 | 10,933 | 11,037 |
| Toplam | 23 | 270 | 21,778 | 22,071 |

a. Hem Pearson'ın ki-karesini hem de ki-kare olabilirlik oranını hesaplayınız. Sonuçları yorumlayınız.

b. Sadece ilk iki sütundaki veri kümelerini kullanarak (kalp krizi geçiren bireylerin verisini) hem Pearson'ın ki-karesini hem de ki-kare olabilirlik oranını hesaplayıp sonuçları yorumlayınız.

c. Ölümcül olan ve olmayan atak sütunlarını birleştirerek, birleştirilmiş sütun ile Atak Yok sütunu için hem Pearson'ın ki-karesini hem de ki-kare olabilirlik oranını hesaplayınız. Sonuçları yorumlayınız.

d. (b) ve (c)'deki Pearson'ın ki-karesini ve ardından (b) ve (c)'deki ki-kare olabilirlik oranını toplayıp her bir sonucu (a)'daki sonuçlar ile kıyaslayınız. Bu sonuçlar size ki-karenin bölümlere ayrılmasıyla ilgili ne ifade etmektedir?

e. Bu sonuçlar size aspirin ve kalp krizi arasındaki ilişki ile ilgili ne gibi bilgi vermektedir?

6.21 Alıştırma 6.20'deki sonuçlar için Cramer'in V'sini ve ilgili olasılık oranlarını hesaplayarak yorumlayınız.

6.22 Alıştırma 6.10'daki veri için olasılık oranı hesaplayınız. Bu değer ne anlama gelmektedir?

6.23 Tablo 6.4 için olasılık oranını hesaplayınız. Bu oran sizin üzerinde çalışılan konu ile ilgili olan anlayışınıza ne katmıştır?

6.24 Alıştırma 6.11'deki veri için SPSS veya başka bir istatistiksel program kullanarak Fisher'ın Koşullu Testini hesaplayınız. Bu Pearson'ın ki-karesi ile ilişkili olasılık ile nasıl kıyaslanabilir?

6.25 Dabbs ve Morris (1990) erkeklerdeki yüksek testosteron düzeyi ile anti sosyal davranışlar arasındaki ilişki üzerine askeri kayıtlardaki arşiv verisini incelemişlerdir. Normal Testosteron grubundaki 4016 erkeğin %10,0'unda, Yüksek Testosteron grubundaki 446 erkeğin ise % 22,6'sında yetişkin suçluluğu (adult delinquency) bulunmaktadır. Bu ilişki manidar mıdır?

6.26 Alıştırma 6.25'teki olasılık oranı kaçtır? Bu değer nasıl yorumlanır?

6.27 Alıştırma 6.25'te açıklanan çalışmada Normal Testosteron grubunun %11,5'i ve Yüksek Testosteron grubunun %17.9'unda çocukluk suçluluğu geçmişi bulunmaktadır.

a. Bu iki değişken arasında manidar bir ilişki var mıdır?

b. Bu ilişkiyi yorumlayınız.

c. Elde edilen bu sonuç, Alıştırma 6.25'ten elde edilen bilgiyi ne ölçüde arttırmıştır?

6.28 Birtakım duygusal problemleri olan gençlere bireyselleştirilmiş bakımın etkisinin incelendiği bir çalışmada, Burchard ve Schaefer (1990), bakıcıların bir günde 40 yetişkinin her birinin belirli davranışlarını var veya yok şeklinde puanlamalarını istemişlerdir. Puanlayıcı güvenirliğini kontrol etmek için iki puanlayıcının her yetişkini puanlamaları istenmiştir. Aşağıdaki tabloda "aşırı kötü söz kullanma" davranışı için hipotetik veri sunulmaktadır.

| | Puanlayıcı A | | |
|---|---|---|---|
| Puanlayıcı B | Var | Yok | Toplam |
| Var | 12 | 2 | 14 |
| Yok | 1 | 25 | 26 |
| | 13 | 27 | 40 |

© Cengage Learning 2013

a. Bu puanlayıcılar için uyum yüzdesi kaçtır?

b. Cohen'ın kappası kaçtır?

c. Kappanın uyum yüzdesinden fark edilir derecede düşük olmasının sebebi nedir?

d. İki istatistiği birbirinden daha da uzaklaştırmak için toplam veri sayısını ($N$) 40 tutarak ham veriyi düzenleyiniz. Bunu nasıl yaparsınız?

6.29 Okula giden pek çok çocuk, cinsel istismarı azaltmak amacıyla uygulanan bir program çerçevesinde "iyi dokunuş-kötü dokunuş" modeli üzerinden birtakım eğitimler almaktadır. Gibson ve Leitenberg (2000) 818 üniversite öğrencisinden daha önce böyle bir eğitim alıp almadıklarına ve istismara uğrayıp uğramadıklarına dair veri toplamıştır. Eğitim alan 500 öğrenciden 43 tanesi istismara uğradıklarını belirtmiştir. Eğitim almamış 318 öğrenciden ise 50 tanesi istismara uğradıklarını söylemişlerdir.

a. Bu veri cinsel istismarı önlemek için uygulanan programın etkililiği ile ilgili ikna edici bir durum sunmakta mıdır?

b. Bu verinin olasılık oranı kaçtır ve bu değer neyi ifade etmektedir?

6.30 Bu kitabın web sayfasında bulunan ve kitabın Appendix Veri Kümesi bölümünde açıklanan Mireault.dat isimli veri kümesi, Mireault ve Bond (1992) tarafından ebeveyn ölümünün etkilerini araştırmak üzere üniversite öğrencilerinden toplanmıştır. Kritik değişkenleri bir süreliğine bir kenara bırakıp öğrencilerin dağılımına bakınız. Veri kümesi öğrencilerin cinsiyet ve üniversitede öğrenim gördükleri fakülte bilgilerini içermektedir.

a. Herhangi bir istatistiksel yöntem ile Cinsiyete karşı Fakülteyi tablo ile gösteriniz.

b. Fakülte kaydının cinsiyetten bağımsız olduğu hipotezin ki-kare testi sonucu nedir?

c. Elde edilen sonuçları yorumlayınız.

6.31 Mireault'un verisindeki değişkenlere bakıldığında, habersiz olduğumuz sistematik farklılıkların olmadığından emin olmalıyız. Örneğin, eğer ölen ebeveynin cinsiyetinin

diğer bir değişkenin çıktısını açıklayacak önemli bir değişken olduğunu bulursak, ölen ebeveyn cinsiyetinin, çocuğun cinsiyeti ile bir şekilde ilişkili olduğunu ve bu iki değişkenin etkilerinin karıştığını sonradan keşfetmek istemeyiz.

a. Bu iki değişken üzerinde bir ki-kare testi yapınız.

b. Elde edilen sonuçları yorumlayınız.

c. Eğer ebeveynin cinsiyeti, çocuğun cinsiyeti ile ilişkili ise ebeveyn cinsiyeti ile diğer değişkenler (örneğin, çocuğun depresyon düzeyi) arasındaki ilişkiye yönelik yapılan yorum ne anlama gelir?

6.32 Zuckerman, Hodgins, Zuckerman ve Rosenthal (1993) 500 bireye istatistiksel konularla ilgili bir dizi sorunun olduğu bir anket uygulamışlardır. Bir soruda, bir okuyucu bir araştırmacıyı örneklemi küçük olduğu için 1.Tip hata yapma olasılığının yüksekliği ile ilgili uyarmaktadır. Araştırmacı ise bu duruma karşı çıkmaktadır. Sorulan "Araştırmacı haklı mıdır?" Cevap verenlerin oranı öğrenci, yardımcı doçent, doçent ve profesör olmak üzere ayrılmıştır. Toplam katılımcı sayısı ve her bir kategorideki birey sayısı aşağıdaki gibidir:

| | Öğrenci | Yardımcı Doçent | Doçent | Profesör |
|---|---|---|---|---|
| Oran | 0,59 | 0,34 | 0,43 | 0,51 |
| Örneklem büyüklüğü | 17 | 175 | 134 | 182 |

© Cengage Learning 2013

(Not: Bu veri 17 öğrencinin %59'unun araştırmacı ile aynı görüşte olduğu anlamına gelmektedir. Gerçek frekanslar hesaplandığında, en yakın tam sayıya yuvarlayınız.)

a. Okuyucuya mı araştırmacıya mı katılıyorsunuz? Niçin?

b. (a) seçeneğindeki soruya verdiğiniz cevaba göre aynı görüşü paylaşmadığınız kişinin yaptığı mantık hatası nedir?

c. Bu problemi bir ki-kare testi için nasıl uygun hal getirirsiniz?

d. Bu veri size cevap veren gruplar arasındaki farklılıkla ilgili hangi bilgiyi vermektedir?

6.33 Bir önceki soruda yer alan Zuckerman ve arkadaşlarının çalışması, fakülte üyelerinin öğrencilere göre bu tarz sorulara olumsuz cevaplar verdikleri için doğruluklarının daha az olduğunu savunmaktadırlar. ("Burada bir şaşırtmaca olabilir.") Bu tarz bir hipotezi test etmek için nasıl bir araştırma deseni kurardınız?

6.34 Hout, Duncan ve Sobel (1987) evli çiftlerin göreceli cinsel tatminleri ile ilgili bir çalışma yapmışlardır. 91 evli çiftin her bireyine "asla"dan "neredeyse her zaman"a kadar değişen dörtlü likert tipi bir ölçekte "Cinsellik benim ve eşim için eğlencelidir." ifadesine cevap vermeleri istenmiştir. İlgili veri aşağıda yer almaktadır:

| | Kadınların Görüşü | | | | |
|---|---|---|---|---|---|
| Erkeklerin Görüşü | Asla | Çok nadir | Çok sık | Neredeyse her zaman | Toplam |
| Asla | 7 | 7 | 2 | 3 | 19 |
| Çok nadir | 2 | 8 | 3 | 7 | 20 |
| Çok sık | 1 | 5 | 4 | 9 | 19 |
| Neredeyse her zaman | 2 | 8 | 9 | 14 | 33 |
| Toplam | 12 | 28 | 18 | 33 | 91 |

© Cengage Learning 2013

a. Bu veri ile ilgili nasıl bir analiz yapılabilir? İki cevabın birbirinden bağımsız olup olmadığı bilgisinden daha fazlasına ulaşılmak istendiği unutulmamalıdır. Eşlerden birinin cevabı yüksekse diğerinin de aynı şekilde olması beklenmektedir.

b. Hem Pearson ki-kare hem de ki-kare olabilirlik oranını kullanınız.

c. Cramer'in V'si ile hangi sonuca ulaşılmıştır?
d. Olasılık oranları nasıldır?
e. Kappa değeri sonucu nedir?
f. Son olarak, "Asla" ve "Çok nadir" kategorileri ile "Çok sık" ve "Neredeyse her zaman" kategorileri birleştirilirse ne olur? Sonuçlar daha açık bir hal alır mı ve hangi koşullar bunu anlamlı kılar?

6.35 Bir önceki soruda eşlerin cinselliği eğlenceli bulma derecelerinin aynı olup olmaması üzerinde durulmuştur. Ancak kadınların "eğlence" ile ilgili farklı bir kesme puanına sahip olduklarını varsayalım. Örneğin, kadınların "Çok nadir" ve "Neredeyse her zaman" düşünceleri, erkeklerin algıladıklarından daha yüksek olabilir. (Kadınlar "ayda birkaç kere" için "Çok sık" tanımlamasını yaparken erkekler "ayda birkaç kere" için "Çok nadir" tanımlaması yapabilirler.) Bu durum sonucu nasıl etkiler? Bu, erkekler ve kadınlar arasındaki bir uyumsuzluk ortaya çıkarır mı?

6.36 Aşağıdaki veri Ramsey ve Shafer'ın (1996) çalışmasından alınmıştır. Bu örnekte mahkûmların avukatları eğer kurban siyah değil de beyaz olduğu durumlarsa, siyah mahkûmların ölüm cezası almalarının daha olası olduğunu göstermeye çalışmışlardır. Avukatlar cezalardaki sistematik ayrımcılığı kanıtlamak için uğraşmışlardır. Georgia eyaleti bu duruma katılmış ancak beyazlara karşı işlenen suçların siyahlara karşı işlenen suçlara göre daha ciddî olduğunu iddia etmiş ve bu sebeple suçlamalardaki farklılığın bundan kaynaklandığını öne sürmüştür. Bu örnek ile ilgili veri aşağıda verilmektedir. Mahkûmların yanında yer alan istatistikçilerin cezalarda bir ayrımcılığın olduğunu savunmaları doğru mudur? Bu hipotezi Mantel-Haenszel sürecini kullanarak test ediniz.

| | | Ölüm Cezası | |
|---|---|---|---|
| Ciddiyet | Kurbanın Irkı | Evet | Hayır |
| 1 | Beyaz | 2 | 60 |
| | Siyah | 1 | 181 |
| 2 | Beyaz | 2 | 15 |
| | Siyah | 1 | 21 |
| 3 | Beyaz | 6 | 7 |
| | Siyah | 2 | 9 |
| 4 | Beyaz | 9 | 3 |
| | Siyah | 2 | 4 |
| 5 | Beyaz | 9 | 0 |
| | Siyah | 4 | 3 |
| 6 | Beyaz | 17 | 0 |
| | Siyah | 4 | 0 |

© Cengage Learning 2013

Beyazlara karşı siyah kurbanlar için ölüm cezası olma olasılık oranını hesaplayınız.

6.37 Fidalgo (2005) iş yerindeki kabadayılık (Evet/Hayır) ile kabadayının cinsiyeti (Kadın/Erkek) arasındaki ilişkiyle ilgili bir veri sunmuştur. Bu veriyi, sonrasında iş düzeyine göre ayırmıştır. İlgili veri aşağıda verilmektedir.

| | | Kabadayılık | |
|---|---|---|---|
| Cinsiyet | İş Kategorisi | Hayır | Evet |
| Erkek | Rehber | 148 | 28 |
| Kadın | | 98 | 22 |
| Erkek | Büro işçisi | 68 | 13 |
| Kadın | | 144 | 32 |
| Erkek | Teknisyen | 121 | 18 |
| Kadın | | 43 | 10 |
| Erkek | Müdür yardımcısı | 95 | 7 |
| Kadın | | 38 | 7 |
| Erkek | Müdür / İdareci | 29 | 2 |
| Kadın | | 8 | 1 |

© Cengage Learning 2013

a. İş kategorilerine göre kabadayılık ile cinsiyet arasında bir ilişki olduğuna dair bir kanıt var mıdır?

b. a sorusundaki analiz için olasılık oranı nedir?

c. İş kategorisine göre şartlandırıldığında kabadayılıkta cinsiyet farklılığına ilişkin bir kanıt var mıdır?

d. c sorusundaki analiz için olasılık oranı nedir?

e. 2x2'den büyük frekans tablosundaki veri kümesi için muhtemelen elinizde Mantel-Haenszel testini yürütecek bir yazılım bulunmamaktadır. Ama standart Pearson ki-kareyi kullanarak her bir cinsiyet için ayrı ayrı iş kategorisi ve kabadayılık arasındaki ilişkiyi inceleyiniz. Bu analizin sonuçlarını açıklayınız.

6.38 Maine hükümeti, 1996 yılında emniyet kemeri kullanımı ve otoban kaza kurbanları ile ilgili bir veri toplamıştır. (Verinin tamamına http://maine.gov/dps/bhs/crash-data/stats/seatbelts.html adresinden ulaşmak mümkündür.)

Psikologlar sıklıkla kendine zarar verme davranışlarının sebepleri üzerine çalışırlar ve aşağıda yer alan veride emniyet kemerinin yaralanma veya ölüm olaylarını önleyip önlemediğine ilişkin değerler yer almaktadır. ("Yolcu" değişkeni gerçekten bir otoban kazasına karışan kişilerin sayısıdır.)

| | Emniyet Kemeri Takılı Değil | Emniyet Kemeri Takılı |
|---|---|---|
| Yolcu | 6307 | 65245 |
| Yaralı | 2323 | 8138 |
| Kaza Kurbanı | 62 | 35 |

Bu veriyi, ölüm ve yaralanmaları önlemede emniyet kemeri takmanın etkililiğini gösterecek şekilde kullanınız.

6.39 Appleton, French ve Vanderpump (1996) sigara içmenin insanlara iyi geldiğini göstermek üzere bir veri sunmuşlardır. 1970'lerin başında sigara içme davranışını değerlendirmişler ve ardından 20 yıl sonra hayatta kalanların verisine bakmışlardır. Kolaylık açısından veriyi halihazırda sigara içen ve hiç sigara içmemiş kadınlarla sınırlandırmışlardır. 582 sigara içicisinden 139 kişi (%24) ölmüştür. 732 sigara içmeyen kişiden ise 230 kişi (%31) ölmüştür. Açıkça içmeyenlere kıyasla daha az oranda sigara içen kişi ölmüştür. Ancak veri yaş gruplarına bölündüğünde aşağıdaki veri elde edilmektedir (Sigara içenler E, içmeyenler ise H ile gösterilmektedir):

| Yaş | 18-24 | | 25-34 | | 35-44 | | 45-54 | | 55-64 | | 65-74 | | 75+ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sigara içme durumu | E | H | E | H | E | H | E | H | E | H | E | H | E | H |
| Ölü | 2 | 1 | 3 | 5 | 14 | 7 | 27 | 12 | 51 | 40 | 29 | 101 | 13 | 64 |
| Sağ | 53 | 61 | 121 | 152 | 95 | 114 | 103 | 66 | 64 | 81 | 7 | 28 | 0 | 0 |

a. Simpson'ın paradoksuna bir örnek olacak şekilde bu sonuçları açıklayınız.

b. Mantel-Haenszel testini bu veriye uygulayınız.

6.40 Tablo 6.11'de yer alan veri kümesi için hesaplanan olasılık oranında olduğu gibi aynı veri için bağıl riski hesaplayınız.



Bölüm 7

# Ortalamalara Uygulanan Hipotez Testleri

## Hedefler

Ölçme ölçeğindeki verinin hipotezlerini test etmek için *t* testi sürecinin tanıtılması ve farklı desenlerde nasıl kullanılabileceğinin gösterilmesi, ortaya çıkan farklılıkların öneminin kestirilme yollarının açıklanması.

## Konular


**7.1** Ortalamanın Örneklem Dağılımı

**7.2** Ortalamalar ile İlgili Hipotezleri Test Etme – $\sigma$ Bilindiğinde

**7.3** $\sigma$ Bilinmediğinde Ortalamayı Test Etme - Tek Örneklem *t* Testi

**7.4** Ortalamalara Uygulanan Hipotez Testleri – Eşleştirilmiş İki Örneklem

**7.5** Ortalamalara Uygulanan Hipotez Testleri – İki Bağımsız Örneklem

**7.6** Varyansın Heterojenliği: Behrens-Fisher Problemi

**7.7** Hipotez Testlerinin Tekrar Gözden Geçirilmesi

BÖLÜM 5 VE 6'DA testleri frekanslı (kategorik) veri ile ilişkilendirerek inceledik. O bölümlerdeki durumlara göre herhangi bir deneyin sonucu, genellikle cevapların her kategorisi için birkaç ara toplam ile açıklanabilir. Bu ve sonraki bölümlerde daha önce ölçme veya nicel veri olarak ifade ettiğim farklı türde veri kümeleriyle ilgileneceğiz.

Ölçme verisini incelerken, konuyla ilgili gruplar arasındaki farklara, iki veya daha çok değişken arasındaki ilişkiye odaklanabiliriz. Soru veya değişkenler arasındaki ilişkileri Bölüm 9, 10, 15 ve 16'ya kadar erteleyeceğiz. Bu bölümde farklara ait sorularla ilgileneceğiz ve daha çok üzerinde duracağımız istatistik, örneklem ortalamasıdır.

Doğum kilosu düşük (DKD) bebekler (genellikle erken doğumlardır) birçok gelişimsel zorluklar açısından risk grubundadırlar. Nurcombe ve arkadaşları (1984) 25 DKD bebeği deney, 31 DKD bebeği ise kontrol grubu olarak alıp deney grubundaki bebeklerin ailelerine DKD bebeklerin ihtiyaçlarının nasıl karşılanacağını gösteren alıştırmalar yapma imkânı sağlamışlar ve bu bebekler iki yaşına geldiklerinde her bebeğin bilişsel yeteneklerini ölçmüşlerdir. Deney grubundaki DKD bebeklerin ortalama puanlarını 117,2, kontrol grubundakilerin ortalama puanlarını ise 106,7 olarak bulduğumuzu varsayalım. Gözlenen ortalamaların farkı, deney grubundaki iki yaşındaki DKD bebeklerin puan ortalamaları kontrol grubundakilerden daha yüksektir sonucuna varmak için yeterli bir kanıt mıdır? Bu özel soruyu daha sonra cevaplayacağız, burada bu bölümde tartışacağımız soru tarzını örneklendirmek için bu probleme değindik.

## 7.1 Ortalamanın Örneklem Dağılımı

**ortalamanın örneklem dağılımı**

**merkezi limit teoremi**

Bölüm 4'ten hatırlayacağınız gibi, herhangi bir istatistiğin örneklem dağılımı, evrenden alınan sonsuz sayıda örneklemden hesaplanması düşünülen istatistiğin dağılımıdır. Bu bölümde *örneklem ortalaması* ile ilgilendiğimiz için **ortalamanın örneklem dağılımı** hakkında bir şeyler bilmemiz gerekir. Örneklem dağılımı hakkındaki tüm önemli bilgi tek bir teorem altında toplanabilir: **merkezi limit teoremi**. Merkezi limit teoremi ortalamaların dağılımı hakkındaki olaylar raporudur. Genişletilmiş haliyle;

> Ortalaması $\mu$ ve varyansı $\sigma^2$ olan bir evrende, ortalamanın örneklem dağılımı (örneklem ortalamalarının dağılımı), ortalaması $\mu$ ($\mu_{\overline{X}} = \mu$), varyansı ($\sigma^2_{\overline{X}}$) $\sigma^2/n$ ve standart sapması ($\sigma$) $\sigma/\sqrt{n}$'ye eşit olacaktır. Örneklem büyüklüğü, n, arttıkça dağılım normal dağılıma yaklaşacaktır.[1]

Merkezî limit teoremi, istatistikteki en önemli teoremlerden biridir. Bu teorem bize sadece örneklem dağılımının ortalaması ve varyansının, verilen herhangi bir örneklem büyüklüğünde ne değer alacağını vermez; aynı zamanda ana evrenin şekli *ne olursa olsun*, *n* arttıkça örneklem dağılımının şeklinin normale yaklaştığını da belirtir. Bu durumun önemi yakın zamanda açıklığa kavuşacaktır.

*n* arttıkça ortalamanın örneklem dağılımının normale ulaşma hızı ana evrenin şeklinin fonksiyonudur. Eğer evrenin dağılımı normal ise ortalamanın örneklem dağılımı *n*'e bağlı olmadan normal dağılım olacaktır. Eğer evren simetrik ancak normal değilse, ortalamanın örneklem dağılımı özellikle evren tek modlu ise küçük örneklemlerde bile neredeyse

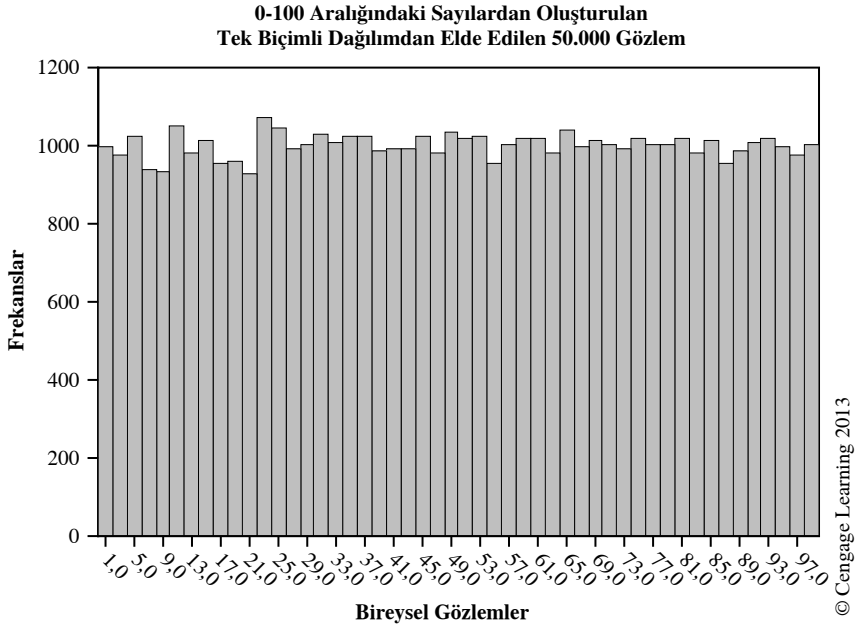
1 Merkezi limit teoremi birçok formda ifade edilmiş olarak bulunabilir. En basit hali; ortalamanın örneklem dağılımının *n* arttıkça normal dağılıma yaklaştığını bildirir. Burada verilen genişletilmiş hali ortalamanın örneklem dağılımı ile ilgili tüm önemli bilgileri içerir.

normal olacaktır. Eğer evren belirgin bir şekilde çarpık ise ortalamanın normal dağılıma yaklaşması için 30 ve üstünde örneklem büyüklüğü gerekebilir. Merkezî limit teoremini açıklamak için, 0 ile 100 arasında tesadüfî olarak alınan sayılardan oluşan sonsuz çoklukta evrenimiz olduğunu düşünelim. 0 ile 100 arasındaki her sayı eş olasılıkta olduğu için bu evren **tek biçimli (dikdörtgen) dağılıma** sahip olacaktır. Şekil 7.1'de bu evrenden elde edilen 50.000 gözlemin dağılımı gösterilmiştir. Dağılımın beklendiği gibi dikdörtgen biçiminde olduğu görülebilir. Tek biçimli dağılımlarda ortalama ($\mu$) ranjın yarısına (50), standart sapma ($\sigma$) ranjın 12'nin karekköküne bölünmesiyle elde edilen sonuca (28,87) ve varyans ($\sigma^2$) 833,33'e eşittir.

**tek biçimli (dikdörtgen) dağılım**

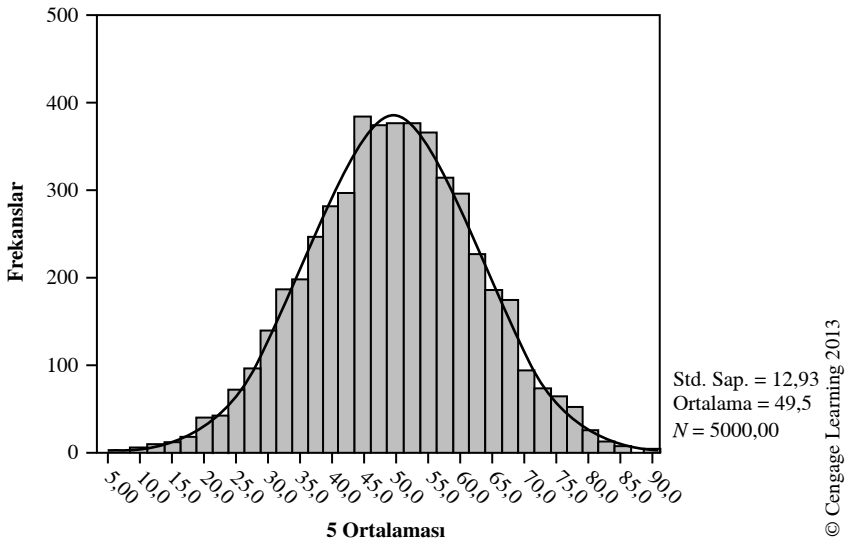
Şimdi bu evrenden örneklem büyüklüğü 5 ($n$ = 5) olan 5.000 örnek seçtiğimizi ve örneklem *ortalamalarının* grafiğini çizdiğimizi düşünelim. Böyle bir örnekleme basit bir bilgisayar programı ile bile yapılabilir; böyle bir işlemin sonucu normal dağılım eğrisi eklenerek Şekil 7.2a'da verilmiştir. (Bu örneği ilk olarak yıllar önce denediğimde ana bilgisayarda örneklemleri oluşturmak yaklaşık olarak 5 dakika sürmüştü. Bugün ise dizüstü bilgisayarımda 1,5 *saniye*mi aldı. Bilgisayar simülasyonu büyük bir sorun değildir.)[2]

Ortalamaların dağılımının tamamen normal olmasa bile, merkezde en yüksek değere ulaştığı, uçlara doğru gittikçe azaldığı açıkça görülmektedir. (Aslında normal dağılım eğrisi veriye tamamıyla uyum sağlamıştır.) Bu dağılımın ortalaması ve standart sapması gösterilmiştir ve değerler $\mu = 50$ ve $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 28,87/\sqrt{5} = 12,91$ olarak bulunmuştur. Bulunan değerler ile merkezî limit teoremiyle kestirilen değerler arasındaki farklılık yuvarlama hatasına ve sonsuz sayıda örneklem alınmamasına dayandırılabilir.

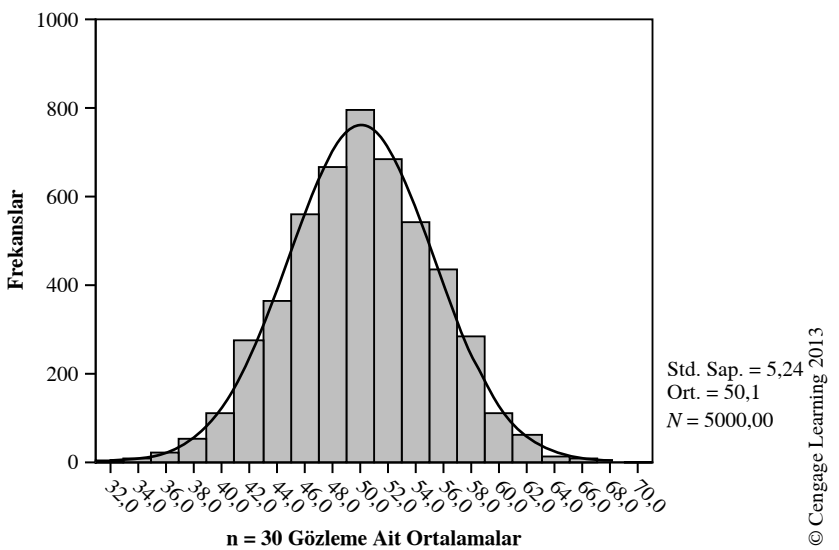



**Şekil 7.1** Tek biçimli dağılımdan elde edilen 50.000 gözlem

2 Herhangi büyüklükteki örneklem için örneklem dağılımı çizmeye fırsat tanıyan, R için yazılmış olan basit program bu kitabın web sitesinde SampDistMean.R ismiyle yer almaktadır.



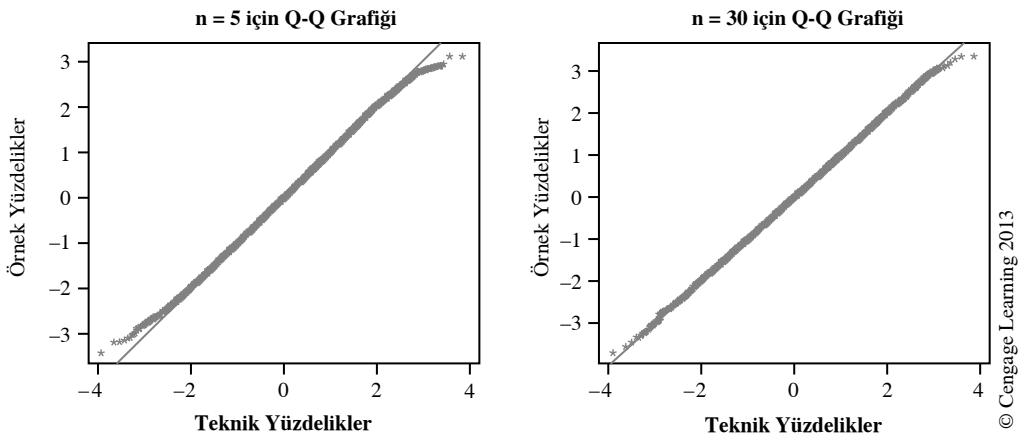

**Şekil 7.2a** $n = 5$ olduğunda ortalamanın örneklem dağılımı




**Şekil 7.2b** $n = 30$ olduğunda ortalamanın örneklem dağılımı

Şimdi tüm süreci tekrarlayalım, ancak bu defa her birinde 30 gözlem bulunan 5.000 örneklem oluşturalım. Bu örnek için elde edilen sonuçlar Şekil 7.2b'de verilmiştir. Merkezî limit teoriminin tahmin ettiği gibi dağılım yaklaşık olarak normaldir, ortalama yine $\mu = 50$'dir ve standart sapma azalarak yaklaşık $\sigma = 28{,}87/\sqrt{30} = 5{,}27$ olmuştur.

Şekil 7.2c'ye bakarak $n = 5$ ve $n = 30$ olduğunda örneklem dağılımının normalliğindeki farklılık hakkında daha iyi fikir edinilebilir. Bu şekil iki dağılım için Q-Q grafiklerini gösterir ve görüldüğü gibi $n = 5$ iken bile dağılım normalden çok uzak değildir ancak $n = 30$ olduğunda dağılımın normale çok daha yakın olduğu görülmektedir.




**Şekil 7.2c** n = 5 ve $n$ = 30 olduğunda örnekleme dağılımları için grafikleri (Q-Q plots)

## 7.2 Ortalamalarla İlgili Hipotezleri Test Etme – *v* Bilindiğinde

Merkezî limit teoremi sayesinde ortalamanın örneklem dağılımına ait önemli tüm özellikleri (şeklini, ortalamasını ve standart sapmasını) biliyoruz. Bu bilgi ışığında ortalamalar ile ilgili hipotezleri test edebiliriz.

Evren ortalaması ile ilgili hipotezleri test ettiğimiz birçok durumda evrenin varyansı hakkında herhangi bir bilgimiz olmaz. (Bu durum, bu bölümün odak konusu olan *t-testinin* temel sebebidir.) Ancak, sınırlı sayıdaki bazı durumlarda $\sigma$ bilinebilir. $\sigma$ bilindiğinde hipotezleri test etme konusunu tartışma, bize normal dağılım hakkında ne bildiğimiz ile $t$ testi hakkında ne bilmek istediğimiz arasında iyi bir geçiş sağlar.

Williamson'ın (2008) şeref (honor) derecesi alan tezinde, depresyondaki ebeveynlerin çocuklarında oluşan başa çıkma davranışlarını incelediği örnek ile başlayacağız. Williamson'ın yaptığı bu çalışmada, ele aldığımız örnek, çalışmanın çok küçük bir bölümüdür ve burada tek örneklem ortalamasını kullanarak yokluk hipotezini test etmeyi anlaşılır kılan bir durum yer almaktadır.

Psikoloji literatüründe çocuğun hayatındaki stresin bazı davranış problemlerine sebep olduğuna dair kanıtlar yer aldığı için, Williamson, depresyondaki ebeveynlerin çocuklarından oluşan örnekleminin yüksek düzeyde problem davranış göstermesini beklemiştir. Bu durumda Achenbach'ın gençlik öz değerlendirme envanterinin kaygı/depresyon alt ölçeği gibi davranışlarla ilgili bir kontrol listesi kullanırsak, ailelerinde depresyon şikayeti olan çocuklardan oluşan örneklemden yüksek puan almaları beklenir. (Gençlik öz değerlendirme envanteri durumu örneklendirmek için uygundur, çünkü evren ortalaması (50) ve standart sapması (10) bilinmektedir.)

Çocukların düşük seviyede depresyon gösterme olasılıkları küçük olsa da, bu olasılığa karşı önlem almak için iki yönlü deneysel hipotez test edilecektir. Test edilecek hipotez; "Bu çocukların kaygı/depresyon puanları normal örneklemden elde edilecek puanlardan farklıdır." şeklinde olacaktır. Ancak deneysel hipotezi doğrudan test edemeyiz. Bunun yerine reddedildiğinde deneysel hipotezin kabul edildiği yokluk hipotezi ($H_0$) test edilecektir. Sıfır hipotezi, "Stresli çocukların evren puan ortalamaları ile normal çocukların evren puan ortalamaları arasında fark yoktur." şeklindedir. Daha özel ifade edecek olursak, $H_0$: $\mu = 50$ ile $H_1$: $\mu \neq 50$ arasında karar vermek istiyoruz.

$H_1$ hipotezini iki yönlü olarak belirledik, çünkü $\mu > 50$ ya da $\mu < 50$ olması durumunda $H_0$'ı reddetmek istiyoruz.

Gençlik Öz Değerlendirme Envanteri'nin evren ortalamasını ve standart sapmasını bildiğimizden, yokluk hipotezi doğru olduğunda örneklem dağılımını belirlemek için merkezi limit teoremini kullanabiliriz. Merkezî limit teoremine göre, bu evrenden ortalamanın örneklem dağılımı elde edildiğinde ortalama ($\mu$) 50'ye , varyans $\sigma^2/n$'ye ve standart sapma (genellikle **standart hata**[3] olarak adlandırılır) $\sigma/\sqrt{n}$'ye eşit olacaktır.

**standart hata**

Williamson (2008), ebeveynlerinden en az birinin depresyon geçmişi olan 166 çocukla çalışmıştır. Bu çocukların hepsine Gençlik Öz Değerlendirme Envanteri'ni uygulamış ve örneklem ortalamasını 55,71, standart sapmasını 7,35 bulmuştur. Biz bu çocukların, ortalaması 50, standart sapması 10 olan normal evrenden geldiklerini savunan yokluk hipotezini test etmek istiyoruz.

Yokluk hipotezimizi test etmeye $z$'nin standart formülüyle başlayabiliriz. Ardından ortalama ve standart hata değerlerini puan ve standart sapma değerlerinin yerine koyabiliriz.

$$z = \frac{X - \mu}{\sigma} \quad \text{ifadesi} \quad z = \frac{X - \mu}{\sigma_{\overline{X}}}$$

biçimine dönüşür. Bu,

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

biçiminde de yazılabilir.

Buradan

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{55,71 - 50}{\frac{10}{\sqrt{166}}} = \frac{5,71}{0,776} = 7,36$$

elde edilir.

Normal dağılım tablosunda $z$'nin en büyük değeri 4,00 olduğu için bu tabloyu kullanamayız. Ancak elde ettiğimiz $z$ değeri 4,00 bile olsaydı 0,000'dan daha düşük bir olasılıkla manidar olacaktı, bu nedenle her durumda yokluk hipotezini reddedebiliriz. (İki yönlü test için tam olasılık 0,00000000000018'dir ve bu sonucun manidar olduğu açıktır.) Bu durumda Williamson'ın, çalışmada yer alan çocukların Gençlik Öz Değerlendirme Envanteri'nin kaygı/depresyon alt ölçeğinden aldıkları puanların tesadüfî örneklemi temsil etmediğine inanmak için her türlü kanıtı vardır. (Jones ve Tukey'nin (2000) daha önce belirttiği gibi, "stres altındaki çocukların puan ortalaması diğer çocuklarınkinden daha yüksektir" bulgusuna ulaşmış olduk.)

Yaptığımız uygulamada olduğu gibi, evren ortalaması bilindiğinde tek örneklem ortalamasının test edilmesi, ortalamaların normal dağılması veya hatayı göz ardı edilebilecek kadar normal dağılması varsayımına dayanır. Birçok ders kitabı örneklemin normal dağıldığını varsayar, ancak bu durum pratikte çok da gerekli değildir. Varsayımda önemli olan ortalamanın örneklem dağılımının normale yakın olmasıdır (Şekil 12.3). Varsayım iki şekilde sağlanabilir; ya (1) örneklem oluşturduğumuz evren normaldir veya (2) örneklem büyüklüğü merkezî limit teoremi yoluyla yaklaşık bir normallik elde edecek kadar büyüktür. Merkezî limit teoremi, N yeterince büyük olduğunda, evren örneklemi normal olmasa bile hipotezi test etme olanağı sağlar. Bu merkezî limit teoreminin sağladığı en önemli avantajlardan biridir.

---

3 Herhangi bir örneklem dağılımının standart sapması bu dağılımın standart hatası olarak anılır. Bu nedenle ortalamanın standart sapmasına ortalamanın standart hatası denir ($\sigma_{\overline{X}}$ şeklinde gösterilir). Ayrıca kısaca değinilecek olan ortalamalar arasındaki farkın standart sapmasına ortalamalar arasındaki farkın standart hatası denir ve $\sigma_{\overline{X}_1 - \overline{X}_2}$ şeklinde gösterilir.

Williamson örneklemin standart sapmasını da elde etmişti. Peki biz o değeri niçin kullanmadık?

Bu soruya verilecek en basit cevap elimizde daha iyi bir değerin olduğudur. Biz *evrenin* standart sapmasını biliyorduk. Gençlik Envanteri ve buna ait puanlar yıllardır özenle geliştirilmiş ve kullanılmaktadır. Normal çocuklar evreninin standart sapmasının 10 veya farkı manidar olmayacak kadar 10'a yakın bir değer olduğunu biliyoruz. Ve biz de örneklemimizin normal çocukların evreninden gelip gelmediğini test etmek istiyoruz. Genellikle evrenin standart sapma değerini bilemeyiz ve örneklemin standart sapmasından kestirmek durumunda kalırız, ancak değeri biliniyorsa evrenin standart sapması kullanılmalıdır. Bir kere, eğer bu çocuklar gerçekten daha yüksek puanlara sahiplerse standart sapma değerleri alt kestirim olacaktır. Çünkü dağılımın sonunda olan standart sapma değerleri, ortalarda olanlara göre daha küçüktür.

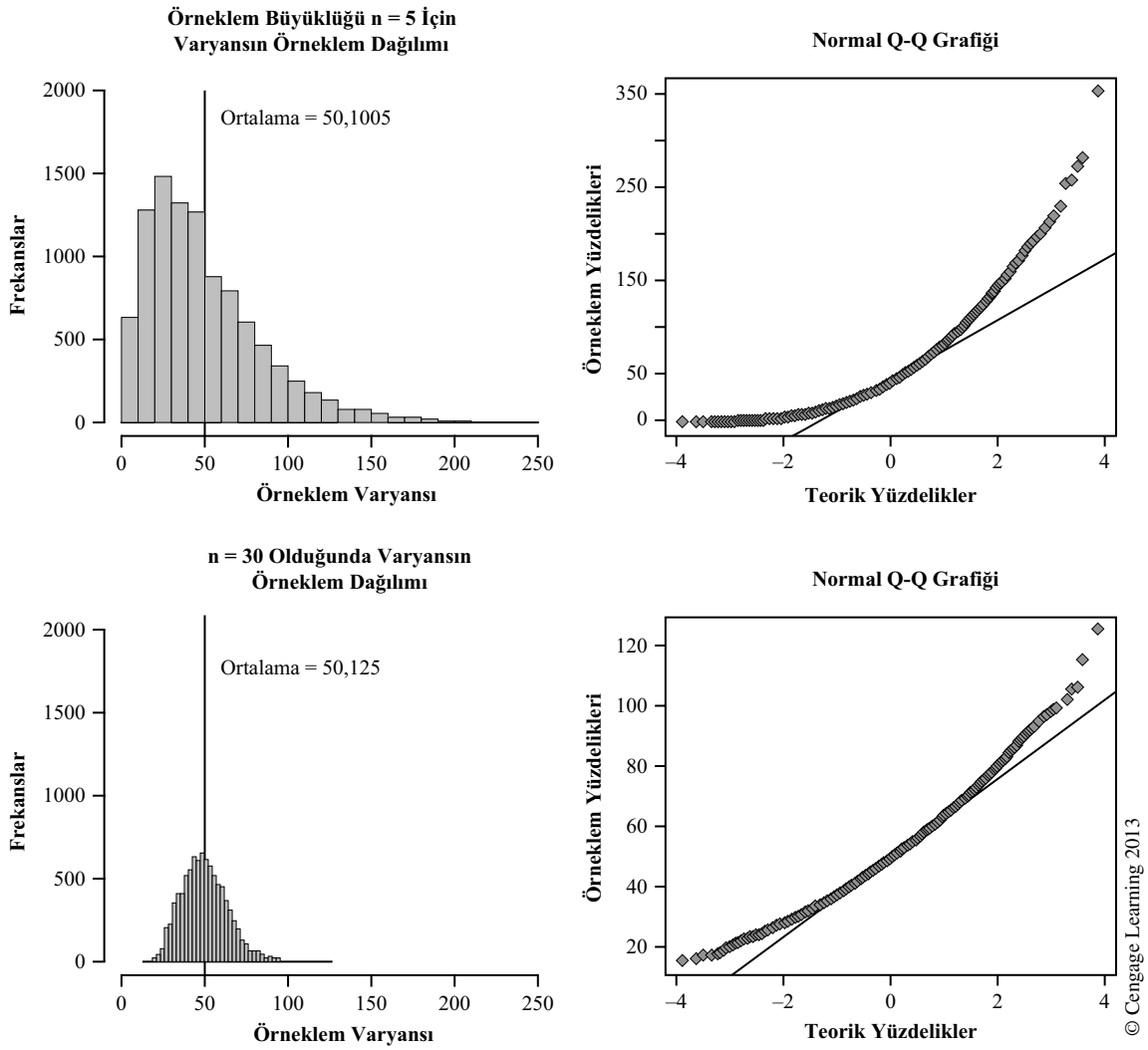
## 7.3 *v* Bilinmediğinde Ortalamayı Test Etme - Tek Örneklem *t* Testi

Yukarıda açıklanan örnek, evrenin standart sapmasının bilindiği sınırlı sayıdaki durum arasından kasıtlı olarak seçilmiştir. Genelde $\sigma$ değerini bilemeyiz ve örneklemin standart sapmasından ($s$) kestirmemiz gerekir. Formülde $\sigma$ yerine $s$ koyduğumuzda testin yapısı değişecektir. Bu durumda cevabın $z$ puanı olduğunu söyleyemeyiz ve elde edilen değerin hesaplanması için $z$ tablosunu kullanamayız. Bunun yerine cevap $t$ ile gösterilecek ve $z$ tablosundan farklı olan $t$ tablosu kullanılarak hesaplanacaktır. $z$'den $t$'ye geçmenin altında yatan neden örneklem varyansının örneklem dağılımı ile ilgilidir.

### $s^2$'nin Örneklem Dağılımı

$t$ testinde $\sigma^2$'nin kestirimi olarak $s^2$ kullanıldığından öncelikle $s^2$'nin örneklem dağılımını incelemeliyiz. Bu örneklem dağılımı karşılaşılacak problemler hakkında öngörümüzün olmasını sağlar. Bölüm 2'de $s^2$'nin, $\sigma^2$'nin *yansız* kestiricisi olduğunu ve tekrarlı örneklemler sonucunda $s^2$'nin ortalama değerinin $\sigma^2$'ye eşit olduğunu görmüştük. Yansız kestirici güzel bir şey olsa da her şey demek değildir. Çünkü *özellikle küçük örneklemlerde $s^2$'nin örneklem dağılımı sağa çarpıktır*. $\mu = 5$ ve $\sigma^2 = 50$ olan normal dağılan bir evrenden n = 5 ve n = 30 olan 10.000 örneklem çektim. Her örneklem için varyansı hesapladım ve Şekil 7.3'te bu 10.000 varyansa ait grafiği çizdim. Normalliğe bakmak için de Q-Q grafiğine başvurdum. Tüm dağılımların ortalaması 50'dir; $s^2$, $\sigma^2$'nin yansız bir kestiricisi olarak yapısını yansıtır. Ancak dağılım özellikle küçük örneklemlerde, bir hayli sağa çarpıktır. Varyansın dağılımının çarpıklığı nedeniyle özellikle küçük örneklemlerde $s^2$, $\sigma^2$'yi olduğundan daha küçük kestirir. Ayrıca bu çarpıklık nedeniyle $t$ değeri, $\sigma$ bilindiğinde elde edilen $z$ değerinden büyük olacaktır.[4]

4 Örneklem büyüklüğünü değiştirdiğinizde ne olacağını kitabın Web sitesinde yer alan SampDistVar.R isimli R programında deneyebilirsiniz.




**Şekil 7.3** $n = 5$ ve $n = 30$ için örneklem varyansının örneklem dağılımı

## *t* İstatistiği

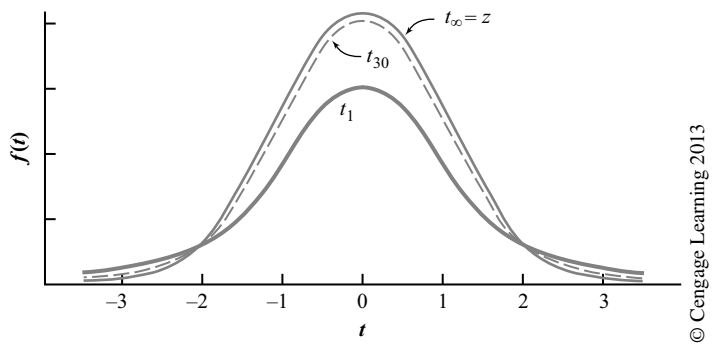
Yukarıda geliştirdiğimiz $z$'yi ele alacağız,

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X}\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

Bu ifadede $\sigma$ yerine $s$ konularak.

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}}$$

bulunur.




**Şekil 7.4** 1, 30 ve ∞ serbestlik dereceleri için *t* dağılımları



Annals of Eugenics

William S. Gosset
(1876–1937)

**Student'ın *t* dağılımı**

Herhangi bir örneklem için $s^2$ değerinin $\sigma^2$ değerinden daha küçük olmayacağını bildiğimizden, paydasının olması gerekenden daha büyük olacağını biliyoruz. Bunun anlamı, *t formülü* ile gerçek ancak bilinmeyen $\sigma^2$ değeri kullanılarak hesaplanan *z*'den daha büyük (mutlak değerce) bir cevap elde edilmeyeceğidir. (Bu durumu gözlemlerin yarısından çoğunun $\sigma^2$ değerinin soluna düştüğü Şekil 7.3'te görebilirsiniz.) Sonuç olarak cevabı *z* puanı olarak ele almak ve *z* tablosunu kullanmak hoş olmayabilir. Ayrıca işlemleri *z* puanı üzerinden yapmak birçok "manidar" sonuç elde etmemize, diğer bir deyişle, %5'in üzerinde 1. Tip hata yapmamıza sebep olur. (Örneğin, *z*'yi hesaplarken, değer ,05 manidarlık düzeyinde ±1,96'yı aştığında $H_0$ reddedilir. $n = 5$ için tekrarlı örneklemler üzerinden $H_0$'ın kabul edileceği durumlar oluşturduğumuzda ve $\sigma^2$ yerine $s^2$ değeri kullandığımızda ±1,96 değerini veya %10 daha fazlasını elde ederiz. Bu durum için $t_{,05}$ kesme puanı ±2,776'dır.)[5]

Bu problemim çözümü 1908 yılında Guinness Bira firmasında çalışan William Gosset[6] tarafından gerçekleştirilmiştir. Gosset 1900'lü yılların başında "Student" takma adını kullanarak birçok önemli çalışma yapmıştır. Gosset, normal dağılımdan örneklenmiş veri kümesinde $\sigma^2$ yerine $s^2$ kullanılırsa günümüzde **Student'ın *t* dağılımı** olarak bilinen özel bir örneklem dağılımının elde edileceğini belirtmiştir. Gosset'in çalışması sonucunda yapılması gereken tek şey; $s^2$ değerini yerine koymak, sonucu *t* cinsinden elde etmek ve *z* değerini normal dağılıma bakarak nasıl yorumluyorsak *t* değerini de kendi dağılımına göre yorumlamaktır.

Ek *t*'de *t* dağılımları tablosu yer almaktadır ve farklı örneklem büyüklükleri için muhtemel *t* dağılımları Şekil 7.4'te gösterilmiştir.

Şekil 7.4'te görüldüğü gibi, *t* dağılımı, örneklemdeki gözlem sayısının bir eksiği olarak tanımlanan serbestlik derecesinin fonksiyonu olarak farklı değerler alır. $n \Rightarrow \infty, p(s^2 < \sigma^2) \Rightarrow p(s^2 > \sigma^2)$ $s^2$'nin örneklem dağılımının çarpıklığı, serbestlik derecesi sayısı arttıkça düzeleceğinden, s'nin, $\sigma$'yı kestirmedeki küçülme eğilimi de ortadan kalkacaktır. Bu nedenle sonsuz büyüklükte serbestlik derecesi sayısı için *t* normal dağılacaktır ve *z*'ye eşit olacaktır.

---

5 Kitabın Sampdistt.R isimli Web sitesinde örneklem büyüklüklerini çeşitlendirebildiğiniz, *t* ve normal dağılımda hipotezi reddetme oranları arasındaki farkı görebileceğiniz programı bulabilirsiniz. Örneğin n=10 için tablodaki t = ±2,262 değeri kullanılarak %4,66 oranında, t = ±1,96 için ise %7,46 oranında $H_0$ reddedilir.

6 William Sealy Gosset gibi istatistik için çok önemli bir kimsenin en azından adının doğru yazılması gerektiğini düşünebilirsiniz. Kontrol ettiğim kitapların neredeyse yarısında ismi "Gossett" olarak yazılıydı. Güvendiğim birçok kişi ise tek "t" ile yazılması gerektiğini söyledi. Gosset'in ilginç yaşam hikayesini http://en.wikipwdia.org/wiki/William_Seally_Gosset adresinde bulabilirsiniz.

Bilinen evren ortalaması için tek örneklem ortalaması testi, örneklemin normal dağılıma sahip bir evrenden olduğu varsayımına dayanır. Bu varsayım önceliklidir, çünkü Gosset *t* dağılımını elde ederken ortalamanın ve varyansın bağımsız olduğunu ve normal dağıldığını varsayar. Ancak uygulamada, örneklem büyüklüğü normale yakın bir dağılım gösterecek kadar büyük olduğunda, elde edilen *t* istatistiği *t* dağılımı ile karşılaştırılabilir. Birçok durum için *n*'nin 25 veya 30 olması yeterlidir ayrıca birçok durum için bundan daha küçük örneklemler bile yeterli olabilir.

## Serbestlik Derecesi

*t* dağılımının, serbestlik derecesinin (*sd*) fonksiyonu olduğunu belirtmiştik. Tek örneklem için, *sd = n-1*'dir; $s^2$'yi hesaplarken örneklem ortalamasını kullandığımız için serbestlik derecesinin birini kaybettik. Daha açık bir anlatımla, varyansı hesaplarken ($s^2$) gözlem sonuçlarının evren ortalamasından farkını $(X - \mu)$ değil, gözlem sonuçlarının kendi ortalamasından farkını $(X - \overline{X})$ kullandık. Çünkü $\left[\sum (X - \overline{X})\right]$ örneklem ortalamasından farkların toplamı her zaman sıfırdır, sadece *n*-1 tane fark serbest olarak değişebilir (farkların toplamı sıfıra eşit olursa *n*'yinci fark hesaplanabilir).

## Doğum Kilosu Düşük Bebeklerin Psikomotor Yetenekleri

Doğum kilosu düşük (DKD) olan bebeklerle ilgili çalışmadan alınan örnek bu noktada faydalı olacaktır, çünkü aynı genel çalışma bu bölümde ve bölümün sonraki kısımlarında sözü edilecek *t* testlerini açıklayacaktır. Nurcombe (1984), DKD bebeklerin anneleri için müdahale programı rapor etmiştir. Bu bebekler, ailelerine özel problemler yaşatırlar, çünkü fiziksel ve gelişimsel problemler yönünden risk altında olmalarının yanında (yüzeysel olarak) tepkisiz ve öngörülemeyendirler. Müdahale programı annelerin bebeklerinin verdiği işaretlerin farkında ve ihtiyaçları konusunda daha duyarlı olmalarını sağlamak için hazırlanmıştır, böylelikle DKD bebeklerde görülebilen gelişim güçlüklerini azaltmak amaçlanmıştır. Çalışmada üç grup bebek kullanılmıştır: DKD deney grubu, DKD kontrol grubu ve doğum kilosu normal (DKN) olan bebeklerden oluşan grup. Son iki grupta yer alan bebeklerin annelerine müdahale programı eğitimi verilmemiştir.

Çalışmada kullanılan bağımsız değişkenlerden biri Bayley Bebek Gelişim Ölçeği'nden elde edilen Psikomotor Gelişim İndeksi'dir (PGİ). Bu ölçek bebeklerin tümüne ilk olarak 6 aylıkken uygulanmıştır. Çünkü 6 aydan önce DKD gruplar arasındaki psikomotor gelişim farklarını görmek mümkün değildir. İki gruptan elde edilen veri kümeleri karşılaştırılmış ve DKD bebeklerin evren ortalaması olan 100 (bu ölçek için) değerinden ne kadar farklılaştığı araştırılmıştır.

Tablo 7.1'de DKD bebeklerin psikomotor gelişim puanları görülmektedir. Tabloda gövde-dal-yaprak gösterimi ve kutu grafiğine de yer verilmiştir. Bu iki gösterim şekli veri kümesinin genel dağılımını incelemek ve uç değerleri görmek için önemlidir.

Gövde-ve-yaprak gösteriminden veri kümesinin, tamamen normal olmasa bile çok çarpık olmayan bir dağılım gösterdiği görülebilir. Dağılımın çarpık olmaması önemlidir, çünkü bu durum ortalamaya ait örneklem dağılımının, normal dağılıma ulaşması daha hızlı olacak demektir. Ancak Q-Q grafiğinde veri kümesinin uçlarında ayrılmanın yoğunlaştığı görülmektedir. Ele alınan örneklem sayısında (56) ortalamanın örneklem dağılımının normal olduğunu varsaymak mantıklıdır ayrıca örneklem sayısı küçük olsaydı Q-Q grafiğinde ortaya çıkardı.[7] Gövde-ve-yaprak grafiğinde görülen ilginç ve beklenmeyen sonuç ise bazı

7 Basit bir yeniden örnekleme çalışması (gösterilmemiştir) ile elde edilen bu şekil için evren ortalamasının örneklem dağılımının normale çok yakın olduğunu gösterebilirsiniz. Kitabın Web sitesinde BootstrapMenas.R isimli uygulamayı görebilirsiniz.

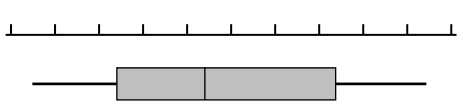
**Tablo 8.1** Doğum Kilosu Düşük (DKD) Bebekler İçin Psikomotor Gelişim İndeksi (PGİ) Verisi

| Ham Veri | | | |
|---|---|---|---|
| 96 | 120 | 112 | 100 |
| 125 | 96 | 86 | 124 |
| 89 | 104 | 116 | 89 |
| 127 | 89 | 89 | 124 |
| 102 | 104 | 120 | 102 |
| 112 | 92 | 92 | 102 |
| 120 | 124 | 83 | 116 |
| 108 | 96 | 108 | 96 |
| 92 | 108 | 108 | 95 |
| 120 | 86 | 92 | 100 |
| 104 | 100 | 120 | 120 |
| 89 | 92 | 102 | 98 |
| 92 | 98 | 100 | 108 |
| 89 | 117 | 112 | 126 |

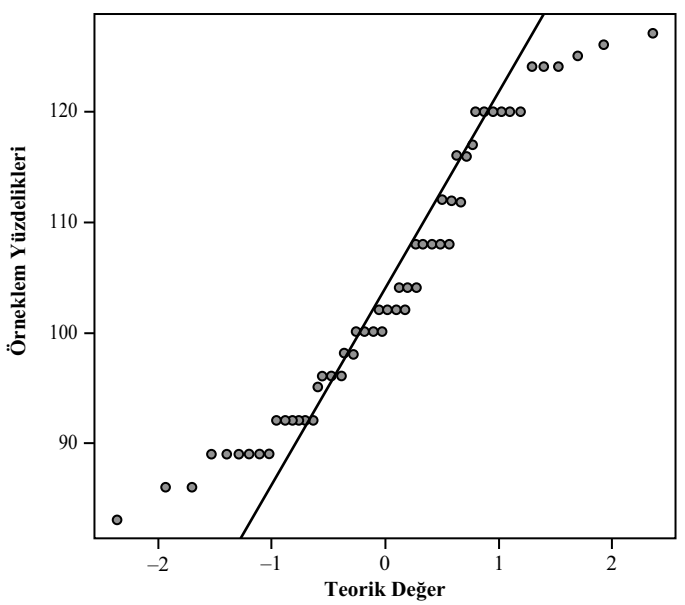
| Gövde-Yaprak Gösterimi | |
|---|---|
| Gövde | Yaprak |
| 8* | 3 |
| 8. | 6 6 9 9 9 9 9 9 |
| 9* | 2 2 2 2 2 2 |
| 9. | 5 6 6 6 6 8 8 |
| 10* | 0 0 0 0 2 2 2 2 4 4 4 |
| 10. | 8 8 8 8 8 |
| 11* | 2 2 2 |
| 11. | 6 6 7 |
| 12* | 0 0 0 0 0 0 4 4 4 |
| 12. | 5 6 7 |

Ortalamama = 104,125
S.S = 12,584
$N$ = 56

**Kutu Grafiği**



**Düşük Doğum Kilosu Verisine Ait Q-Q Grafiği**




© Cengage Learning 2013

puanların baskın olmasıdır. Örneğin beş adet 108 puan yer almakta ve 104 ile 112 arasında başka herhangi bir değer bulunmamaktadır. Benzer şekilde, altı 120 puan olup, 117 ve 124 arasında başka hiç değer yoktur. Ayrıca altı 89 puanı dikkate alınmazsa tek sayılarda ciddi bir eksiklik vardır. Verinin genel analizi, bu garipliklerin ortaya çıkmasını ve kaynaklarının izini sürmeyi gerektirir. Puanların gruplar halinde olma eğiliminin herhangi bir nedeninin olup olmadığını araştırmak için puanlama sürecinin incelenmesi gerekir. Bu durumun sebebi muhtemelen ham puanlar ölçek puanlarına dönüştürülürken yapılmış bir işlem hatası olabilir, yine de kontrol edilmelidir. (Puanlama el ile kontrol edildiğinde bu tuhaflıkların beklenilen durumlar olduğu görülebilir.) Tukey'nin açımlayıcı faktör analizinin (AFA) üstünlüklerinden biri de bu tuhaflıkların ortaya çıkmasını sağlamasıdır. Sonuç olarak kutu grafiği incelendiğinde endişelenilecek uç değerlerin olmadığı görülebilir.

Tablo 7.1'deki veri kümesinden DKD bebeklerin psikomotor gelişim puanları ortalamasının 104,125 olduğunu görebiliriz. Normlara göre psikomotor gelişim puanları için evren ortalaması 100 olmalıdır. Eldeki veriye göre kontrol edilmesi gereken ilk nokta örneklem ortalamasının evren ortalaması olan 100'den manidar olarak farklılaşıp farklılaşmadığıdır. Bu problemin çözümü için *t* testi kullanılır.

Elde edilen değerler *t* eşitliğinde kullanılarak

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$
$$= \frac{104,125 - 100}{\frac{12,584}{\sqrt{56}}} = \frac{4,125}{1,682}$$
$$= 2,45$$

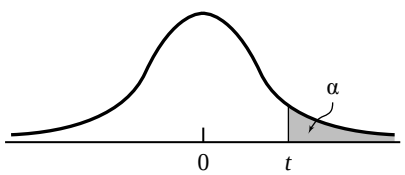
bulunur. Bulunan *t* değeri, bu küme, ortalaması $\mu$ = 100 olan bir evrenden örneklendiyse serbestlik derecesi 56 - 1 = 55 için *t* dağılımının içinde yer almalıdır - eğer veri $\mu$ = 100 olan bir evrenden örneklendi ise.

2,45 olarak bulunan *t* değeri, örneklem dağılımına dayanarak değerlendirilmeden anlam kazanmaz. Bunun için Ek *t*'de *t*'nin kritik değerleri verilmiştir. *z* değerlerinin tersine her serbestlik derecesi için farklı bir kritik *t* değeri vardır. *t* tablosu, ki-kare dağılımı gibi normal dağılım tablosundan (*z*) yapı olarak farklılaşır. *t* tablosunda her *t* değeri için, çok fazla yer kaplayacağı düşünülen alt ve üst alan yerine ,05 ve ,01 gibi kritik alanların kesme puanları verilmiştir. İki yönlü ,05 düzeyinde işlem yaptığımız için gerekli olan kritik *t* değeri 5/2 = %2,5'tur. Kritik değer genel olarak $t_{\alpha/2}$ veya bizim örneğimiz için $t_{,025}$ olarak gösterilir. Ek *t*'den $t_{,025}$ için kritik değere bakılmış ve 2,009 bulunmuştur. (Ek *t*'nin kısaltılmış şekli Tablo 7,2'de verilmiştir.) (Serbestlik derecesini göstermek için $t_{,025}(50)$ şeklinde gösterime de gidilebilir.) Hesaplanan *t* değeri kritik değerden daha büyük olduğu için iki yönlü ,05 düzeyinde $H_0$ reddedilir. "DKD bebeklerden oluşan örneklemimiz genel evrenden farklıdır." sonucuna varılır. Gerçekte örneklem ortalaması istatistiksel olarak manidar bir şekilde evren ortalamasından büyüktür. Bu durum iki yönlü test kullanmanın sağladığı avantajı ortaya koyar, çünkü grubun puanlarının evren ortalamasının altında kalması beklenirdi. Bu durum ayrıca puanlama sürecimizi de fazla puan verilip verilmediği yönünden gözden geçirmemiz gerektiğini öne sürer. Ancak birçok başka çalışmada da benzer şekilde daha yüksek ortalama elde edildiği görülmüştür. [bkz. Achenbach, C. T. Howell, Aoki ve Rauh (1993).]

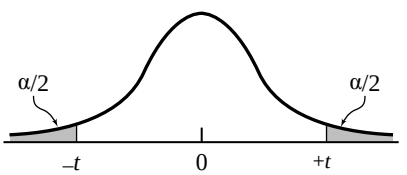
**NOT**

Ki-kare testinde gördüğümüz ve varyans analizinde tekrardan göreceğimiz gibi, bu kitapta ele alınan testlerin çoğunda pozitif test istatistikleri elde edilmiştir. Ancak *t* testinde hem pozitif hem de negatif sonuçlar elde edilebilir ve elde edilen *t* değerinin çok büyük ya da çok küçük olmasına göre yokluk hipotezini ret etmek isteriz. Bu durum iki yönlü testlerin yorumlanmasında karışıklığa yol açabilir. Yukarıda iki yönlü test için kritik değeri $t_{,025}(50) = 2{,}009$ olarak gösterdim. Bir diğer ifadeyle *t* dağılımında 50 serbestlik derecesi için üst %2,5'un sınırı 2,009 ve simetrik olarak alt %2,5'in sınırı -2,009'dur. Benzer şekilde durum $t_{,05}(50) = \pm 2{,}009$ olarak da gösterilebilir. Her iki gösterim türü de doğrudur. Kitap içinde bu konuda tutarlı davranılmaya çalışılmış ve zaman zaman unutulsa da $t_{,025}(sd) = 2{,}009$ gösterim şekli kullanılmıştır. Ayrıca aksini söylemedikçe kitaptaki tüm testler ,05 düzeyinde olacaktır ve sonuçlarda iki yönlü pozitif kesme puanı rapor edilecektir. (Diğer yönün kesme puanı bu değerin negatifidir.)

**Tablo 7.2** *t* dağılımının yüzdelik puanları



Tek Yönlü Test



İki Yönlü Test

| Tek Yönlü Test İçin Manidarlık Düzeyleri | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ,25 | ,20 | ,15 | ,10 | ,05 | ,025 | ,01 | ,005 | ,0005 |
| İki Yönlü Test İçin Manidarlık Düzeyleri | | | | | | | | | |
| *sd* | ,50 | ,40 | ,30 | ,20 | ,10 | ,05 | ,02 | ,01 | ,001 |
| 1 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 | 636,62 |
| 2 | 0,816 | 1,061 | 1,386 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 31,599 |
| 3 | 0,765 | 0,978 | 1,250 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 12,924 |
| 4 | 0,741 | 0,941 | 1,190 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 8,610 |
| 5 | 0,727 | 0,920 | 1,156 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 6,869 |
| 6 | 0,718 | 0,906 | 1,134 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 5,959 |
| 7 | 0,711 | 0,896 | 1,119 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 5,408 |
| 8 | 0,706 | 0,889 | 1,108 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 5,041 |
| 9 | 0,703 | 0,883 | 1,100 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,781 |
| 10 | 0,700 | 0,879 | 1,093 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,587 |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
| 30 | 0,683 | 0,854 | 1,055 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,646 |
| 40 | 0,681 | 0,851 | 1,050 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 3,551 |
| 50 | 0,679 | 0,849 | 1,047 | 1,299 | 1,676 | 2,009 | 2,403 | 2,678 | 3,496 |
| 100 | 0,677 | 0,845 | 1,042 | 1,290 | 1,660 | 1,984 | 2,364 | 2,626 | 3,390 |
| ∞ | 0,674 | 0,842 | 1,036 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,291 |

*Kaynak:* Tablodaki değerler kitabın yazarı tarafından hesaplanmıştır.

© Cengage Learning 2013

## İşler Değişiyor

Birçok insanın beklentisinin aksine istatistiksel süreç zamanla değişir. Bu şu anlama gelir; birkaç yıl içinde profesyonel çalışmalarda, kitaplarınızda hiç görmediğiniz şeyler göreceksiniz. (Bunu daha önce belirttim, şimdi tekrar belirtiyorum.) Şu an, bu gelişimlerden birini örneklemek için uygun bir zamandır, çünkü şimdi veri normal dağılmadığında *t* testi uygulama konusu ele alınacaktır.

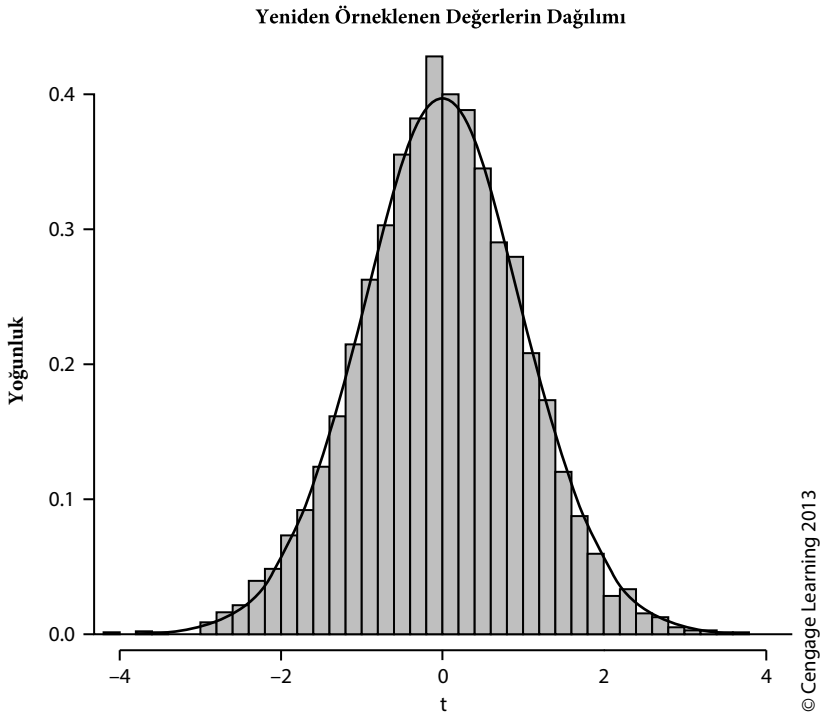
En büyük gelişim bilgisayar programlarının inanılmaz hızına bağlı olarak ortaya çıkmıştır. Birkaç yıl önce "….'yı yapabildiğini düşünsene" dediğimiz hesaplamaları bugün kolaylıkla yapabiliyoruz. Örneğin *t*-testini açıklarken "evrenin normal dağıldığı ve bu dağılımdan çok büyük örneklemleri ele aldığımız varsayılır, her örneklem için *t* değeri hesaplanır, sonuçta elde edilen grafik çizilir ve buna *t*'nin örneklem dağılımı denir idi". Peki, ya normal dağılan bir evrenimiz yoksa ve elimizdeki örneklem de Tablo7.1'deki gibi normal olmayan bir dağılıma sahipse? Tablo 7.1'deki gibi normal dağılmayan büyük bir evrene sahip olduğumuzu düşünelim. Böyle bir evreni örneklemdeki her veriyi birçok kez çoğaltarak ve aynı zamanda evrenin ortalamasının yokluk hipotezindeki gibi 100'e eşit olmasını sağlayarak elde edebiliriz. Gerçekte yapılan iş büyük bir evren oluşturmak değil orijinal verideki değerlerden bazılarını ekleyerek çok büyük bir evren oluşturmaktır. Yapılan bu iş aynı etkiye sahiptir ve kesinlikle daha kolaydır.

Şimdi bu evrenden tesadüfî olarak örneklediğimiz 56 gözlemi ele aldığımızı ve bu örneklemin ortalamasını hesapladığımızı varsayalım. Ardından bu süreci 10.000 kere tekrarlayalım. Bu durumda elde edilen ortalamanın orijinal örneklemden elde edilen ortalama olan 104,125'ten yüzde kaç oranında büyük olduğunu sorabiliriz. Elde edilen ortalama 104,125'i düşük oranda aşıyorsa bile yokluk hipotezini reddedebiliriz. Çünkü eğer evren ortalaması gerçekten 100 ise ortalamanın 104,125 olması nadiren ortaya çıkan bir durumdur. Bir adım öteye giderek, her örneklem için sadece ortalamayı hesaplamak yerine *t* istatistiklerini de hesaplayalım. *t* değerini hesaplamak herhangi bir farklılık yaratmayacaktır, çünkü örneklem ortalaması 104,125'ten büyükse ilgili *t* değeri de 2,4529'dan büyük olacaktır. Ancak *t* değerini hesaplayarak normallik varsayımını sağlamayan yeniden örneklenen dağılımın *t* grafiği ile varsayımı sağlayan Student'a ait *t* örnekleme dağılımı grafiğini çizebiliriz. Sonuçlar Şekil 7.5'te verilmiştir. *t* değerinin 2,4529'a eşit ya da büyük olup olmadığı veya örneklem ortalamasının 104,125'ten ne kadar büyük olduğunu hesapladığımızda oran ,0146 çıkmıştır ve bu durum yokluk hipotezini reddedebileceğimiz anlamına gelir.

**özyükleme**

Verdiğimiz örnekte yapılan işe **özyükleme (bootstrapping)** denir. Yeniden örneklemenin başka yolları da vardır ve ileride bu yöntemler üzerinde de durulacaktır ancak özyüklemede gözlenen veriyi çoğaltarak örneklem elde edilir ve bu durum sonsuz bir evrenden örnekleme ile aynı etkiye sahiptir. Özyükleme yöntemi genel olarak tekrar eden örneklemlerde bazı istatistiklerin değişkenliğini kestirmek için kullanılır, ancak burada yaptığımız gibi bazı hipotezleri test etmede de kullanılabilir. Bu örneği burada vermemin nedeni, artık birçok istatistiksel yazılım simülasyon, özyükleme, "yeniden örnekleme" veya "tesadüfî hale getir" gibi seçenekler sunuyor ve bu seçeneklerin tümü aşağı yukarı verdiğimiz örnekte yapılan işi yapıyor. 20 yıl sonra Student'ın *t* testinin "antika" olarak anılacağını söylesek yanlış yapmış olmayız.[8]

8 Standart *t* testi yerine kullanılabilecek başka ve daha iyi tekrar örnekleme yöntemleri vardır, bunlara geleceğim. Şimdi sadece, problemi çözmek için mantıklı alternatif yolların bulunduğunu göstermek istiyorum.




**Şekil 7.5** Düşük kilolu doğumlar örneği için teorik *t* dağılımı ile birlikte verilmiş *t* değerinin deneysel örnekleme dağılımı

## Ay Yanılsaması

Kaufman ve Rock'ın (1962) ay yanılsaması[9] üzerine yaptıkları çalışma üzerinden ikinci bir örneği vermek yararlı olacaktır. Ay yanılsaması, psikologları yıllardır etkileyen bir konudur ve Ay'ı ufukta gördüğümüzde tepede olduğundan çok daha büyük algılamamız olarak ifade edilir. Kaufman ve Rock bu yanılsamayı, Ay ufuk çizgisinde olduğunda daha büyük görünüş uzaklığında olması varsayımına dayandırır. Art arda yapılan birçok deneyin bir parçası olarak, Ay yanılmasını kestirebilmek için katılımcılardan ufukta beliren Ay ile tepedeki Ay'ın büyüklüklerini karşılaştırmaları istenmiştir. (Bu deney için gerçek bir Ay değil özel cihazlarla elde edilmiş yapay bir Ay kullanılmıştı.) Öncelikle sormamız gereken sorulardan biri, bu yapay aletlerin Ay yanılsamasını oluşturup oluşturamadığı, bir diğer anlatımla ufuktaki ve tepedeki Ay'ın büyüklüklerinin karşılaştırılması için çok daha büyük bir düzenek gerekli olup olmadığı olmalıdır. Aşağıdaki 10 ölçüme ait veri kümesi Kaufman ve Rock'ın çalışmasından alınmıştır ve veride ufukta beliren Ay (değişken) ile tepedeki Ay'ın çap büyüklükleri oranı yer almaktadır. Oran 1 olduğunda Ay yanılsaması olmadığı ancak 1'den farklı olduğunda Ay yanılsaması olduğu sonucuna varılacaktır. (Örneğin oran 1,50 ise ufuktaki Ay'ın çapı tepedeki Ay'ın çapının 1,50 katı olacaktır.) Yanılsamanın varlığının kanıtı olarak $H_0 : \mu = 1{,}00$ olan sıfır hipotezi reddedilmeli ve $H_A: \mu \neq 1{,}00$ alternatif hipotez kabul edilmelidir. (Burada sıfır hipotezi evren ortalaması olarak sıfırdan farklı bir değer konulduğu için beklenmedik bir durum söz konusudur. Ancak formülde $\mu$ için doğru değer yerine konulduğunda bu durum herhangi bir probleme neden olmaz.)

---

9 Bu konu üzerinde son zamanlarda yapılmış çalışma Lloyd Kaufman ve oğlu James Kaufman tarafından Proceedings of the National Academy of Sciences dergisinin Ocak 2000 sayısında basılmıştır.

Veri kümesi için $n = 10$, $\overline{X}$=1,463 ve $s = 0{,}341$'dir. $H_0 : \mu$ =1,00 olduğuna göre $t$ değeri;

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$= \frac{1{,}463 - 1{,}000}{\frac{0{,}341}{\sqrt{10}}} = \frac{0{,}463}{0{,}108}$$

$$= 4{,}29$$

Ek $t$'den $10 - 1 = 9$ serbestlik derecesi ve iki yönlü $\alpha$ = ,05 için kritik değerin $t_{,025}(9) = 2{,}262$ olduğu görülür. Elde edilen $t$ değeri 4,29'dur. 4,29 değeri 2,262 değerinden büyük olduğu için yokluk hipotezi ,05 düzeyinde reddedilir ve gerçek ortalamaların oranının 1'e eşit olmadığı sonucuna varılır. Hâttâ yapılan deneyden, beklendiği gibi, gerçek ortalamanın oranı 1'den büyüktür. (Çocukluktan beri bildiğimiz şeylerin bilimle onaylanması her zaman rahatlatıcı olmuştur, ancak elde edilen sonuçlar ayrıca Kaufman ve Rock'ın hazırladığı deneysel düzeneğin de olması gerektiği gibi çalıştığını göstermiştir.) Teknolojiyi sevenler için http://www.danielsoper.com/statcalc/calc40.aspx adresinde olasılık hesaplama makinesi yer almaktadır ve yapılan hesaba göre iki yönlü olasılık ,001483'tür.

## Evren Ortalamasının Güven Aralığı

Güven aralığı, basit hipotez testlerinin önüne geçen deneysel sonuçların anlamını yaymak için en kullanışlı yoldur. Ay yanılsaması örneğinde evren ortalaması, ufuktaki Ay ile tepedeki Ay'ın çaplarının oranı idi ve deneydeki veri, evren ortalamasının gerçek kestirimiyle ilgilenildiği durumlar için mükemmel bir örnekti. Bilindiği gibi örneklem ortalaması ($\overline{X}$) burada μ'in yansız kestirimidir. Parametreye ait özel bir kestirim varsa buna **nokta kestirim** denir. Ayrıca evrenin gerçek ortalamasının yüksek olasılıkla alabileceği sınır değerlerini tanımlayan bir aralık kestirimi söz konusudur. Bu durumda gerekli olan şey evren ortalamasının **güven sınırları**nı belirlemektir. Bu sınırlar **güven aralığı**nı kapsar.[10] Bölüm 3'te gözlemler için olası sınırlardan bahsedilmişti. Parametrelerin güven sınırları belirlenirken de aynı gerekçeler söz konudur.

**nokta kestirim**

**güven sınırları**

**güven aralığı**

Eldeki veri kümesinin, evren ortalamasını içeren güven sınırlarının belirlenmesi isteniyorsa, gözlenen örneklem ortalaması ile $t$ testi yaptığımızda $H_0$'ın reddedilmemesi için evren ortalamasının ne kadar büyük veya küçük olması gerektiğini sormalıyız. Örneğin Ay yanılması ile ilgili evren ortalamasının 1,00'e eşit olduğunu varsayan yokluk hipotezini test ettiğimizde, bu hipotez reddedilmişti. Peki ya yokluk hipotezi, evren ortalamasının 0,93'e eşit olduğunu varsayarsa? Bu durumda yine yokluk hipotezi reddedilir. Evren ortalamasının değerini yokluk hipotezini reddetmeyecek noktaya kadar düşürmeye devam edebiliriz ve yokluk hipotezinin ret edilmediği bu değer ,05 manidarlık düzeyinde evren ortalamasının en küçük değeridir. Daha sonra evren ortalamasının değeri büyütülür (örneğin, 2,2) ve bu büyütmeye yokluk hipotezinin reddedilmediği noktaya kadar devam edilir. Yokluk hipotezinin reddedilmediği bu değer ,05 düzeyinde evren ortalamasının en büyük değeridir. Bu durumda evren ortalamasının alt ve üst sınırları arasına düşen her değer için yokluk hipotezi kabul edilir. Güven sınırlarının bu şekilde belirlenebilmesinin yanında, aynı cevabı verebilecek daha kısa bir yol vardır.

---

[10] Güven aralığı ve güven sınırları benzer kavramlar olarak kullanılır. Ancak güven sınırları güven aralığının başladığı ve bittiği noktalardır. Birbirlerinin yerine kullanıldığında bu ikisi karıştırılmamalıdır.

Kolay yol, tek örneklem *t* testi formülü ile başlar:

$$t = \frac{x - \mu}{s_{\overline{X}}} = \frac{x - \mu}{\frac{s}{\sqrt{n}}}$$

Ay yanılması örneğindeki veriden $\overline{X}$ = 1,463, $s$ = 0,341 ve $n$ = 10 olduğunu biliyoruz. Ayrıca *t* testi için iki yönlü kritik *t* değerinin $\alpha$ = 0,05 manidarlık düzeyi için $t_{,025}(9)$ = 2,262 [veya $t_{,05}$ (9) ± 2,2621) olduğunu da biliyoruz. Bu değerler *t* eşitliğinde yerine konulup eşitlik *t* yerine $\mu$'ye göre çözülerek ve *t*'nin kritik değeri için evren ortalamasının alacağı değerler hesaplanacaktır.

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \qquad \pm 2,262 = \frac{1,463 - \mu}{\frac{0,341}{\sqrt{10}}} = \frac{1,463 - \mu}{0,108}$$

Bu ifade $\mu$'ye göre düzenlenerek

$\mu$ = ±2,262 (0,108) + 1,463 = ± 0,244 + 1,463

ve eşitlik +0,244 ve -0,244 için ayrı ayrı yazılarak

$\mu_{\text{üst}}$ = +0,244 + 1,463 = 1,707

$\mu_{\text{alt}}$ = -0,244 + 1,463 = 1,219

elde edilir. Bu durumda %95 güven sınırları 1,219 ve 1,707 olarak yazılabilir ve güven aralığı da

$GA_{,95}$ =1,219 ≤ $\mu$ ≤ 1,707

biçiminde yazılabilir. Evren ortalaması için bu aralığın dışında kalan değerler elde edildiğinde yokluk hipotezi reddedilirken bu sınırlar içerisindekiler $\mu$ için kabul edilir. Güven aralığının genel ifadesi aşağıdaki gibidir:

$$GA_{1-\alpha} = \overline{X} \pm t_{\alpha/2}(s_{\overline{X}}) = \overline{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Verilen örnekte %95 olasılıklı güven aralığı söz konusudur, çünkü $\alpha$ = ,05 manidarlık düzeyinde iki yönlü kritik değer kullanılmıştır. %99 sınırları için $t_{,01/2} = t_{,005}$ = 3,250'dir. Bu durumda %99 güven aralığı aşağıdaki gibi yazılır:

$$GA_{,99} = \overline{X} \pm t_{,01/2}(s_{\overline{X}}) = 1,463 \pm 3,250(0,108) = 1,112 \leq 1,814$$

## Güven Aralığı Nedir?

"Evren ortalaması için %95 olasılıklı güven aralığı 1,219 ≤ $\mu$ ≤ 1,707'dir." ifadesinin anlamı nedir? Her iki kitabın yedi ayrı basımı için bu soru zihinleri karıştırmıştır. Ay yanılsaması örneği için gerçek ortalama oranının %95 olasılıkla 1,219 ile 1,707 aralığında ve %99 olasılıkla 1,112 ile 1,812 aralığında olduğunu söylemek oldukça çekicidir. Ancak birçok istatistikçi güven aralığının bu şekilde ifade edilmesine karşıdır. İstatistikçilere göre *deney ve hesaplamalar yapılmadan* önce evren ortalaması için %95 olasılıkla aralığının formu şu şekildedir:

$$\overline{X} \pm t_{,025}(s_{\overline{X}})$$

Ancak evren ortalaması bilinmemesine rağmen sabit bir özelliktir ve veri kümesi elde edildiğinde değeri ya özel aralık olan 1,219 ile 1,707 arasındadır ($p$ = 1,00) ya da değildir ($p$ = ,00). Bu düşünceye karşı çıkan istatistikçiler "aralığın belirlenirken ,95 olasılıkla evren ortalamasını kapsayacak şekilde yapılandırıldığı" fikrinde ısrar edeceklerdir. Söylenilen bu fikirler kulağa aynı gibi gelebilir ancak değildir ve bu fikri savunanlar oldukça katı görüşlüdürler Ancak Good (1999) farklı bir noktaya değinmiştir ve güvenlik düzeyinin *aralık*la değil *metot* ile ilgili olduğunu savunmuştur. Farklı bir bakış açısı ile;

$$\overline{X} \pm t_{,025}(s_{\overline{X}})$$

bir tesadüfî değişkendir (bir deneyden diğerine farklı değerler alabilir) ancak 1,219'dan 1,707'e kadar olan aralık değişken değildir ve bu nedenle de herhangi bir olasılıktan bahsedilmesi mümkün değildir. Diğer yandan olasılık kavramını kişisel inanışlar üzerinden ortaya çıkaran Bayes istatistikçileri "evren ortalaması ,95 olasılıkla 1,219 ile 1,707 arasındadır" demekten çok büyük bir mutluluk duyarlar ancak aralık için "**inanılır aralık**" terimini kullanırlar. Birçok açıdan güven aralığı ile inanılır aralık benzer özellikler gösterir; ancak bazı durumlarda olasılık kavramının doğası gereği farklılaşır. Sonuç olarak bu kitap için Bayes istatistikçilerinin savundukları fikirde gitme kararı aldım.[11] Gelenekçilerin yolunda gidildiğinde öyle cümlelerle karşılaşılıyor ki durumu açıklamaktan çok, karışıklığa yol açıyor. Onlar burada anlatılan istatistikleri eleştireceklerdir; ancak öncelikle açıklananları anlamaları lâzımdır. Bu konu hakkında yazılmış iyi çalışmalar için Dracup (2005) ve Masson ve Loftus (2003) önerilir.

**inanılır aralık**

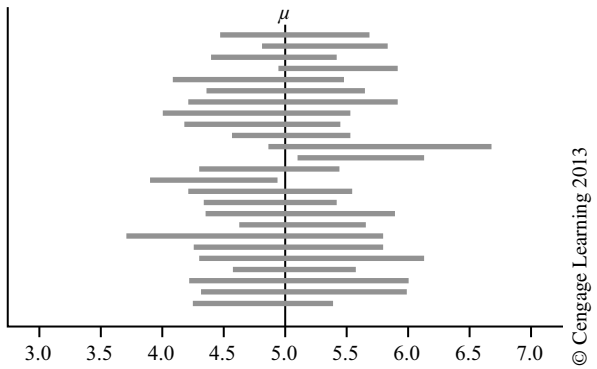
Unutulmamalıdır ki hesaplanan güven aralıkları Ay yanılsamasının olmadığı anlamına gelen 1,00 değerini ne %95 ne de %99 manidarlık düzeyinde içerir. ,95 manidarlık düzeyinde Ay yanılsaması olmadığı zaten bilinmektedir, çünkü bu düzeyde $t$ testi yapıldığında yokluk hipotezi reddedilmiştir.

Güven sınırlarını yorumlamanın bir başka yolunu eklemeliyim. $\mu$ parametresi deneyden deneye farklı değerler almadığı için değişken değildir. $\mu$ bir sabittir, ancak deneyden deneye farklılık gösteren güven aralığıdır. Bu nedenle parametre bir çubuk/kazık, güven sınırlarını hesaplama işi ise ona halka atma olarak düşünülebilir. %95 olasılıkla halkalar kazığa geçecektir; %5 olasılıkla ($\mu$) ıskalanacaktır. Güven ifadesi direğin halkayla dolma olasılığı değil halkanın hedefe ulaşma olasılığıdır.

Güven sınırlarına ait grafik Şekil 7.6'da gösterilmiştir. Şekli oluşturmak için ortalaması ($\mu$) 5 olan bir evrenden $n$ = 4 olan 25 örneklem çektim. Her örneklem için evren ortalamasının %95 güven sınırları hesaplanmış ve grafikte gösterilmiştir. Örneğin ilk örneklem için sınırlar 4,46 ve 5,72 (şekildeki birinci doğru) ikinci örneklem için ise 4,83 ve 5,80 bulunmuştur. μ'nun değeri bilindiğinden ve değeri 5 olduğundan, bu noktayı dik bir çizgi ile gösterdim. Örneklem 12 ve 14'ün sınırlarının 5'i içermediğine dikkat edilmelidir. Güven sınırlarının %95 olasılıkla 5 değerini kapsadığı varsayıldığı için 25 örneklemden ikisinin sınırları aşması mantıklı görülebilir. Ayrıca güven sınırlarının farklı genişliklerde olduğuna dikkat edilmelidir. Bu farklılaşmanın temel sebebi aralığın genişliğinin örneklemin standart sapmasının bir fonksiyonu olmasıdır ve bazı örneklemlerin standart sapmaları diğerlerine göre daha büyüktür.

---

11 Pek çok saygın istatistikçi güven aralığını burada açıkladığım şekilde tanımlar ve bunu yaparken yüzleri bile kızarmaz.




**Şekil 7.6** Evren ortalaması 5 ($\mu = 5$) olan 25 örneklemden elde edilen güven aralıkları

## Uç Durumları Belirleme

Bölüm 3'te uç durumları belirlemek için normal dağılım özelliklerinin kullanılabileceğini gördük. Örneğin Achenbach'ın çocuk davranışları kontrol listesi ile çalışıyorken evren parametreleri olarak ortalamayı 50 ve standart sapmayı 10 aldık. Puanı, ortalamanın 1,64 standart sapma üzerinde çıkan çocuklar için uç değer ifadesi kullanılır. Bu durumda;

$$z = \frac{X - \mu}{\sigma}$$

$$1,64 = \frac{X - 50}{10}$$

$$X = 1,64(10) + 50 = 64,4$$

Problem Davranış Kontrol Listesi'nden 64,4'ten büyük puan alan çocuklar uygulama için adaylardır.

Tom Achenbach çalışmasında, evren ortalamasını 50 ve standart sapmasını 10 olarak kullanmanın herhangi bir sakıncası olmayacak kadar büyük örneklemler kullandığı için problemi bu şekilde çözebiliriz. Benzer durum normları çok büyük örneklemler üzerinden elde edilmiş Beck Depresyon Envanteri ve Wechler Zekâ Testi için de geçerlidir. Ancak pek çok durumda, örneğin nöropsikolojide, normların kestirildiği örneklemlerin küçük olduğu birçok durum söz konudur. Örneğin İşitsel Sözel Öğrenme Testi'nde (Geffen, Moar, O'Hanlon, Clark ve Geffen; 1990) normlar 153 kişi üzerinden kestirilmiştir ancak puanlar üzerinde cinsiyet ve yaş değişkenlerinin önemli etkileri vardır (yaş 16'dan 86'ya değişmektedir) ve bu değişkenlere göre normları belirlemek istediğinizde örneklem büyüklüğü 10 olur. 10 puanın ortalamasını ve standart sapmasını ele almak ve bu değerleri evren parametreleri olarak değerlendirmek akıllıca olmayabilir. (Küçük örneklemlerin varyans dağılımlarının pozitif kayışlı olduğunu hatırlayınız. Bu durumlarda *s* için alt kestirimi elde ederiz ve buna bağlı olarak *z* değerini büyük buluruz.) Benzer şekilde son zamanlarda tek denekli çalışmalar çoğalmıştır ve bu çalışmaların büyük örneklemler üzerinden normlarını elde etmek çok zaman alıcıdır. Peki bu durumlarda ne yapılmalıdır?

Crawford ve Howell (1998) yaptıkları çalışma ile bu probleme eğilmişler ve *z* hesaplarken *t* dağılımının kullanılabileceğini göstermişlerdir. (Ayrıca Crawford, Gartwaite ve Howell'ın (2009) yaptıkları çalışmaya da bakınız.) Bu çalışmaya göre küçük normatif örnekleme ait ortalama ($\overline{X}$) ve standart sapma (*s*) değerleri alınabilir ve kişinin aldığı puan aşağıdaki formülde yerine konulduğunda *t* değeri bulunur.

$$t = \frac{X_1 - \overline{X}}{s\sqrt{\frac{n+1}{n}}}$$

Burada *t, n* – 1 serbestlik derecesindedir.

Örneğin uzaysal hafıza testinin norm değerlerinin kestirileceği 11 uygulamadan oluşan bir örneklemimiz olduğunu, ortalamanın 45 ve standart sapmanın 9 olarak hesaplandığını düşünülelim. Bu testten 31 puan almış bir hasta için *t* değeri;

$$t = \frac{X_1 - \overline{X}}{s\sqrt{\frac{n+1}{n}}} = \frac{31-45}{9\sqrt{\frac{11+1}{11}}} = \frac{-14}{9\sqrt{\frac{12}{11}}} = \frac{-14}{9,4} = -1,49$$

Normatif örneklemin %95'inden düşük puan alındığında, bu puanın bir uç değer olduğu söyleniyorsa, bu hipotez (*n*-1) = 10 serbestlik derecesi ile tek yönlüdür. 0,95 düzeyinde 10 serbestlik derecesi için kritik değer $t_{,05(10)} = 1,81$'dir, bu durumda ele alınan durum uç değer değildir.

Bu noktada, ele alınan durumun ideal örneklemden daha küçük ortalamaya sahip olan evrenden gelip gelmediğinin araştırılmadığının altı çizilmelidir. Burada tek bir durum söz konusudur ve varsayılan evrenden gelip gelmediği düşünülmemiştir. Evren ne olabilir? Bunun yerine elde edilen puan kadar uç bir değerin ideal evrende olma olasılığı araştırılmıştır.

## Son Örnek: Terapatik Dokunuşla İşimiz Bitmedi

Son bölümde terapatik dokunuşla ilgili bir örnek üzerinde tartıştık ve bağımsızlık varsayımı sağlanmadığı için yokluk hipotezini ki-kare testi ile yoklamanın sağlam temellere dayanmadığı yorumunu yaptık. Bu bölümde yokluk hipotezini test etmenin farklı yolu üzerinde duracağız. Bu yol çalışmanın yazarlarının veriyi incelerken izledikleri yola benzerdir.

Terapatik dokunuş deneyi 28 oturumdan oluşmaktadır ve her oturumda denek hangi elin üzerinde olduğuna dair 10 tahmin yürütür. Denek 10 deneme içinden doğru bildiği her durum için 1 puan olmak üzere toplamda 0 ile 10 arasında puan alır. Şans başarısı için 10 deneme üzerinden 5 doğru kestirimin ortalaması dikkate alınır. Aşağıda her oturumdan elde edilmiş veri ve hem elle hesaplama hem de R veya S-PLUS programları ile yapılan analizler yer almaktadır.

| Oturum | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Doğru Sayısı | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |

| Oturum | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Doğru Sayısı | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 8 |

Ortalama = 4,393
$s$ = 1,663
$n$ = 28

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{4,393 - 5,00}{\frac{1,663}{\sqrt{28}}} = \frac{-0,607}{0,314} = -1,93$$

R programı için bilgisayar çıktısı
veri: doğru sayısı
$t$ = –1,913,18, $sd$ =27, $p$-değeri = 0,06395
alternatif hipotez: gerçek ortalama 5'e eşit değildir.
%95 olasılıklı güven aralığı:
3,747978 – 5,037737
örneklem kestirimleri:
$X$ ortalama
4,392857

Denekler 10 üzerinden 5 doğru cevap verdiklerinde 5 beklenen değerine göre şansla doğru cevaplamışlardır. Bu nedenle yokluk hipotezi $\mu = 5$, alternatif hipotez $\mu \neq 5$ olacaktır. Bu durumda, $t = -1{,}93$, $sd = 27$ ve $p = ,064$ olduğundan yokluk hipotezini reddedemeyeceğiz. Deneklerin şans başarısının dışında telepati ile doğru cevapladıkladıklarına dair bir kanıtımız yoktur. Hatta denekler şansa dayalı doğru cevaplama oranının bile altında doğrular elde etmişlerdir.

Manidarlık testi araştırılan temel soruya cevap olsa bile evren ortalamasının güven aralığını hesaplamak yararlı olacaktır. Duruma ait güven aralığı aşağıdaki gibi hesaplanabilir:

$$GA = \overline{X} \pm t,025 s_{\overline{X}}$$
$$GA = X \pm 2{,}052\,(0{,}314)$$
$$= 4{,}393 \pm 0{,}644$$
$$= 3{,}750 < \mu < 5{,}038$$

Beklenildiği gibi güven aralığı $\mu = 5$ değerini kapsamaktadır.

## Tek örneklem *t*-testi için SPSS

İstatistiksel problemleri çözmek için genellikle bilgisayar programları kullanılır. Bunlardan biri *t* değerini hesaplamak için SPSS programını kullanmaktır. Gösterim 7.1'de Ay yanılsaması verisi üzerinden tek örneklem *t* testi ve güven aralıklarını hesaplamak için SPSS'in nasıl kullanılacağı örneklendirilmiştir. Ay yanılsaması verisine ait *t* değerini hesaplamak için SPSS'te sırasıyla **Analyze/Compare Means/One Sample t Test** sekmeleri tıklanır ve açılan arayüzde bağımlı değişken belirlenir. SPSS ile elde edilen sonucun elle hesaplamayla elde edilenle yaklaşık olarak aynı olacağı unutulmamalıdır. Ayrıca, tablodan elde edilen kritik değer ile karşılaştırma yerine SPSS ile I. Tip hatanın olasılığının **(*p* düzeyi)** kesin değeri hesaplanabilir. I. Tip hatanın olasılık değeri ,05'ten küçük ise SPSS'e göre gerçek olasılık ,0020'dir. Birçok bilgisayar programı bu mantık üzerine çalışır.

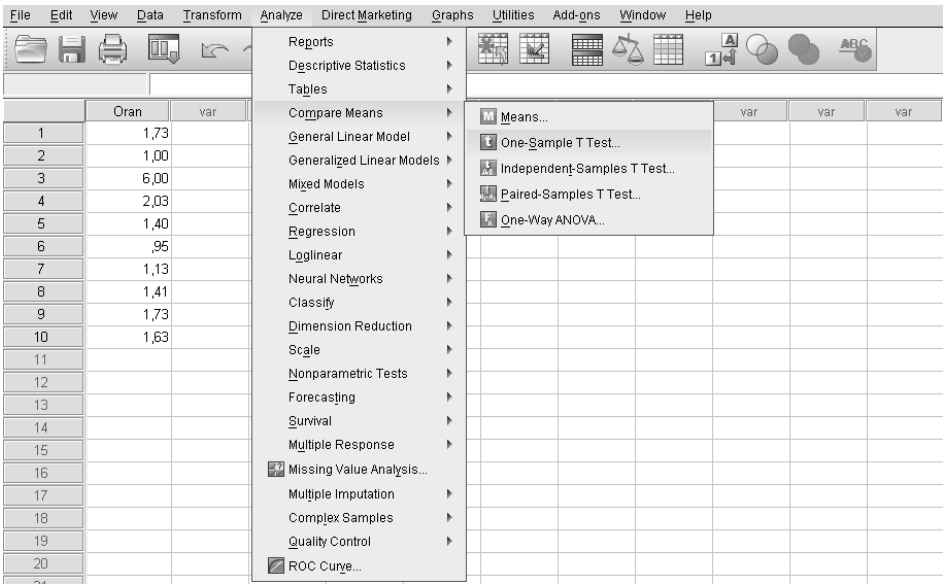
***p* düzeyi**

Ancak güven aralığı hesaplamada, el ile hesaplamayla SPSS ile hesaplama arasında fark vardır ve her ikisi de doğrudur. El ile hesaplamada Ay yanılsaması ortalamasının kestirim değeri olarak 1,463 değeri kullanılmışken SPSS'te 1,463 ile gerçek yanılsama ortalaması olan 1,00 arasındaki fark kullanılmıştır. Bir diğer deyişle, elle hesaplama yapılırken 1,463 çevresinde, SPSS ile hesaplamada 0,463 çevresinde (1,463 – 1,00 = 0,463) sınırlar elde edilmiştir. Bu nedenle SPSS ile elde edilen güven sınırları el ile hesaplama ile elde edilenlerden 1,00 değerinde daha küçüktür. İki sürecin hesaplanmasındaki farklılıklar anlaşıldığında sonuçların farklılığı da anlaşılır olacaktır.[12]

# 7.4 Ortalamalara Uygulanan Hipotez Testleri – Eşleştirilmiş İki Örneklem

Bölüm 7.3'te tek bir örneklem ortalamasının olduğu ($\overline{X}$) ve bu ortalamanın evren ortalamasından ($\mu_0$) manidar şekilde farklı olup olmadığının test edildiği durumları ele almıştık. Başka bir deyişle, Bölüm 7.3'te örneklemin alındığı evren ortalamasının ($\mu_1$) yokluk hipotezinde verilen değere ($\mu_0$) eşit olup olmadığını araştırmıştık. Bu bölümde

[12] SPSS ile güven aralıklarını hesaplamak için sırasıyla **Analyze/Descriptive statistics/Explorer** sekmeleri kullanılmalıdır.




**Tek-Örneklem İstatistikleri**

| | N | Ortalama | Std. Sapma | Ortalamanın Std. Hatası |
|---|---|---|---|---|
| **Oran** | 10 | 1,4630 | ,34069 | ,10773 |

**Tek-Örneklem Testi**
Test değeri = 1

| | t | Sd. | Manidarlık (2-yönlü) | Ortalama Fark | %95 Farkın Güven Aralığı | |
|---|---|---|---|---|---|---|
| | | | | | Alt Sınır | Üst Sınır |
| **Oran** | 4,298 | 9 | ,002 | ,4630 | ,2193 | ,7067 |

**Gösterim 7.1** Tek örneklem *t* testi ve güven aralığı için SPSS çıktısı

**eşleştirilmiş örneklemler**

**tekrarlı ölçümler**

**bağımlı örneklem**

**eşleştirilmiş örneklem t-testi**

**eşleştirilmiş iki örneklem**in (aynı kişilerin iki durum için ölçümlerinin yapıldığı bazı durumlarda **tekrarlı ölçümler**, bazı durumlarda ilişkili örneklemler veya **bağımlı örneklemler** olarak isimlendirilir) var olduğu durumlar ve iki ortalamanın farkı incelenecektir. Yapılan bu işleme **eşleştirilmiş örneklem t-testi** adı verilir.

## Anoreksiya Tedavisi

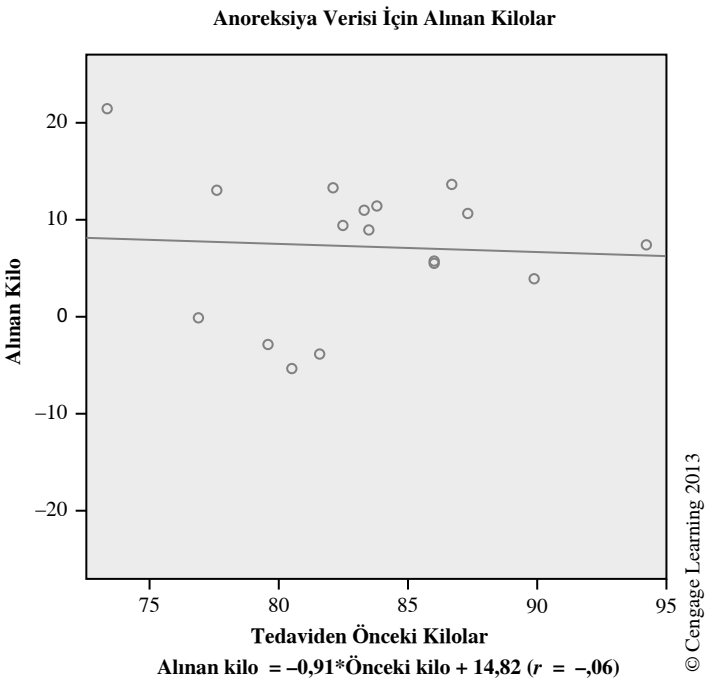
Everitt, 1994 yılında yayımladığı raporda aile terapisinin anoreksiya tedavisindeki etkileri üzerinde durmuştur. Yapılan deneyde 17 kız vardır ve denekler tedaviden önce ve sonra tartılmışlardır. Tablo 7.3'te kızların kiloları bulunmaktadır. (Birim olarak pound kullanılmıştır. 1 pound = 0,45359237 kg.) Fark sütunu tedavi sonrasında ölçülen kiloların tedaviden önceki kilolarından farkını gösterir. Bu nedenle farkın negatif olması kilo kaybına, pozitif olması deneğin kilo aldığına işaret eder.

**Tablo 7.3** Everitt'in kilo almaya ait veri kümesi

| Kişi | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Önce** | 83,8 | 83,3 | 86,0 | 82,5 | 86,7 | 79,6 | 76,9 | 94,2 | 73,4 | 80,5 |
| **Sonra** | 95,2 | 94,3 | 91,5 | 91,9 | 100,3 | 76,7 | 76,8 | 101,6 | 94,9 | 75,2 |
| **Fark** | 11,4 | 11,0 | 5,5 | 9,4 | 13,6 | -2,9 | -0,1 | 7,4 | 21,5 | -5,3 |

| Kişi | 11 | 12 | 13 | 14 | 15 | 16 | 17 | Ortalama | Std. sapma |
|---|---|---|---|---|---|---|---|---|---|
| **Önce** | 81,6 | 82,1 | 77,6 | 83,5 | 89,9 | 86,0 | 87,3 | 83,23 | 5,02 |
| **Sonra** | 77,8 | 95,5 | 90,7 | 92,5 | 93,8 | 91,7 | 98,0 | 90,49 | 8,48 |
| **Fark** | -3,8 | 13,4 | 13,1 | 9,0 | 3,9 | 5,7 | 10,7 | 7,26 | 7,16 |

© Cengage Learning 2013




**Şekil 7.7** Anoreksiya hastası 17 kızın aile terapisinden önceki ve sonraki kiloları arasındaki ilişki

Öncelikle yapılması gereken iş, her ne kadar bizi bir an için *t* testinden uzaklaştırsa da, tedaviden önceki kilo değerleri ile alınan kilolar arasındaki ilişkiyi belirlemektir. Böylelikle iki değişken arasında ilişki var mı, varsa ne kadar doğrusal bir ilişki olduğu belirlenir. Şekil 7.7'ye göre, alınan kilolar ile tedaviden önceki kilolar arasındaki ilişki, eğim sıfıra yakın olduğu için doğrusaldır. Bu durumda terapiden önce deneğin kilosunun ne kadar olduğu terapiden sonra ne kadar kilo kaybedeceği ya da alacağını etkilemez. (Regresyon doğrusu ve regresyon eğimi konularının üzerinde Bölüm 9'da durulacaktır.)

Cevabı merak edilen temel soru terapi seanslarının deneklere kilo aldırıp aldırmadığıdır. Ancak burada deneysel bir sorun yer almaktadır, çünkü kilo alma terapiden bağımsız

olarak zamanla da oluşabilecek bir durumdur. Ancak Everitt'in deneyinde yer alan diğer veri incelendiğinde, aile terapisi almayan anoreksiya grubun aynı süre içerisinde hiç kilo almadığı görülmüştür ve böylelikle zaman değişkeninin deney için önemli olmadığı kanıtlanmıştır. Kişilerin kiloları terapiden önce ve sonra hesaplanmış ve ortalamaları sırasıyla 83,23 ve 90,49 lbs. olarak bulunmuştur. Bu durum, "terapi ortalama 7 lbs. kilo kazanmayı sağlıyor" şeklinde yorumlanabilir. Ancak hâlâ bu farkın evrendeki gerçek farktan mı yoksa şanstan mı kaynaklandığının açıklanması gereklidir. Bunun için evrende terapiden önceki değerlerin terapiden sonraki değerlere eşit olduğunu savunan yokluk hipotezi test edilmelidir. Bir diğer deyişle $H_0 : \mu_{ö} = \mu_{s}$ eşitliği test edilmelidir.

## Fark Puanları

**fark puanları**

**kazanılan puanlar**

Veri kümesi terapiden önce ve sonra olmak üzere iki örneklem puanı şeklinde ifade edilebileceği gibi, her bir kişinin önceki ve sonraki puanları $X_1$ ve $X_2$ ile gösterilip farkı alınarak tek bir puan türünde ifade edilebilir. Bu puan türüne **fark puanları** veya **kazanılan puanlar** denir; bu puanlar Tablo 7.3'ün üçüncü satırında gösterilmiştir. Tablodaki fark puanları treapide alınan kilo cinsinden değerlerini gösterir. Eğer terapinin anoreksiya üzerinde bir etkisi yoksa terapiden önce ve sonra ortalama kilonun değişmemesi gerekir. Bazı katılımcıların terapiden sonraki ölçümlerde öncekine göre daha kilolu, bazılarının ise daha zayıf olması söz konusudur; ancak ortalamaları alındığında önemli bir fark olmaması beklenir.

Eldeki verinin fark puanlarından oluştuğu bilindiğine göre yokluk hipotezi, "fark puanlarının evren ortalaması ($\mu_F$) sıfırdır" şeklinde ele alınabilir. $\mu_F = \mu_1 - \mu_2$ olduğu için yokluk hipotezi $H_0 : \mu_F = \mu_1 - \mu_2 = 0$ şeklinde gösterilebilir. Bu durumda, daha önce üzerinde durulan, tek örnekleme dayalı bir hipotez test edilecektir.

## *t* İstatistiği

Şu an, bir önceki bölümde olduğu gibi, tek örnekleme ait veri ve "evren ortalaması sıfıra eşittir" varsayımında bulunan bir yokluk hipotezi ($\mu = 0$) testi elde edilmiştir. Ancak bir önceki bölümde yapılan *t* testinden farkı, verinin fark puanları ve ortalama ve standart sapmasının bu fark puanlarına dayalı olmasıdır. Hatırlanacak olursa, *t* istatistiği evren ortalaması ile örneklem ortalaması arasındaki farkın, ortalamanın standart hatasına oranıdır. Bu durumda fark puanlarının ortalaması, $\overline{F}$ fark puanlarının ortalaması, $s_F$ fark puanlarının standart sapması ve *n* fark puanları sayısı olmak üzere *t* değeri aşağıdaki formül ile elde edilir:

$$t = \frac{\overline{D} - 0}{s_{\overline{D}}} = \frac{\overline{D} - 0}{\frac{s_D}{\sqrt{n}}}$$

Tablo 7.3'teki bilgiye göre, fark puanlarının ortalaması 7,26 ve standart sapması 7,16'dır. Bu değerler yukarıdaki formülde kullanılarak *t* değeri aşağıdaki gibi hesaplanır:

$$t = \frac{\overline{F} - 0}{s_{\overline{F}}} = \frac{\overline{F} - 0}{\frac{s_F}{\sqrt{n}}} = \frac{7,26 - 0}{\frac{7,16}{\sqrt{17}}} = \frac{7,26}{1,74} = 4,18$$

## Serbestlik Derecesi

Eşleştirilmiş örneklem *t* testinin serbestlik derecesi tek örneklem *t* testi için ifade edilen serbestlik derecesi ile tamamen aynıdır. Fark puanları ile çalıştığımızdan *n* fark sayısına (veya aynı değere sahip gözlem çifti sayısına veya bağımsız gözlem sayısına) eşit olacaktır. Fark puanlarının varyansı fark puanları evreninin varyansını kestirmek için kullanıldığından ve örneklem varyansı örneklem ortalaması kullanılarak hesaplandığından ortalamanın serbestlik derecesinden 1 eksilecektir ve serbestlik derecesi *n*-1 olacaktır. Bir diğer deyişle *sd* = gözlem çiftleri sayısı eksi 1'dir.

Bu örnekte 17 fark puanımız vardır, bu nedenle serbestlik derecemiz 16'dır. Ek *t* incelendiğinde ,05 manidarlık düzeyinde iki yönlü *t* değerinin 2,12 olduğunu görürüz. Hesaplamalardan elde ettiğimiz *t* istatistiği (4,18) bu değerden büyük olduğu için $H_0$ reddedilir ve fark puanlarının, evren ortalaması sıfır ($\mu_F$= 0) olan bir evrenden örneklenmediği sonucuna varırız. Buna göre yapılan terapi sonrasında anoreksiyalı deneklerde kilo kazanımı olmuştur. Sonuçlara göre, terapi programının başarılı olduğu düşünülse de, bu durumun normal bir büyüme olma olasılığı da unutulmamalıdır. Ancak her ne sebeple olursa olsun ikinci ölçümdeki değerlerin daha yüksek olması sebebiyle yokluk hipotezi ($H_0 : \mu_F = \mu_1 - \mu_2 = 0$)[13] reddedilir.

## Güven Aralıkları

Terapi öncesi ve sonrasında yapılan ölçümler arasında manidar bir fark çıkması, terapinin iyi ve faydalı olduğunun bir göstergesidir, ancak yine de farkın güven aralığını belirlemek gerekir. Bunun için tek örneklem *t* testinde yapılan işlemlerin aynısı yapılır; *t* denklemi yazılır, 16 serbestlik derecesi için kritik değerler tablodan bulunur ve evren ortalaması için denklem çözülür.

$$t = \frac{F - \mu}{s_F/\sqrt{n}}$$

$$\pm 2,12 = \frac{7,26 - \mu}{7,16/\sqrt{17}} = \frac{7,26 - \mu}{1,74}$$

$$\mu = \pm 2,12 \times 1,74 + 7,26 = \pm 3,69 + 7,26$$

$$7,26 - 3,69 \leq \mu \leq 7,26 + 3,69$$

$$3,57 \leq \mu \leq 10,95$$

Elde edilen sonuçlara göre, terapinin en az etki gösterdiği kişide bile 3,5 poundluk kilo artışı yaşanmıştır. Bir diğer anlatımla terapi öncesi 83,23 olan ortalama %4 oranında artmıştır.

## Tekrar Ay Yanılsaması Örneği

Bağımlı örneklem *t* testi için ikinci örnek olarak Kaufman ve Rock'ın (1962) yaptığı Ay yanılsaması çalışmasına döneceğiz. Ay yanılsamasının neden oluştuğu ile ilgili önemli bir hipotez Holway ve Boring (1940) tarafından ortaya atılmıştır. Holway ve Boring'e göre Ay yanılsaması vardır, çünkü Ay ufukta iken gözlemci Ay'a kendi göz seviyesinde bakar

---

[13] 19 Ekim 2010'da New York Times'ta Lock, Le Grange, Agras, Moye, Bryson ve Jo'nun (2000) yaptığı anoreksiya tedavisinde aile terapisi ile alternatif terapinin karşılaştırıldığı çalışmaya yer verilmiştir. Yazarların elde ettiği sonuçlara göre Everitt'in çalışmasındaki gibi aile terapisi kilo kaybını azaltmada daha etkili bulunmuştur.

**Tablo 7.4** Tepedeki Ay gözler yukarı kaldırıldığında ve göz seviyesinde olduğunda elde edilen Ay yanılsaması değerleri

| Gözlemci | Gözler yukarı kaldırıldığında | Göz seviyesinde | Fark (F) |
|---|---|---|---|
| **1** | 1,65 | 1,73 | –0,08 |
| **2** | 1,00 | 1,06 | –0,06 |
| **3** | 2,03 | 2,03 | 0,00 |
| **4** | 1,25 | 1,40 | –0,15 |
| **5** | 1,05 | 0,95 | 0,10 |
| **6** | 1,02 | 1,13 | –0,11 |
| **7** | 1,67 | 1,41 | 0,26 |
| **8** | 1,86 | 1,73 | 0,13 |
| **9** | 1,56 | 1,63 | –0,07 |
| **10** | 1,73 | 1,56 | 0,17 |
| | | | $\overline{F} = 0{,}019$ |
| | | | $s_F = 0{,}137$ |
| | | | $s_{\overline{F}} = 0{,}043$ |

ancak tepede iken gözlemci Ay'a bakmak için kafasını -ve doğal olarak gözlerini- yukarı kaldırmalıdır. Holway ve Boring Ay'a bakarken gözlemcinin göz seviyesinin farklı olmasının Ay yanılsamasına neden olduğunu savunmuşlardır. Ancak Kaufman ve Rock bu görüşe katılmamışlardır. Holway ve Boring'e ait bu Ay yanılsaması hipotezini test etmek için iki yapay Ay'dan (biri ufukta, diğeri tepede) oluşan ve deneklerin tepedeki Ay'a bakmak için gözlerini kaldırıp kaldırmadığını ölçen bir düzenek hazırlamışlardır. Hazırlanan düzeneklerin birinde deneklerin tepedeki Ay'a bakması için başlarını kaldırmaları gerekmezken (örneğin, yatarak görebilecekleri bir düzenek) diğerinde gözlerini yukarı kaldırmak durumundadırlar. (Ufuktaki Ay her zaman göz seviyesinde görülecek şekilde tutulmuştur.) Her iki durum için de bağımlı değişken olarak ufuktaki ve tepedeki Ay'ın gözlenen çaplarının oranı alınmıştır (oranın 1,00 olması yanılsama olmadığını gösterecektir.) Eğer Holway ve Boring hipotezlerinde haklılarsa, ufuktaki Ay sabit tutularak yapılan bu deneyde gözlerin yukarı kaldırılması gereken düzenekte göz seviyesinde gözlem yapılabilen düzeneğe göre daha fazla yanılsama görülmesi gerekir. Deneye ait veri kümesi Tablo 7.4'te verilmiştir.

Bu örnekte kontrol etmek istediğimiz yokluk hipotezi, göz seviyesinin değiştirildiği düzeneklerde elde edilen ortalamaların eşit olduğu şeklindedir. İlişkili gözlemlerle ilgilendiğimiz için fark puanları ile çalışacağız; yokluk hipotezi $H_0 : \mu_F = 0$ olacak; ,05 manidarlık düzeyinde iki yönlü test kullanılacaktır ve alternatif hipotez $H_1 : \mu_F \neq 0$'dır.

Elde edilen değerler ilişkili örneklem *t* testi denkleminde yerine konursa aşağıdaki sonuca ulaşırız:

$$t = \frac{\overline{F} - 0}{s_{\overline{F}}} = \frac{\overline{F} - 0}{\frac{s_F}{\sqrt{n}}}$$

$$= \frac{0{,}019 - 0}{\frac{0{,}137}{\sqrt{10}}} = \frac{0{,}019}{0{,}043}$$

$$= 0{,}44$$

Ek *t*'den $t_{,025}$ (9) = 2,262 buluruz. Hesaplanan *t* değeri (0,44) kritik değerden (2,262) küçük olduğu için yokluk hipotezi kabul edilecek ve Ay yanılsamasının Ay'a bakarken olabilecek göz seviyesi farkından etkilenmediği sonucuna varılacaktırır.[14] (Bu sonuç ayrıca Holway ve Boring'in hipotezinin diğer testini de içerir. Çünkü Holway ve Boring'e göre tepedeki Ay'a göz seviyesinde bakılırsa yanılsama olmaz. Elde dilen sonuçlar, bu hipotezin aksine, bu şartlar altında bir yanılsama olduğunu göstermiştir. Göz seviyesinde de yanılsama olduğunun manidarlık testi bir sonraki bölümde açıklanacak yöntemler ile kestirilebilir.)

## Etki Büyüklüğü

Bölüm 6'da yapılan deneyde görülen etkinin büyüklüğünü anlamak için etki büyüklüğü ölçümlerine değindik. İlişkili ölçümlerin farkına baktığımızda da, etki büyüklüğü hesaplayabiliriz ve hesaplanmalıdır. Uygulamada birçok dergi farka ait *t* değerleri ve güven aralığının yanında etki büyüklüğünün de verilmesi konusunda ısrarcılardır. Eşleştirilmiş örneklemlerde etki büyüklüğü kısa zaman içinde göreceğimiz gibi biraz karışıktır.

## Ölçümlerin *d*-Ailesi

Etki büyüklüğü için birçok farklı ölçüm söz konusudur, ancak bu başlık altında Kline'ın (2004) bahsettiği etki büyüklüğü katsayılarını tavsiye ederim. Bölüm 6'da yaptığım gibi gruplar arası farka dayalı (*d*-ailesi) ölçümler ile değişkenler arası ilişkilere dayalı (*r*-ailesi) ölçümler arasındaki farka değineceğim. Ancak bu bölümde kısmen daha az bilgilendirici olduğu için, kısmen de Bölüm 11'deki varyans analizi konusu içinde bahsetmek daha kolay ve daha mantıklı olacağı için, *r*-ailesinin üzerinde çok durmayacağım. *d*-ailesi ve *r*-ailesinin karşılaştırıldığı ilginç bir çalışma için McGrath ve Meyer'ın (2006) makalesi incelenebilir.

### Cohen'in *d*'si

Ölçümlerin isimlendirilmesinde bir karmaşa söz konusudur ve bu karışıklığın giderilmesi için Kline'ın (2004) yaptığı çalışma önerilmiştir. Bu noktada en çok kullanılan yaklaşım olan ve Kline'ın teknik olarak yanlış ifade ettiği **Cohen'ın d'si** üzerinde duracağım. Hedges ve Glass tarafından ortaya atılan ölçümler birbirine çok benzerdir ve birçok durumda birbirinin yerine kullanılabilir. Tüm bu ölçümler, ortalamalar arasındaki farkın standart sapmaya bölünmesiyle elde edilir.

Anoreksiya tedavisi ile ilgili veri, farkı anlatabilmek için iyi bir örnektir, çünkü herkesin beş ya da 10 kg alma ya da verme hakkında bilgisi vardır. Daha önce elde edilen sonuçlara göre Everitt'in verisinde fark manidardır ($t = 4{,}18$, $p < ,05$) ve kızlar ortalama 7,26 pound almışlardır. Bu durum ortalama ağırlıkları 83,23 pound olan kızlar için önemlidir. Ancak alınan kiloyu yüzde cinsinden vermek ve terapi sonrasında kızların kilosunda %9'luk (7,26/83,23) artış olduğunu söylemek daha anlaşılır olacaktır.

Bir alternatif yöntem de kazanılan kiloyu standart sapma cinsinden vermektir. Bu fikir Cohen'a aittir ve elde ettiği istatistiksel değere *d* adını vermiştir bu istatistik aşağıdaki gibi yazılır:

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

---

[14] Jones ve Tukey'nin (2000) yaptıkları çalışmaya göre, iki bakış farkı arasında küçük bir fark olabileceği ortaya atılmıştır; ancak örnek olarak ele alınan çalışmada, bu sonucu destekleyecek yeterince kanıt yoktur.

Bu eşitlikte pay iki evren ortalamasının farkı, payda ise evrene ait standart sapmadır. Denklem var olan örneğe göre yeniden düzenlendiğinde pay kazanılan ortalama kilo ($\mu_{\text{Sonraki}} - \mu_{\text{Önceki}}$) payda ise tedavi öncesi kiloların evrenine ait standart sapması olacaktır. (*t* değeri ve güven aralıkları hesaplanırken kullanılan fark puanlarının standart sapması yerine tedavi öncesi ağırlıkların evrenine ait standart sapmanın kullanıldığına dikkat ediniz.) İstatistiksel hesaplamalarda parametrelerden ziyade hesaplanan değerlerle işlem yapmak için evren puanları yerine örneklem ortalamasını ve standart sapmasını kullanırız. Bu bizi aşağıdaki ifadeye götürür:

$$\widehat{d} = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1}} = \frac{90,49 - 83,23}{5,02} = \frac{7,26}{5,02} = 1,45$$

Elde edilen değerin, *d*'nin kestirimi olduğunu belirtmek için *d*'nin üstüne şapka (^) koydum ve paydaya terapi öncesi puanların standart sapmasını yazdım. Elde edilen kestirim bize, aile terapisine katılan kızların yaklaşık olarak terapiden önceki kilolarının bir buçuk standart sapması kadar kilo kazandığı yorumunu söyler.

Ele alınan örnekte *d* yerine ortalama kazanılan ağırlık üzerinden konuşmak daha anlaşılırdır; çünkü ağırlık somut bir değişkendir. Ancak yapılan bu deney ağırlık yerine kızların özsaygıları üzerine yapılmış olsaydı terapinin sonunda "7,26 öz saygı puanı kazandılar" şeklinde elde edilen bir sonuç anlaşılır olmayacaktı. Böyle bir durumda "yaklaşık olarak bir buçuk standart sapma değerinde öz saygı puanı kazanıldı" demek sonucu daha anlaşılır kılacaktır.

Ancak etki değeri konusu bu kadar basit değildir; çünkü problemin sonucuna ulaşmak için birçok alternatif yol vardır. Bunlardan biri, terapi öncesi puanların standart sapmasının değil, terapi öncesi ve sonrası puanların standart sapmalarının ortalamasının kullanılmasıdır. Ancak kazanç ölçülüyorsa orijinal ağırlıklar üzerinden hesap yapmak kulağa daha mantıklı geliyor, çünkü araştırılan konu kızların başladığı noktaya göre nerede bitirdikleridir. Benzer şekilde standart sapmanın ortalamasının kullanılmasının daha mantıklı olduğu durumlar da söz konusudur.

*d* hesaplanırken paydada fark puanlarının standart sapmasının kullanılması da ayrıca olasıdır, ancak bu durumda tamamen başka bir değer ölçülmüş olur. Kline (2004), bu yaklaşımı tartışmış ve "ölçüm puanları düşünüldüğünde kaynak orijinal standart sapma ise standartlaştırılmış ortalamadaki değişimi (bu değere ait standart sapmayı) kullanmak daha mantıklıdır" sonucuna varmıştır. Ancak burada akılda tutulması gereken önemli nokta, bu görüş daha çok değişmeyen fikirlere sahip olan araştırma alanlarına dayanmıştır ve tek başına daima en iyi olan herhangi bir ölçüm yoktur. İstatistikte tamamen sabit olan birçok konu vardır. Örneğin, tek örneklem için güven aralığı tek bir yolla hesaplanır. Ancak araştırmacının sonuca ulaşmak için birçok yolunun olduğu durumlar söz konusudur ve araştırmacının hangi yolu seçeceği, yapılan deney için hangi yolun daha mantıklı olduğuna bağlıdır. Hangi yol seçilirse seçilsin, hangi standart sapmayı kullandığınızı okuyucuya belirtmek önemlidir.

## Eşleştirilmiş Örneklemler Hakkında Daha Fazlası

Eşleştirilmiş örneklem *t* testinin kullanıldığı birçok durumda aynı kişilerden elde edilen iki veri kümesini kullanabiliriz. Örneğin, 20 kişiye kan bağışında bulunduktan önce ve sonraki kaygı düzeylerini sorabiliriz. Veya 20 engelli bireyin engel düzeylerini belirleyen iki farklı puanlama sistemi kullanarak elde ettiğimiz puanları karşılaştırabiliriz. Her iki örnekte de 20 sayı kümesi ve her birey için iki veri elde edilecektir ve bu iki küme birbiri ile ilişkilidir. Kan bağışı örneği ele alınacak olursa, bu deneyde kişilerin kaygı seviyeleri farklı olacaktır. Kimileri ne olduğunun önemi olmadan sürekli bir kaygı içerisindeyken bazı kişiler hiçbir şey için kaygılanmaz. Bu nedenle kişilerin kaygı düzeyleri arasında kan bağışı yapmadan

ve yaptıktan sonra bir ilişki olması gereklidir. Bir diğer ifadeyle kişinin bağıştan önceki kaygı düzeyini bilirsek, bağıştan sonraki kaygısının hangi düzeyde olacağı ile ilgili anlamlı bir tahmin yürütebiliriz. Benzer şekilde engelli bireylerin bazılarına ait engel durumu büyükken, bazı kişiler orta seviyede engellidir. Eğer engelli bireyin puanlama sistemlerinin birinden yüksek puan aldığı bilinirse, diğer sistem kullanıldığında da yüksek bir puan alması beklenir. İki veri kümesi arasındaki ilişki muhtemelen hiçbir zaman mükemmel olmayacaktır. Ancak bu durum veri kümelerinin eşleştirilmiş mi, yoksa ilişkili mi olduğu konusunda tahmin yürütmeyi kolaylaştırır.

Yukarıda sözü edilen iki örnekte de, deneyde yer alan her bireyin iki farklı puanının olduğu durumları seçtim. İlişkili örneklemler için en genel yol budur ancak tek yol değildir. Örneğin, evlilik doyumu ile ilgili yapılan bir çalışmada eşlere, evliliklerinde ne kadar doyuma ulaştıkları sorulabilir. Böylelikle kadınların erkeklerden daha az veya daha fazla doyuma ulaşıp ulaşmadığı test edilir. (Bu bölümün alıştırmalar kısmında benzer bir örnek verilecektir.) Böyle bir örnekte her bireyin tek bir puanı olacaktır ancak her çift bir birim olarak ele alındığında puan çifti söz konusu olacaktır. Eşlerden biri evlilikte mutsuzsa diğerinin de mutsuz olacağını ya da tam tersini varsaymak mümkündür ve bu nedenle eşlerin puanları ilişkilidir.

Birçok deneysel desen ilişkili örneklem içerir. Yapılan çalışmalarda genel bir yapı vardır ve puan çiftlerinden birini bilmek çok fazla olmasa bile diğeri hakkında bilgi sahibi olunmasını sağlar. Böyle bir durumda örneklemlerin eşleştirilmiş olduğu söylenebilir.

## Kayıp Veri

İdeal olarak eşleştirilmiş örneklemlerde her değişkenin her bir durumu veya durum çifti için bir puanına sahibiz. Eğer birey ön testte yer almış ise son testte de yer alır. Eğer çiftlerden biri yapılan deneye katılıp veri sağlamışsa diğeri de katılır. Veri toplama işi bittiğinde eşleştirilmiş puanlardan oluşan bir veri kümesi oluştururuz. Ancak yapılan deneyler bazı durumlarda istediğimiz düzeyde gitmeyebilir.

Örneğin, çocuklarının gösterdiği problem davranışlar için kontrol listesinin anneler ve babalar tarafından ayrı ayrı doldurulduğunu varsayalım. Yapılan çalışma için beklentinin, annelerin çocuklarının problemli davranışları hakkında babalardan daha duyarlı davrandıklarını düşündüğümüzde, annelerin puanlarının babalara göre daha yüksek çıkması yönündedir. Genel olarak ebeveynlerin her ikisi de formu doldurur. Ancak annelerin doldurup babaların doldurmadığı 10 ve babaların doldurduğu annelerin doldurmadığı 5 form olduğunu varsayalım. Böyle bir durumda normal işleyiş içerisinde genellikle herkesin yapacağı en uygun şey, her iki ebeveyn tarafından doldurulmamış 15 formu çıkarmak ve kalanlara eşleştirilmiş örneklem *t* testini uygulamaktır. Ancak eşleştirilmiş örneklem *t* testinde kayıp verinin tesadüfî olduğu ve sistematik olmadığını varsayarak tüm verinin kullanılabileceği alternatif bir yol vardır. (Burada annesi tarafından çok az problemli davranış gösterdiği belirtilen ancak babasından veri toplanamamış çocuğa ait ya da babası tarafından çok problemli davranışı olduğu belirtilen ancak annesinden veri toplanamamış çocuğa ait verinin olmadığı varsayılmalıdır.)

Bhoj (1978) önerdiği yaratıcı testte, önce her iki puanı (anne ve babaya ait) ele alarak eşleştirilmiş *t* testi uygulamıştır; daha sonra sadece annelerin ve sadece babaların cevap verdiği veriye bağımsız örneklem *t* testi (konuya daha sonra değinilecektir) uygulamıştır ve en sonunda bu iki istatistiği birleştirmiştir. Birleştirilerek elde edilmiş *t* değeri özel tablolara göre değerlendirilmelidir. Bu tablolar Wilcox'ın 1986 yılında yaptığı çalışmada mevcuttur ve birleştirilmiş istatistiğe ait kritik değerler Wilcox'ın (1987a) yaptığı çalışmada tartışılmıştır. Eğer iki ya da üç kayıp veri söz konusu ise böyle bir testi kullanmak uygun değildir, ancak kayıp veri sayısı fazla ise birleştirilmiş istatistik kullanmak uygundur. Daha ayrıntılı bilgi için Wilcox'a (1987b) ait çalışma incelenebilir.

### Eşleştirilmiş Örneklemler *t* Testi İçin Kullanılan Bilgisayar Programları

Eşleştirilmiş örneklemleri analiz etmek için kullanılan neredeyse tüm bilgisayar programlarında, öncelikle karşılaştırılmak istenilen değişkenler arasındaki farktan oluşan yeni bir değişken elde edilir. Daha sonra bu farkın sıfır olan evren ortalamasından farklı olup olmadığını test etmek için tek örneklem *t* testi uygulanır. Alternatif olarak SPSS gibi bazı bilgisayar programları ilişkili örneklemler üzerinden *t* değeri hesaplamaya olanak sağlar. Daha önce yapılan uygulamaya çok benzer olduğu için burada tekrar etmeyeceğim.

### Bağımlı *t* Testi Sonuçlarını Raporlama

Everitt'in anoreksiya için aile terapisi çalışmasının sonuçlarını yazmak istediğimizi düşünelim. Yazılan raporda istatistiksel manidarlık testlerinin sonuçlarının yanı sıra ilgili örneklem istatistiklerine (ortalama, standart sapma, örneklem sayısı) de yer vermeliyiz. Rapora ayrıca terapi sonucunda kazanılan ortalama kiloya ait güven sınırları ve kestirdiğimiz etki büyüklüğü değerimizi (*d*) de eklemek isteriz. Bu bilgiye dayanarak sonuçları şu şekilde raporlayabiliriz:

> Everitt anoreksiya sorunu yaşayan kızların kilo almalarına ait aile terapisinin etkisini araştırmıştır. Yapılan çalışmada öncelikle 17 kızın terapiden önce kiloları ölçülmüştür, daha sonra kendilerine ve ailelerine aile terapisi uygulanmış ve en sonunda kızların terapi sonrası kiloları elde edilmiştir.
>
> Terapi sonunda $N = 17$ kızın aldıkları ortalama 7,26 ve standart sapma 7,16 pounddur. Alınan kilolara ait iki yönlü *t* testi istatistiksel olarak manidardır ($t(16) = 4{,}18$, $p < {,}05$). Bu sonuca göre terapi sonunda kızlarda kilo artışı gözlenmiştir. Ortalama kilo alış %95 güvenle 3,57 ile 10,95 arasındadır. Böylece aralığın ucunda olan bireylerde bile kilo alışı gözlendiği söylenebilir. 1,45 olarak bulunan Cohen's *d* katsayısına göre, kızların terapi öncesi kilolarına ek olarak, yaklaşık 1,5 standart sapma değerinde kilo aldıkları söylenebilir. Yapılan çalışmada aile terapisinin anoreksiya tedavisinde önemli bir etkisi olduğu sonucuna varılmıştır.

## 7.5 Ortalamalara Uygulanan Hipotez Testleri - İki Bağımsız Örneklem

En sık kullanılan *t* testlerinden biri, iki bağımsız grubun ortalamaları arasındaki farkın karşılaştırılmasıdır. Normal şartlar altında ve karanlıktaki farelerin basit bir sisteme ulaşmak için yaptıkları girişim sayılarının ortalamaları arasındaki farkı karşılaştırmak isteyebiliriz. Veya bir grup lise öğrencisine bir aktif durum cümlesini, diğer bir grup lise öğrencisine de olumsuz bir pasif cümleyi tekrarlatarak öğrencilerin cümleleri hatırlama ortalamalarını karşılaştırmak isteyebiliriz. Veya bireyleri yardıma ihtiyacı olan başka bir bireyin olduğu bir ortama sokarak, bireylerin yalnız olduğunda veya grup içinde bulunduğunda yardım etme eğilimlerini karşılaştırmak isteyebiliriz. Verilen tüm örneklerde gruplarda yer alan deneklerin her zaman mümkün olmasa da (örneğin, cinsiyete göre karşılaştırmalarda), gruplara tesadüfî olarak atandığını varsayarız.

İki bağımsız grupla yürütülen deneylerde örneklem ortalamalarının belli miktarlarda farklılaştığı gözlenir. Bu noktada önemli olan bu farkın iki örneklemin farklı evrenlerden geldiği sonucuna varacak kadar büyük olup olmadığı sorusudur. Jones ve Tukey'ye (2000) göre, sorulması gereken soru şudur: "İki örneklem ortalaması arasındaki fark evren ortalamasındaki farkı tanımlayabilecek kadar manidar büyüklükte midir?" Duruma uygun bir örneğe geçmeden önce örneklemler arasındaki farkların dağılımı ve buradan çıkacak *t* testi incelenmelidir.

## Ortalamalar Arası Farkın Dağılımı

**ortalamalar arasındaki farkların dağılımı**

Birinci evrenin ortalaması ($\mu_1$) ile ikinci evrenin ortalaması ($\mu_2$) arasındaki farkı test etme ile ilgilendiğimize göre yokluk hipotezini $H_0 = \mu_1 - \mu_2 = 0$ veya $\mu_1 = \mu_2$ şeklinde kurabiliriz. Yazılan bu yokluk hipotezinin test edilmesi işi, bağımsız örneklemlerin ortalamaları arasındaki farkı içerdiği için öncelikle **ortalamalar arasındaki farkların dağılımı**nın incelenmesi önemlidir. $X_1$ ve $X_2$ ile ifade edilen iki evrenimizin, bunların ortalamalarının $\mu_1$ ve $\mu_2$, varyanslarının $\sigma^2_1$ ve $\sigma^2_2$ olduğunu varsayalım. $X_1$ evreninden $n_1$ ve $X_2$ evreninden $n_2$ büyüklüğünde birer örneklem çekelim, bu örneklemlere ait ortalamaları ve bu iki ortalama arasındaki farkı hesaplayalım. Örnekleme her iki evrende de bağımsız olduğu için örneklem ortalamaları da bağımsız olacaktır. Bu sürecin sonsuz sayıda tekrarı ile elde edilecek sonuç Şekil 7.8'de gösterilmiştir. Şeklin alt kısmında yer alan ilk iki sütun $X_1$ ve $X_2$'ye ait örnekleme dağılımını, üçüncü sütun ise ortalamalar arası farkın örnekleme dağılımını $(\bar{X}_1 - \bar{X}_2)$ gösterir. En çok üçüncü sütun üzerinde duracağız, çünkü ilgilendiğimiz esas konu ortalamalar arası farkın test edilmesidir. Bu dağılımın ortalaması $\mu_1 - \mu_2$ şeklinde gösterilebilir. Genel olarak **varyansların toplamı yasası** olarak adlandırılan farkın dağılımına ait varyans kısaca şöyle tanımlanır:

**varyansların toplamı yasası**

> İki bağımsız değişkenin toplamının veya farkının varyansı bu değişkenlerin varyanslarının toplamına eşittir.[15]

Merkezî limit teoreminden $\bar{X}_1$'nın dağılımının varyansının $\sigma^2_1/n_1$ ve $\bar{X}_2$'nin dağılımının varyansının da $\sigma^2_2/n_2$ olduğunu biliyoruz. Değişkenler (örneklem ortalamaları) bağımsız olduğu için bu iki değişkene ait farkın varyansı, varyansları toplamına eşittir.

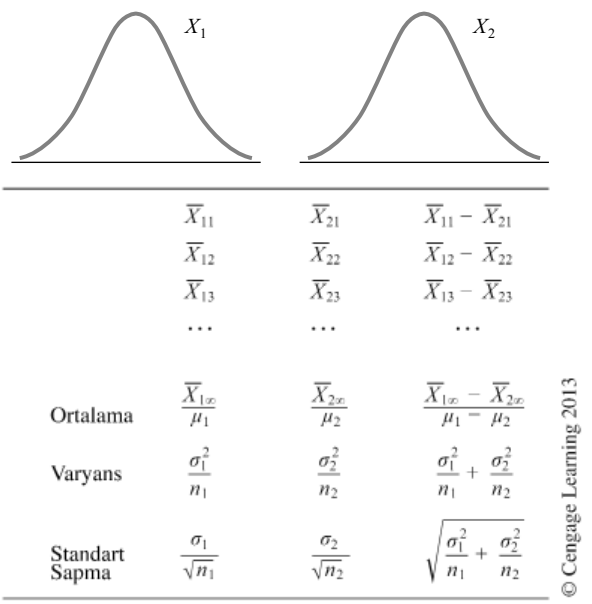
$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$$

Ortalamalar arasındaki farkın ortalamasının ve varyansının bulunmasıyla ihtiyaç olan bilinmeyenlerin çoğu elde edilmiş olur. Ortalamalar arası farkın örneklem dağılımına ait genel biçimi Şekil 7.9'daki gibidir.

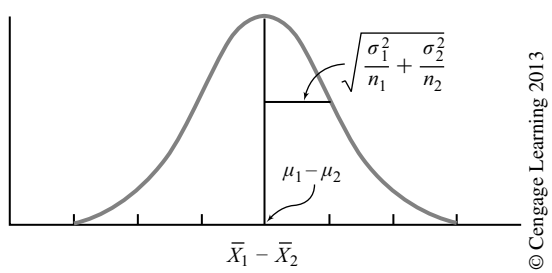
Bu dağılıma ait değinilecek son nokta dağılımın şeklidir. İstatistikteki önemli teoremlerden birine göre, toplamı ya da farkı normal dağılım gösteren iki bağımsız değişkenin kendileri de tek başlarına normal dağılım gösterir. Şekil 7.9'da ortalamanın örneklem dağılımı arasındaki farkın dağılımı gösterildiği ve ortalamanın örneklem dağılımı normal olduğu için Şekil 7.9'daki dağılım kendi başına en az yaklaşık olarak normal olmak zorundadır.

---

[15] Yasanın tamamı ele alındığında, değişkenlerin bağımsız olma durumu dışarıda bırakılır ve iki değişkenin toplamlarının ya da farklarının varyansı $\sigma^2_{X_1 \pm X_2} = \sigma^2_1 + \sigma^2_2 \pm 2\rho\sigma_1\sigma_2$ ile gösterilir. Denklemde yer alan "±" işareti için, değişkenlerin toplamı alınıyorsa "+", farkı alınıyorsa "–" olarak alınacağını ifade eder. $\rho$ ifadesi ise iki değişken arasındaki ilişkiyi verir ve değişkenler bağımsız olduğunda bu değer sıfırdır. ($\rho \neq 0$ ise değişkenler birbirinden bağımsız değildir, bu durum bizi de ilgili örnekleri ayrı ayrı incelemeye zorlar.)




**Şekil 7.8** İki evrenden örnekleme yapıldığında elde edilen ortalama ve ortalamalar arası farkın grafikle gösterimi




**Şekil 7.9** Ortalamalar arası farkın örneklem dağılımı

## *t* İstatistiği

**ortalamalar arası farkın standart hatası**

Yapılan açıklamalar ile ortalamalar arası farkın örneklem dağılımı hakkında bilgi sahibiyiz, bu nedenle artık ilgili test süreci üzerinde ilerleyebiliriz. Bir an için evren varyanslarının ($\sigma_i^2$) bilinmediğini varsayalım. Daha önce $z$ (dağılımdaki herhangi bir noktanın) istatistiğinin, dağılımın ortalaması arasındaki farkın dağılımın standart hatasına bölünmesi ile elde edileceğine değinmiştik. Şimdi yapılacak istatistiksel işlemde, örneklem ortalamaları arasındaki fark ($\overline{X}_1 - \overline{X}_2$), örneklem dağılımının ortalaması ($\mu_1 - \mu_2$) ve **ortalamalar arası farkın standart hatası**[16] kullanılır**.**

[16] Herhangi bir örneklem dağılımına ait standart sapma değerinin dağılımın standart hatası olarak isimlendirildiği unutulmamalıdır.

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Böylece

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}}$$
$$= \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

yazabiliriz.

Tek örneklem *t* testi açıklanırken belirtildiği gibi $\sigma$ = ,05 (iki kuyruklu) için kritik değer ± 1,96'dır.

Yukarıdaki formül *t* testinin kaynağını göstermek dışındaki amaçlar için kullanışlı değildir, çünkü evren varyansı nadir olarak bilinir. Bu problem tek örneklem *t* testinde yapıldığı gibi evren varyansı yerine örneklem varyansı kullanılarak çözülebilir. Tek örneklem *t* testinde olduğu gibi, böyle bir durumda da sonuçlar *z* yerine *t* cinsinden aşağıdaki biçimde yazılır:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1-\bar{X}_2}}$$
$$= \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Yokluk hipotezi "iki evren ortalaması arasındaki fark sıfırdır" şeklinde kurulduğu için denklemde yer alan $\mu_1 - \mu_2$ yerine "0" konur. Bu durumda eşitlik şu hale gelir:

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{s_{\bar{X}_1-\bar{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

## Varyansları Birleştirme

Yukarıda elde edilen *t* denklemi iki değişkene ait örneklem büyüklükleri eşit ise uygundur. Farklı örneklem büyüklükleri için denklemde bazı değişiklikler yapmak gerekir. Bu değişiklikler evren varyansının kestirilmesini kolaylaştıracak şekilde olmalıdır, çünkü evren varyansının kullanılması gereken durumlarda bazı karışıklıklar söz konusudur. Örneklem büyüklükleri farklı olan değişkenlerin varyanslarının birleştirilmesiyle ilgili yöntemlerden söz edilecektir, ancak bu bölümün sonunda, bu noktaya geri dönülecek ve bu yöntemin ne kadar iyi olduğu tartışılacaktır.

Bağımsız iki örneklemde *t* testinin kullanabilabilmesi için gerekli varsayımlardan biri örneklemlerin evren varyanslarının eşit olması $\sigma_1^2 = \sigma_2^2$ (yokluk hipotezinin doğru veya yanlış olduğuna bakılmaksızın örneklemlerin varyansları eşit olan örneklemlerden geldiği) varsayımıdır. Varsayım, örneklem büyüklüklerinin eşit olup olmamasına ($n_1$ ve $n_2$ eşit) bakılmaksızın gereklidir. Böyle bir varsayım çoğunlukla akla uygun gelir. Deneye genel olarak eşit

olan iki değişkenle başlarız ve daha sonra değişkenlerden birine veya ikisine birden sonuçları etkileyecek bir uygulama yaparız. Böyle bir durumda değişkenlerin varyanslarının durumu etkilemeyeceğini varsaymak daha doğrudur. (Toplam puana sabit bir değeri eklemek ya da çıkarmanın -deney için uygulama etkisinin- varyansa etkisi olmadığını unutmayınız.) İki değişkene ait evren varyanslarının eşit olduğu varsayıldığı için bu genel varyans $\sigma^2$ şeklinde gösterilebilir.

Elde edilen veri kümesi için genel varyansın ($\sigma^2$) $s_1^2$ ve $s_2^2$ şeklinde iki ayrı kestirimi vardır. İki ayrı kestirim yerine varyansa daha yakın bir değer elde etmek için $s_1^2$ ve $s_2^2$ 'nin ortalamasını almak daha uygun görünmektedir. Ancak kestirimlerin herhangi biri çok daha fazla gözlem ile elde edilmiş olabileceğinden varyans değerlerinin basit aritmetik ortalamasını almak doğru değildir. Böyle bir durumda elde edilmesi gereken ortalama, örnekleme varyanslarının serbestlik dereceleri ($n$-1) ile ağırlıklandırıldığı, **ağırlıklı ortalama**dır. Bu yeni kestirime $s_B^2$ denilirse eşitlik aşağıdaki gibi olacaktır:

**ağırlıklı ortalama**

$$s_B^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Yukarıdaki eşitliğin payında serbestlik dereceleri ile ağırlıklandırılmış varyanslar, paydasında ise serbestlik dereceleri toplamı yer almaktadır.

İki örneklem varyansının ağırlıklı ortalaması genel olarak **birleşik varyans kestirimi** olarak isimlendirilir. Açıklanan birleşik varyansı $t$ eşitliğinde yerine koyarak şöyle yazabiliriz.

**birleşik varyans kestirimi**

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{s_{\overline{X}_1 - \overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_B^2}{n_1} + \frac{s_B^2}{n_2}}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{s_B^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Yukarıda elde edilen ve daha önce elde eden $t$ ifadeleri incelendiğinde, her iki denklemde de $t$ değeri, örneklem ortalamaları arasındaki farkın, ortalamalar arasındaki farkın standart hatasına bölünmesiyle elde edilmiştir. Eşitlikler arasındaki tek fark standart hatanın kestirilmesi şeklidir. Her iki değişkene ait örneklem büyüklükleri eşit olduğunda varyansların birleşip birleşmediğinin önemi kalmaz ve her iki durumda da eşit sonuç elde edilir. Ancak örneklem büyüklükleri eşit değil ise birleşik varyans kestirimi farklılığa yol açar.

## Serbestlik Derecesi

$t$ değeri hesaplamalarında iki örneklem varyansı ($s_1^2$ ve $s_2^2$ ) söz konusudur. Bu varyansların her ikisi de ilgili örneklem ortalamalarına ait sapmaların karesidir ve bu nedenle her biri için serbestlik derecesi ($n$-1)'dir. Her iki örneklem düşünüldüğünde toplam serbestlik derecesinin değeri $(n_1-1) + (n_2-1) = (n_1 + n_2 - 2)$ olacaktır. Böylece bağımsız iki örneklem $t$ testi için serbestlik derecesi $(n_1 + n_2 - 2)$ olur.

## Homofobi ve Cinsel Uyarılma

Adams, Wright ve Lohr (1996) yaptıkları çalışmalar ile homofobinin bireylerin bilinçaltlarında yatan homoseksüel olma kaygıları ile ilişkili olabileceğine dayanan psikoanalitik hipotezleri incelemişlerdir. Araştırmacılar 64 heteroseksüel erkeğe homofobi ölçeği uygulamışlar ve aldıkları puana göre bireyleri homofobik ve homofobik olmayan diye iki sınıfa ayırmışlardır. Daha sonra homofobik ve homofobik olmayan heteroseksüel bireylere homoseksüel ve heteroseksüel cinsel içerikleri olan erotik videolar izletilmiş ve bireylerin cinsel uyarılma dereceleri kaydedilmiştir. Adams ve arkadaşları, "eğer homofobi bireylerin bilinçaltlarında yatan homoseksüel olma kaygıları ile ilgili ise homofobik bireylerin homofobik olmayan bireylere göre videolara daha yoğun uyarılma tepkisi vermesi gerektiğini" savunmuşlardır.

**Tablo 7.5** Adams ve arkadaşlarının yaptığı çalışmadan alınan homofobik ve homofobik olmayan bireylerin cinsel uyarılma derecesine ait veri kümesi

| **Homofobik** | | | | | | **Homofobik olmayan** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 39,1 | 38,0 | 14,9 | 20,7 | 19,5 | 32,2 | 24,0 | 17,0 | 35,8 | 18,0 | -1,7 | 11,1 |
| 11,0 | 20,7 | 26,4 | 35,7 | 26,4 | 28,8 | 10,1 | 16,1 | -0,7 | 14,1 | 25,9 | 23,0 |
| 33,4 | 13,7 | 46,1 | 13,7 | 23,0 | 20,7 | 20,0 | 14,1 | -1,7 | 19,0 | 20,0 | 30,9 |
| 19,5 | 11,4 | 24,1 | 17,2 | 38,0 | 10,3 | 30,9 | 22,0 | 6,2 | 27,9 | 14,1 | 33,8 |
| 35,7 | 41,5 | 18,4 | 36,8 | 54,1 | 11,4 | 26,3 | 5,2 | 13,1 | 19,0 | -15,5 | |
| 8,7 | 23,0 | 14,3 | 5,3 | 6,3 | | | | | | | |
| **Ortalama** | | | 24,0 | | | **Ortalama** | | | 16,50 | | |
| **Varyans** | | | 148,87 | | | **Varyans** | | | 139,16 | | |
| ***n*** | | | 35 | | | ***n*** | | | 29 | | |

© Cengage Learning 2013

Verilen örnek içinde sadece homoseksüel videolar ile ilgili olan veri incelenecektir. (Çalışmadan elde edilen veri kümesinde heteroseksüel ve lezbiyen görüntüler içeren videolarda grup ayrımı yoktur.) Tablo 7.5'te yer alan veri Adams ve arkadaşlarının yaptıkları çalışma ile aynı ortalama ve birleşik varyans değerine eşit olacak şekilde düzenlenmiştir.[17] Bağımlı değişken 4 dakikalık videonun sonunda bireylerin uyarılma derecesidir ve büyük değerler daha fazla uyarılma olduğunu gösterir.

Herhangi bir istatistiksel testi uygulamaya, hatta veri toplamaya başlamadan önce teste ait bazı durumları belirlemeliyiz. Bunlardan öncelikli olanı sıfır ve alternatif hipotezlerin kurulmasıdır:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Alternatif hipotez iki yönlüdür. (İki yönlü test kullanılır ve $\mu_1 < \mu_2$ veya $\mu_1 > \mu_2$ olduğu durumlarda yokluk hipotezi reddedilir.) Biz iki yönlü hipotez testini kullanacağız. Kitapta bulunan diğer örneklerle uyum sağlaması açısından $\alpha$ değerini ,05 alacağız. Ancak unutulmamalıdır ki verilen bu kararlar kesinlik taşımaz ve değişebilir. (Jones ve Tukey'nin (2000) bu paragrafı nasıl yazdığını düşünün. Burada yazılandan nerede farklı olurdu ve onların yaklaşımı niçin daha açık hale getirilebilirdi?)

Yukarıdaki yokluk hipotezine göre *t* değerini aşağıdaki gibi hesaplayabiliriz:

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{s_{\overline{X}_1 - \overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_B^2}{n_1} + \frac{s_B^2}{n_2}}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{s_B^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$H_0 = \mu_1 - \mu_2 = 0$ yokluk hipotezi test edildiği için $\mu_1 - \mu_2$ eşitlikten çıkarılmıştır. Elde edilen örneklem varyansları birleştirilebilir, çünkü varyanslar, homojenliği için kaygı duyulmayacak kadar birbirlerine yakındır. Varyanslar birleştirebilir ve birleşik varyansı aşağıdaki gibi hesaplayabiliriz:

$$s_B^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{34(148,87) + 28(139,16)}{35 + 29 - 2} = 144,48$$

[17] Negatif puanlara yer vermemek için tabloda yer alan değerlere gerçekte olan değerlerden farklı olarak 12 puan eklenmiştir. Ancak bu durumun sonuca bir etkisi yoktur.

Birleştirilmiş varyansın değeri, birinci örneklemin varyansına, ikinciye göre daha yakındır. Çünkü birinci örneklemdeki eleman sayısı (35) ikincininkinden (29) fazladır, bu nedenle daha fazla ağırlıklandırılmıştır.

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_B^2}{n_1} + \frac{s_B^2}{n_2}}} = \frac{(24{,}00 - 16{,}50)}{\sqrt{\frac{144{,}48}{35} + \frac{144{,}48}{29}}} = \frac{7{,}50}{\sqrt{9{,}11}} = 2{,}48$$

Verilen örnekte homofobik grup için $n_1 - 1 = 34$ *sd*, homofobik olmayan grup için $n_2 - 1 = 28$ sd olup toplam serbestlik derecesi $n_1 - 1 + n_2 - 1 = 62$'dir. Ek *t*'deki tablodan $t_{,025}(62) \cong 2{,}003$ (doğrusal interpolasyon ile) olarak buluruz. Hesaplanan *t* değeri kritik *t* değerinden daha büyük olduğu için yokluk hipotezini reddederiz ve gözlemlerin yapıldığı örneklem ortalamaları arasında manidar fark olduğu sonucuna ulaşırız. Bir diğer ifade ile $\mu_1 \neq \mu_2$ ve pratikte $\mu_1 > \mu_2$ olduğu için homofobik bireyler, homofobik olmayan bireylere göre homoseksüel içerikli videolara daha fazla uyarılma tepkisi göstermiştir. (Jones Tukey'ye (2000) göre burada yapılanı nasıl yorumlarsınız?

## $\mu_1 - \mu_2$ Farkı İçin Güven Aralığı

Evren ortalamaları arasındaki farkın yokluk ($H_0$: $\mu_1 - \mu_2 = 0$) hipotezini test etmenin ve kısaca elde edeceğimiz etki büyüklüğünün yanında $\mu_1$ ve $\mu_2$ için güven sınırlarını belirlemek de önemlidir. Güven sınırlarını belirleme mantığı tek örneklem *t* testinde yapılan ile tamamen aynıdır. Hesaplamalarda, ortalama ve ortalamanın standart hatası yerine ortalamalar arası *fark* ve *farkın* standart hatasını kullanırız ve bunun dışında bir farklılık söz konusu değildir. Evren ortalamaları arasındaki farkın %95 olasılıklı güven sınırlarını aşağıdaki şekilde ifade edebiliriz:

$$GA_{,95} = (\overline{X}_1 - \overline{X}_2) \pm t_{,025}\, s_{\overline{X}_1 - \overline{X}_2}$$

Homofobi çalışması için güven aralığı aşağıdaki gibi yazılabilir:

$$GA_{,95} = (\overline{X}_1 - \overline{X}_2) \pm t_{,025}\, s_{\overline{X}_1 - \overline{X}_2} = (24{,}00 - 16{,}5) \pm 2{,}00\sqrt{\frac{144{,}48}{35} + \frac{144{,}48}{29}}$$

$$= 7{,}50 \pm 2{,}00(3{,}018) = 7{,}5 \pm 6{,}04$$

$$1{,}46 \leq (\mu_1 - \mu_2) \leq 13{,}54$$

Geleneksel istatistikçilere göre homofobik ve homofobik olmayan katılımcılar arasındaki uyarılma farkı ,95 olasılıkla hesaplanan aralıktadır. Benzer şekilde Bayes istatistikçilerine göre de evren ortalamaları arasındaki gerçek fark ,95 olasılıkla bu aralık içindedir. Aralık geniş olmasına rağmen sıfır değerini içermemektedir. Bu durum yokluk hipotezinin reddedilmesiyle tutarlıdır ve homofobik bireylerin homoseksüel içerikleri olan videolardan homofobik olmayan bireylere göre daha fazla uyarıldıkları sonucunu destekler. Ancak güven aralığının bu kadar geniş olması göz ardı edilemez. Gruplar arası fark istatistiksel olarak manidar bulunmuştur, ancak hâlâ farkın ne kadar büyük olduğu ile ilgili belirsizlik vardır. Ayrıca unutulmamalıdır ki bağımlı değişken tesadüfen seçilmiş bir örneklemden elde edilen cinsel uyarılma derecesidir. Hesaplanan güven aralığı dar olsa bile sonucun gerçekte ne anlama geldiğini yorumlamak zordur. Gruplar arasındaki uyarılma farkının 7,5 olduğunu söylemek anlamlı bir bilgi değildir, çünkü bu farkın büyük bir fark mı yoksa küçük mü olduğu bilinmemektedir. Bu farkın ne anlama geldiğini öğrenmenin gerçek bir yolu yoktur, çünkü birimlerin değeri sezgisel yollarla anlaşılmaz. Ancak bir sonraki bölümde yapılacağı gibi, ölçümler standartlaştırıldığında daha çok bilgilendirici olabilir.

## Etki Büyüklüğü

Hesaplanan güven aralığı, her ne kadar istatistiksel olarak homofobik grubun homofobik olmayan gruba göre daha fazla uyarılma gösterdiği sonucunu verse de, iki grup arasındaki uyarılma farkı ile ilgili önemli ölçüde belirsizlik olduğunu göstermiştir. Gruplar arasındaki farkın büyüklüğü hakkında bilgi sahibi olmak için daha önce verdiğimiz bilgiye geri döneceğiz. Bu noktada Cohen'ın *d* etki büyüklüğü ölçüsünü kullanacağız ve eşleştirilmiş örneklemlerde olduğu gibi ortalamalar arasındaki farkı standart sapmaya böleceğiz. Yapılan işlem sonucu elde edilen katsayıyı *d* olarak adlandıracağız. Ancak burada kullanılacak olan standart sapma, her iki evrenden kestirilen standart sapmadır. Veri kümesinin elde edildiği evrene ait standart sapmanın en iyi kestirimi için önce örneklemlere ait varyanslar birleştirilir ve daha sonra elde edilen birleşik varyansın karekökü alınır.[18] (Birbirinden çok farklı varyanslar söz konusu ise, hangi örnekleme ait olduğu belirtilmek şartıyla örneklemlerden birine ait standart sapma değeri kullanılabilir.)

Homofobi çalışmasına ait veri dikkate alındığında;

$$\widehat{d} = \frac{X_1 - X_2}{s_B} = \frac{24,00 - 16,50}{12,02} = 0,62$$

Elde edilen bu değer, iki grup arasındaki farkı standart sapma cinsinden belirtir ve homofobik katılımcıların ortalama uyarılma düzeyleri, homofobik olmayanlara göre yaklaşık 2/3 standart sapma değeri kadar daha büyüktür. Bu oldukça büyük bir farktır.

## Etki Büyüklüğünün Güven Sınırları

Ortalamalar ve ortalamalar arasındaki farkın güven sınırlarını belirlediğimiz gibi, etki büyüklüğünün de güven sınırlarını belirleyebiliriz. Ancak yapılması gereken işlemler daha zordur ve merkezî olmayan *t* dağılımı ile ilişkilidir. (Bölüm 10, Kesim 10.1'de bu konuya değinilmiştir.) Güven sınırlarının belirlenebileceği yazılımlar vardır (Cumming ve Finch, 2001) ve http://www.uvm.edu/~dhowell/methods8/Supplements/Confidence eklentisinde etki büyüklüğünün güven aralığı ile ilgili dokümanlar yer almaktadır. Dokümanlarda, kitapta altı çizilen yöntemlerin dışında sınırları belirlemenin önemine de vurgu yapan çalışmalar bulunmaktadır. Cumming ve Finch'in (2001) yazdığı yazılım kullanılarak homofobi çalışması için elde edilen etki büyüklüğü (*d*) güven aralığı değerleri oldukça geniş olan 0,155 ile 1,125'tir. Aynı zamanda güven aralığı için alt sınır olan değer bile manidar bir şekilde büyüktür. R ve MBESS programlarını kullanarak güven aralıklarını hesaplamak için kitabın Web sitesinden KellyCI.R isimli dosya kullanılabilir.

Dikkat edilmesi gereken önemli noktalar: Homofobi örneğinde ölçümlere ait birimler geniş ölçüde keyfîdir ve 7,5 birimlik uyarılma farkının kendine özgü bir anlamı yoktur. Bu durumda fark değerini araştırmacıya daha fazla anlam ifade etmesi açısından standart sapma cinsinden belirtmek daha doğrudur. Ancak orijinal birimlerin anlamlı olduğu birçok durum söz konusudur ve böylesi durumlarda ölçümü standartlaştırmak gereksizdir. Standartlaştırılmış birimler yerine ortalamalar arasındaki farkı veya ortalamaların oranını tercih edebiliriz. Ay yanılsaması çalışması böyle bir durum için örnek verilebilir. Ufuktaki Ay'ın tepedeki Ay'dan bir buçuk kat daha büyük gözükmesi anlamlandırılabilir ve aradaki farkın standartlaştırılmış birimler ile ifade edilmesinin hiçbir avantajı yoktur. İki grubun ortalamaları arasındaki farkı ifade ederken önemli olan araştırmacı ya da okuyucu için far-

[18] Evren değeri için birleştirilmiş standart sapma kavramı ilk olarak Hedges (1982) tarafından ortaya atılmıştır. Bu nedenle bu kavram Hedge'in g'si olarak anılır.

kın değerinin bir anlam ifade etmesidir. Bu nedenle gruplar arasındaki farkın ifade edildiği bazı durumlar için *d* gibi standartlaştırılmış ölçümler daha uygun iken bazı durumlar için ortalamalar arası fark gibi ölçümler daha uygundur.

Dikkat edilmesi gereken ikinci nokta, literatürde etki büyüklüğü hakkında yer alan bilgidir. Elde edilen etki büyüklüğü değeri sadece yayımlanmış çalışmalara (Sterling, 1959; Lane ve Dunlap, 1978; Brand, Bradley, Best ve Stoica, 2008) dayandırılarak yorumlandığında, etki büyüklüğü değerinin kestiriminin olduğundan yüksek yorumlanması ihtimali vardır. Çünkü çalışmaların yayımlanması için sonuçların manidar olması gerektiğine yönelik kesin bir eğilim vardır ve bu nedenle etki değeri manidar olmayan çalışmalar için örnekler yetersiz kalmıştır. (Rosenthal (1979) bu durum için, manidar sonuçların olmadığı ve basılmayan çalışmaların yer aldığı "dosya çekmecesi problemi" tanımlaması yapmıştır.) Lane ve Dunlap (1978) ,25 düzeyinde gerçek etki büyüklüğünü kullanarak ve ortalamalar arası farkın 4 puan olduğu (standart sapma = 16) basit bir örnekleme çalışması yapmışlardır. Her iki değişken için de, örneklem büyüklüğünün $n_1 = n_2 = 15$ olduğu çalışmada ortalamalar arası farkı 13,21 bulmuşlardır. Sadece sonuçlara bakıldığında elde edilen bu fark ,05 düzeyinde manidardır. Bunlara ek olarak çalışmada, örneklem standart sapmaları dikkat çekecek kadar düşük kestirilmiştir. Bu, güven aralıklarının daha dar çıkmasına neden olmuştur. Sadece yayımlanmış çalışmalar ile araştırma yapılıyorsa bu bilgiler akılda bulundurulmalıdır.

Son olarak hipotezleri test ederken ağırlıklandırılmış ortalamaların ve birleştirilmiş varyansların kullanılmasının etki büyüklüğü konusunu daha önemli hale getirdiğini not etmeliyim. Algina, Keselman ve Penfield (2005) tarafından belirtildiğine göre, Cohen'ın *d'si* gibi katsayılar bu tarz istatistiklerin kullanılması ile geliştirilebilir. Benzer durum güven sınırları için de geçerlidir.

Bir sonraki bölümde göreceğiniz gibi, Cohen etki büyüklüğünü "büyük, orta ve küçük" olarak değerlendirilebilecek temel sınırlar çizmiştir. Cohen'a göre *d* değeri ,20 olduğunda etkili ancak küçük, ,50 olduğunda orta düzeyde etki ve ,80 için büyük etkidir. Cohen bu sınırlar için kesinlik belirtmemiştir ancak yine de bu sınırlar rehber olması açısından kullanışlıdır.

## Sonuçları Raporlama

Bağımsız iki örneklem *t* testi için sonuçları raporlama, bağımlı örneklem *t* testinde yapılana çok benzerdir. Adams ve arkadaşlarının homofobi üzerine yaptıkları çalışmada biri homofobi ölçeğinde yüksek, diğeri düşük puan almış iki grup katılımcı vardır. Gruplara cinsel içerikli homoseksüel videolar izlettirildiğinde homofobik grup daha yüksek düzeyde cinsel uyarılma göstermiştir (ortalamalar arası fark 7,5 birim olarak bulunmuştur). Ortalamalar arası fark için yapılan *t* testine göre, sonuç istatistiksel olarak manidardır ($p < ,05$) ve ,62 bulunan etki büyüklüğüne göre iki grup birbirinden neredeyse 2/3 standart sapma değerinde farklılaşmaktadır. Ancak evren ortalamaları arasındaki farkın güven sınırları oldukça geniş bulunmuştur ($1,46 \leq (\mu_1 - \mu_2) \leq 13,54$) bu nedenle bulunan farkın büyüklüğü kesin bir yargıya ulaşmak için yetersiz kalmıştır.

## SPSS Analizi

Adams ve arkadaşlarının yaptıkları çalışmaya ait SPSS analizleri Tablo7.6'da verilmiştir. SPSS analizlerinde ilk olarak varyansların eşitliği için Levene testi sonuçları yer almaktadır. Bu testi kısaca açıklayacağız. Levene testi sadece varyansların homojenliği varsayımını test etmek için yapılır. Bu testi kısaca tartışacağız; fakat Levene testi, sadece varyansların homojenliği varsayımını yoklayan bir testtir. Varyanların eşit olduğu hipotezini ret etmeyeceğiz ($p = ,534$). Dolayısıyla varyansların eşit olduğu varsayımından endişe etmeyeceğiz, varyansların eşit olduğunu kabul edeceğiz. Şimdi tablonun en alt satırına odaklanacağız.

**Tablo 7.6** Adams ve arkadaşlarının (1966) yaptıkları çalışmanın SPSS analizleri

**Grup İstatistikleri**

| GRUP | | N | Ortalama | Standart sapma | Ortalamanın standart hatası |
|---|---|---|---|---|---|
| **Uyarılma** | Homofobik | 35 | 24,0000 | 12,2013 | 2,0624 |
| | Homofobik olmayan | 29 | 16,5034 | 11,7966 | 2,1906 |

Bağımsız Örneklemler t Testi

| | **Levene Testi** | | **Ortalamaların Eşitliği İçin t Testi** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Farkın %95 Güven Aralığı | |
| | F | Mnd. | t | Sd | Mnd. (iki yönlü) | Ortalama Fark | Farkın Standart Hatası | En Düşük | En Yüksek |
| **Varyanalar eşit ise** | ,391 | ,534 | 2,484 | 62 | ,016 | 7,4966 | 3,0183 | 1,4630 | 13,5301 |
| **Varyanslar eşit değil ise** | | | 2,492 | 60,495 | ,015 | 7,4966 | 3,0087 | 1,4794 | 13,5138 |

SPSS ile elde edilen *t* değeri bizim hesapladığımız ile aynıdır, ilgili *t* değeri (,016) ,05'ten küçüktür. Bu nedenle yokluk hipotezi reddedilmiştir. Ayrıca SPSS analizi ile ortalamalar arası fark ve farkın standart hatası değerleri hesaplanmıştır. Bunun yanında SPSS çıktısında el ile hesaplanan değerleri onaylayacak şekilde, %95 olasılıklı ortalamalar arası farkın güven aralığı yer almaktadır. Bu değerler burada bulunanla uyumludur.

## Çalışılmış İkinci Örnek

Joshua Aronson "klişe" (stereotype) adını verdiği geniş çaplı araştırmasında, klişe grup üyelerinin davranışlarının grupları hakkında yeteneksiz oldukları yönünde negatif fikir geliştirebileceğini düşündükleri durumlarda daha yoğun bir baskı hissettiklerini ortaya atmıştır (Aronson, Lustina, Good, Keough, Steele ve Brown, 1998). Çalışmada klişe korkularının bireyin bireysel zamanlarda gösterebileceği performansını etkileyeceği ve genellikle düşüreceği yönünde hipotez kurulmuştur. Benzer çalışmalar bazı alanlarda önyargılı bir şekilde yetersiz olarak ün yapmış etnik gruplar ile gerçekleştirilmiştir. Ancak Aronson ve arkadaşları çalışmayı bir adım daha öteye taşımışlar ve klişe korkuları olmayan beyaz ırktan erkeklerden oluşan bir grup da almışlar ve klişe baskısının performanslarını düşürüp düşürmediğini incelemişlerdir.

Aronson ve arkadaşları iki bağımsız grupla çalışmışlardır. Gruplarda matematik becerileri yüksek olan lise öğrencilerine yer vermişlerdir. Kontrol grubu olarak 11 öğrenci seçilmiş ve zor bir matematik sınavını tamamlamaları istenmiştir. Deney grubunda ise 12 öğrenci yer almıştır ve öğrencilere Asyalı öğrencilerin diğer öğrencilere göre matematikte daha yetenekli olduğu ve yapılan testin amacının bu farkın nedenini araştırmak olduğu söylenmiştir. Aronson'a göre beyaz öğrencilere Asyalıların matematikte daha başarılı olduğunu söylemek öğrenciler üzerinde klişe baskısı oluşturacak ve öğrencilerin başarısını düşürecektir.

Tablo 7.7'deki veri, Aronson ve arkadaşlarının çalışmasında yer alan veri ile benzer ortalama ve standart sapma değeri verecek şekilde oluşturulmuştur. Bağımlı değişken doğru cevaplanan soru sayısıdır.

**Tablo 7.7** Aronson ve arkadaşlarının yaptıkları çalışmaya (1998) ait veri kümesi

| Kontrol Grubu | | | | Deney Grubu | | | |
|---|---|---|---|---|---|---|---|
| 4 | 9 | 12 | 8 | 7 | 8 | 7 | 2 |
| 9 | 13 | 12 | 13 | 6 | 9 | 7 | 10 |
| 13 | 7 | 6 | | 5 | 0 | 10 | 8 |
| Ortalama = 9,64 | | | | Ortalama = 6,58 | | | |
| Std. sapma = 3,17 | | | | Std. sapma = 3,03 | | | |
| $n_1 = 11$ | | | | $n_2 = 12$ | | | |

Öncelikle yokluk hipotezini, manidarlık düzeyini ve tek ya da iki yönlü testten hangisinin kullanılacağını belirlemeliyiz. Her iki grubun testte eşit düzeyde iyi performans gösterdiğini test etmek istiyoruz, bu nedenle yokluk hipotezini $H_0$: $\mu_1 = \mu_2$ olarak ifade ederiz. Manidarlık düzeyini daha önce olduğu gibi yine ,05 alacağız. Son olarak gruplardan her ikisinin de daha iyi performans göstermesi olası olduğu için iki yönlü test kullanacağız.

Daha sonra birleşik varyansın kestirimini hesaplama ihtiyacındayız.

$$s_B^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{10(3,17^2) + 11(3,03^2)}{11 + 12 - 2}$$

$$= \frac{10(10,0489) + 11(9,1809)}{21} = \frac{201,4789}{21} = 9,5942$$

Son olarak elde edilen birleşik varyansın kestirimini kullanarak *t* değerini hesaplarız.

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_B^2}{n_1} + \frac{s_B^2}{n_2}}} = \frac{(9,64 - 6,58)}{\sqrt{\frac{9,5942}{11} + \frac{9,5942}{12}}} = \frac{3,06}{\sqrt{1,6717}} = \frac{3,06}{1,2929} = 2,37$$

Verilen örnek için serbestlik derecesi $n_1 + n_2 - 2 = 21$'dir. Ek *t*'den 21 serbestlik derecesi için $t_{,025}$ kritik değeri 2,080 olarak bulunur. 2,37 > 2,080 olduğundan yokluk hipotezini reddederiz ve iki evren ortalamasının eşit olmadığı sonucuna varırız.

## Sonuçları Raporlama

Yapılan deneyin sonuçlarını aşağıdaki gibi yazabiliriz:

> Aronson ve arkadaşları 1998 yılında yaptıkları bir çalışmada, bireylerin gerçekte olumsuz yargıların bulunduğu bir klişeye ait olmasa bile klişe baskısı yapıldığında matematik performanslarının düşeceğine dayalı hipotezi test etmişlerdir. İki grup katılımcıdan zor bir matematik sınavı sorularının cevaplandırılması istenmiştir. Katılımcılar matematik başarıları yüksek olan ve matematik başarısını önemseyen beyaz erkeklerden oluşmaktadır. İlk gruba ($n = 11$) matematik sınavı verilmiştir ve soruları cevaplandırmaları istenmiştir. İkinci gruba ($n = 12$) sınavdan önce Asyalı erkeklerin beyaz erkeklerden daha yüksek matematik başarısına sahip oldukları ve bu sınavın temel amacının bu durumun nedenini belirlemek olduğu bilgisi verilmiştir. Her iki grup için de cevaplandırılması istenen sorular aynıdır. Elde edilen sonuçlara göre kontrol grubu ortalama 9,64, deney grubu ise ortalama 6,58 soruyu doğru cevaplamışlardır. Standart sapma değerleri sırasıyla 3,17 ve 3,03'tür. ,99 bulunan etki büyüklüğü değerine (*d*) göre, iki grup soruları doğru cevaplama açısından yaklaşık 1 standart sapma değeri kadar farklılaşmaktadır.

Grupları karşılaştırmak için *t* testi kullanılmıştır. *t* değeri 2,37 olarak hesaplanmış ve p < ,05 olduğu için manidar bulunmuştur. Elde edilen bu sonuca göre, klişe baskısı performansı düşürmektedir. %95 düzeyinde ortalamalar arası farkın güven aralığı $0{,}317 \leq \mu_1 - \mu_2 \leq 5{,}7488$'dir, aralık oldukça geniştir ancak örneklem büyüklüklerinin 11 ve 12 olduğu unutulmamalıdır. Grupları karşılaştırmanın bir diğer yolu ise deney grubunun kontrol grubuna göre %32 oranında daha az doğru cevap verdiğidir.

## 7.6 Varyansların Heterojenliği: Behrens-Fisher Problemi

**varyansların heterojenliği**

Varyansların homojenliği iki bağımsız örneklem için *t* testinin varsayımlarından birinin **varyansların homojenliği** ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) olduğunu gördük. Daha özel olmak için, $H_0$ doğru ve *varyanslar homojen olduğunda* varyanslar birleştirilir ve oran

$$t = \frac{(X_1 - X_2)}{\sqrt{\frac{s_B^2}{n_1} + \frac{s_B^2}{n_2}}}$$

$n_1 + n_2 - 2$ serbestlik derecesinde *t* dağılımına sahiptir. Varyansların homojenliği varsayımı karşılandığında herhangi bir sorun söz konusu değildir ve bu bölümde açıklanacak olan bilgiye gerek yoktur. Ancak varsayım karşılanmadığında ne yapılması gerekir? Eğer varyanslar heterojen ise oranı *t* değeri olarak kabul etmek mümkün değildir. Böyle bir durumda neyse ki çözümü olan bir sorun ile karşı karşıya kalınacaktır.

**heterojen varyanslar**

Öncelikle $\sigma_1^2 = \sigma_2^2 = \sigma^2$ değilse ortalama varyansları birleştirmek (ortalamasını almak) anlamsız olacaktır; çünkü varyansları birleştirmemizin ana sebebi, aynı miktarı kestirdiğimizin varsayılmasıdır. Varyanslar **heterojen** olduğunda ilk vazgeçilecek iş varyansları birleştirmektir ve $s_1^2$, $s_2^2$ heterojen varyansları belirtmek üzere *t'* değerini aşağıdaki gibi tanımlayacağız:

$$t' = \frac{(X_1 - X_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Yukarıdaki belirttiğim gibi *t´*, $n_1 + n_2 - 2$ serbestlik derecesinde *t* dağılımı *göstermeyebilir*. *t´*nün örneklem dağılımı biliniyorsa elde edilen sonucu yorumlamada herhangi bir sorun olmayacaktır. Tek yapılması gereken, örneklem dağılımında $t'$ değerini hesaplamaktır. Neyse ki evrensel bir ittifak olmamasına rağmen *t´* değeri için yaklaşık bir dağılım bilinmektedir.

### *t'* İçin Örneklem Dağılımı

**Behrens-Fisher problemi**

***t'*** için örneklem dağılımı çalışmaları ilk olarak Behrens tarafından başlatılmış ve daha sonra Fisher tarafından genişletilmiştir. Bu nedenle varyansların heterojenliği konusu **Behrens-Fisher problemi** olarak anılır. Behrens ve Fisher'ın yaptıkları çalışmalara dayanarak *t'* dağılımı ve tablosu elde dilmiştir. Ancak tablolar sadece bazı serbestlik derecelerini içerdiği ve kaynakların çoğunda bulunmadığı için kullanılır hale gelmemiştir, bu kitapta da verilmeyecektir.

**Welch-Satterthwaite çözümü**

Bu soruna alternatif bir çözüm önerisi birbirinden bağımsız olarak Welch (1938) ve Satterhwaite (1946) tarafından geliştirilmiştir. **Welch-Satterthwaite çözümü** bilhassa önemlidir ve varyans analizi konusunda bu denkleme tekrar dönülecektir. Bu denklem ile *t´*, *t* dağılımının bilinmeyen serbestlik derecesinde akla uygun bir üyesi olduğu düşünülebilir. Bu durumda problem uygun serbestlik derecesi *sd'nin, sd'* olarak ifade edilen uygun bir çözümünü bulmaktır. Bu çözüm aşağıdaki gibidir:

$$sd' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Serbestlik derecesi ($sd'$) yukarıdaki denklemden elde edilen sonuca en yakın tam sayı olarak alınır.[19] Bu yaklaşıma göre ($sd'$)'nün alacağı en küçük değer ($n_1$ - 1) ve ($n_2$ - 1)'den küçük olan kadardır; en büyük değer de ($n_1 + n_2$ - 2)'dir. Matematiksel olarak ifade edilecek olursa bu durum Min( ($n_1$ - 1), ($n_2$ - 1) ≤ $sd'$ ≤ ($n_1 + n_2$ - 2) biçiminde yazılabilir.

Bu kitapta temel olarak Welch-Satterthwaite yuvarlamasına güveneceğiz. Yaklaşım varyans analizinde ortaya çıkabilecek sorunların çözümüne karşı kolaylıkla uygulanabilir ve diğer çözümlerinden daha garip değildir.

## Varyansların Heterojenliğinin Test Edilmesi

Çalışmanın başında varyansların heterojen olup olmadığını nasıl bilebiliriz? Çünkü, $\sigma_1^2$ ve $\sigma_2^2$ varyanslarını açık olarak bilmiyoruz (bilseydik $t$ için çözmeyecektik); iki örneklem varyansının ($s_1^2$ ve $s_2^2$) farkını kullanarak test etmek için bir yol bulmalıyız.

Varyansların heterojenliğinin test edilmesi için birden fazla yol vardır. En kolaylarından biri Levene (1960) tarafından ortaya atılmıştır. Levene, her $X$ yerine $i$ ve $j$ sırasıyla $j$'inci gruptaki $i$ deneğini temsil etmak üzere, $X_{ij}$ değerinin ortalamadan farkının mutlak değerini $-d_{ij} = \left| X_{ij} - \overline{X} \right|$ −veya ortalamadan farkının karesini $-d_{ij} = (X_{ij} - \overline{X})^2$ −koymayı, daha sonra elde edilen $d_{ij}$ değerleriyle standart iki örneklem $t$ testi yapılmasını önermiştir. Levene testi mantıksal olarak anlamlıdır, çünkü bir grupta değişkenlik daha fazla ise mutlak değeri ya da karesi de büyük olacaktır. $t$ değeri manidar bulunursa iki grubun varyanslarının farklı olduğu sonucuna varılır.[20] Birçok alternatif yaklaşım vardır (Brown ve Forsythe, 1974; O'Brien, 1981) ancak standart yazılımlarla uygulamaları yapılamamaktadır. Levene testi varyansların heterojenliği için en çok kullanılan istatistiktir.[21]

Yukarıda açıklanan yöntemler büyük olan örneklem varyansının küçük olan örneklem varyansına oranı olan $F$ testine alternatif olarak önerilmiştir. İstatistiksel yazılımlar yapılıyor olmasına rağmen, $F$ testi, verinin dağılımının normal olmamasından fazlasıyla etkilenir, böyle durumlarda kullanılmamalıdır.

## Heterojen Varyanslarla Yapılan t Testinin Dayanıklılığı

**dayanıklı**

$t$ testinin, **dayanıklı** olduğunu belirtmiştim. Dayanıklılık, gerekli varsayımların orta düzeyde sağlanmadığında etkilenmemesi durumudur. Bağımsız örneklem $t$ testinde karşı-

19 Welch (1947) daha sonra serbestlik derecesinin aradaki fark göz ardı edilebilir seviyede olmasına rağmen aşağıdaki şekilde elde edilmesinin daha uygun olduğunu tavsiye etmiştir.

$$\left[ \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 + 1}} \right] - 2$$

20 Yapılan Levene testinde çok açık bir sorun söz konudur. Farkın mutlak değeri veya karesi alındığında elde edilecek değer ve değere ait dağılım mutlaka pozitiftir, bunun dağılımı pozitif kayışlı (çarpık) olacaktır. Bu probleme rağmen test çok iyi işlemektedir.

21 Levene testi ile ilgili çalışmalara http://www.ps.uci.edu/~markm/statistics/eda35a.pdf adresinden ulaşılabilir.

lanması gereken iki önemli varsayım ve bir koşul vardır. Varsayımlar ortalamalar arası farkın örnekleme dağılımının normalliği ve varyansların homojenliğidir. Koşul ise, örneklem sayılarının eşit olması veya olmaması durumudur. Varyansların homojenliği karşılanmadığında yapılması gerekenler üzerinde duruldu ancak varyansların heterojen olmasına rağmen standart süreç uygulanırsa *t* değerinin nasıl değişeceği önemli bir sorudur.

Birçok kişi *t* testi varsayımlarının her ikisinin veya tek tek sağlanamamasının *t* testi üzerine etkilerini araştırmıştır. Genel olarak varılan sonuca göre, eşit örneklem büyüklüklerinde, varyansların homojenliği varsayımının karşılanmaması çok küçük bir etki yaratır – $\alpha$ = ,05 manidarlık düzeyinde $\alpha$'nın gerçek değerinin yaklaşık ±,02 kadarı. Bunun anlamı, eşit olmayan varyanslarla bir durum oluşturulduğunda *yokluk hipotezinin doğru olması* durumunda, *t* hesaplandığında elde edilecek değer normal şartlar altında elde edilecek örneklem *t* değerinin %3 ile %7'si arasında sapma gösterecektir. Bu düzeydeki hata göz ardı edilebilir. Benzer bir durum normallik varsayımı sağlanmadığında da geçerlidir. Değişkenlere ait dağılımlar dikkat çekici şekilde çarpık (özellikle ters yönde) ve varyansları eşit değilse ciddî problemler ortaya çıkar.

Değişkenlere ait örneklem büyüklükleri eşit değilse sonuçları yorumlamak daha zordur. Örneklem büyüklükleri önemli sayılacak kadar farklı ise Welch-Satterthwaite yaklaşımının kullanılmasını tavsiye ederim. Bu yaklaşımı kullanarak az şey kaybedilir fakat çok şey kazanılabilir.

Wilcox (1992) çok yönlü dağılımlara sahip grup ortalamalarını karşılaştırırken dengelenmiş örneklemler kullanmanın önemi üzerinde çalışmıştır. (İlginç bir şekilde istatistikçiler dengelenmiş örneklemler konusuyla yakından ilgilenirken psikologlar veya davranış bilimciler konuyu çok önemsememişlerdir.) Wilcox'ın yaptığı çalışmalar ile standart yaklaşımlarla karşılaştırılınca dengelenmiş örneklemlerle daha güçlü sonuçlar elde edildiğini göstermiştir. Yeniden örneklemeye dayanan nonparametrik yaklaşımlara kitabın başka bölümlerinde değinilmiştir. Bu yaklaşımlar evren ile ilgili varsayımları gerektirmeyen güçlü tekniklerdir. Yakın gelecekte yeniden örneklemeye dayanan istatistikler ve ilgili süreçler, başlıca istatistiksel analizler olarak yerini alacaktır.

## Varyansların Homojenliğini Test Etmeli miyiz?

Bu sorunun cevabı, ilk bakışta, çok açık bir şekilde "tabii ki test edilmelidir" gibi görülebilir. Peki gerçekten öyle midir? Hayes ve Cai (2007) homojenlik testinin gereksiz olduğunu ve araştırmacıyı yanlış yöne sevk ettiğini ileri sürmüşlerdir. Geleneksel bakışa göre varyasların homojenliği test edilmelidir. Testin sonucunda varyanslar homojen ancak eşit değilse varyans kestirimleri birleştirilir ancak varyanslar heterojense birleştirmeden ayrı ayrı kullanılır. Hayes ve Cai yaptıkları çalışma ile homojenlik testine dayanarak varyansların ayrı ayrı veya birleştirilerek kullanılmasına karar vermenin ne kadar doğru olduğunu araştırmışlardır. Çalışmanın sonucunda varyansları her durumda ayrı ayrı kullanmanın daha iyi sonuçlar verdiği ortaya çıkmıştır. Varyansların ayrı ayrı kullanılması ile hata oranı ,05 düzeyine daha yakın olacaktır. Bu durumda daha önce önerilenin aksine varyansların ayrı ayrı kullanılması daha doğru olacaktır. Bu yaklaşım ile birkaç serbestlik derecesi hesaba katılmamış olacaktır, ancak sonuçta daha kârlı olunacaktır. Bu probleme Bölüm 12'de tekrar dönülecektir.

## Bir Uyarı

Welch, Satterthwaite, Behrens ve Fisher homojen olmayan varyanslar içeren ortalamaya dayalı testleri geliştirirken araştırmacılara düşünülen kadar çok iyilik yapmamış olabilirler. Venables (2000) geliştirilen bu testlerin, tecrübesiz kullanıcılara, varyans değerleri çok farklı

olsa bile, uygulamış oldukları testlerin çok yararlı olduğu hissi vermesine değinmiştir. Kısaca Venables'a göre varyansların farklı olması durumuna bir çözümün olması araştırmacıları rahatlatan bir etmendir. Ancak eşit olmayan varyanslar konusu gerçekten dikkat edilmesi gereken bir durumdur. Bazı deneylerde varyanslar ortalamadan daha önemli olabilir. Örneğin, anoreksiya tedavisinde aile terapisinin bilişsel davranış terapisi ile karşılaştırıldığı bir deneyin sonucunda ortalamalar aynı ancak varyanslar farklı olabilir. Böyle bir deney için odaklanılması gereken nokta terapilerden biri bazı bireyler üzerinde etkiliyken diğer bireyler üzerinde etkili olmadığını gösteren varyans değeridir. Venables ayrıca terapilerden birinin diğerine göre daha yüksek ortalamaya sahip olmasının, daha yüksek varyans değerine sahip olmadığı sürece önemli olmadığına değinmiştir. Venables çalışmanın sonunda varyansların heterojenliğini düşünmeden *t* testine devam edebiliyor olmanın araştırmacıyı daha rahat hissettirdiği sonucuna varmıştır. Bölüm 11'de yer alan varyans analizi konusu açıklanırken varyansların eşitliği söz konusu olmasa bile, bu durumun sorun olmayacağına dayanarak işlemler yapılacaktır.

Tavsiye etmek istediğim, varyansların heterojenliğinin düzeltilmesi gereken bir problem olarak görülmediği değildir, aksine varyansların farklılığının ne anlatmak istediğinin daha önemli olduğudur.

## 7.7 Hipotez Testlerinin Tekrar Gözden Geçirilmesi

Bölüm 4'te hipotez testi sürecinin üzerinde durduk. Geleneksel yaklaşımda öncelikle yokluk hipotezinin kurulduğunu, ardından yokluk hipotezinin doğru olup olmadığının test edildiğini, daha sonra istatistiksel işlemlerin yapıldığını belirttim. Bir diğer ifadeyle yokluk hipotezinin olma olasılığı hesaplandığını, olasılık düşükse hipotezin reddedildiğini ifade ettim.

Bölüm 4'te ayrıca, probleme farklı bir bakış açısı getiren Jones ve Tukey'nin (2000) önerisine de özet olarak baktık. Bu bölümde incelenen örneklerden bazıları da bu öneri çerçevesinde incelenebilir. Örneğin, Adams ve arkadaşlarının (1996) yaptıkları homofobi çalışmasına Jones ve Tukey'nin bakış açısı ile bakılabilir. Homofobi örneği için geleneksel yaklaşımın adımlarının burada tekrarlanmasına gerek yoktur. Araştırmadan elde edilen sonuca göre her iki koşula ait evren ortalamalarının eşit olduğunu savunan yokluk hipotezi reddedilmiştir. Çalışma sonucundaki ortalamalara ait değer incelendiğinde virgülden sonra 10'uncu basamaktaki değerde görülen fark bir kenara konulsa bile, homoseksüellik hakkında tutumu farklı olan bireylerin erotik videolara verdikleri tepkilerin de farklı olduğu görülmüştür. Hangi grubun daha büyük tepki verdiği bilinmese bile, evrenlerden birinin ortalamasının diğerine göre daha büyük olduğu açıktır ve bu sonuç önemlidir. Elde edilen bu sonuca göre yokluk hipotezi reddedilir ve üç farklı olasılık ortaya çıkar. Bu olasılıklardan birincisi $\mu_H < \mu_{HO}$ olmasıdır, ikincisine göre $\mu_H > \mu_{HO}$ olabilir ya da üçüncü olasılık ise sonucu oluşturacak yeterince kanıt bulunamaması durumudur.

Şimdi bu ihtimallerin hatalarına bakalım. Aslında $\mu_H < \mu_{HO}$ olabilir. Fakat biz homofobik olmayan bireylerin daha çok uyarıldığı kararına vararak bunun tersi bir sonuca ulaştık. Bu Jones ve Tukey'nin "tersine çevirme" dedikleri durumdur ve bu hatanın yapılması olasılığı, eğer tek yönlü hipotez testini kullanıyorsak $\alpha$ = ,05'tir. Alternatif olarak $\mu_H > \mu_{HO}$ olabilir, fakat homofobik olmayan deneklerin daha az uyarıldıkları kararına varırsak yine hata yaparız. Tek yönlü testte bu hatayı yapma olasılığı yine ,05'tir. Bu iki hatayı aynı anda yapmamız mümkün değildir. Çünkü hipotezlerden biri doğrudur. Tek yönlü testin kullanılması halinde (her iki yön için) $\alpha$ = ,05 manidarlık düzeyinde hata oranı %5'tir. Bu tartışmanın bizi götürdüğü sonuç tek yönlü test kullanıldığında sonuçlar elde edilene kadar

endişelenmeye gerek olmadığıdır. Bu düşünce, geleneksel düşünceden büyük ölçüde ayrılmaktadır. Verilen örnekte 62 serbestlik derecesi için kritik değer yaklaşık 1,68, hesaplanan *t* değeri ise 2,48'dir. Bu değerlere dayanarak homofobik bireylerin homofobik olmayan bireylere göre daha fazla tepki verdiği söylenebilir. Dikkat edilmesi gereken noktalardan biri, burada aktarılan hata konusu içerisinde "1. Tip hata" kavramından bahsedilmediğidir. Çünkü 1. Tip hata türü gerçek yokluktan bahseder ancak bu örnekte gerçek yokluk söz konusu değildir.

## Anahtar Terimler

Ortalamanın örneklem dağılımı (7.1)
Merkezî limit teoremi (7.1)
Tek biçimli (dikdörtgensel) dağılım (7.1)
Standart hata (7.2)
Student *t* dağılımı (7.3)
Özyükleme (7.3)
Nokta kestirimi (7.3)
Güven sınırları (7.3)
Güven aralığı (7.3)
İnanılır aralık (7.3)
*p* düzeyi (7.3)
Eşleştirilmiş örneklemler (7.4)
Tekrarlı ölçümler (7.4)
İlişkili örneklemler (7.4)
Fark puanları (7.4)
Kazanılan puanlar (7.4)
Cohen'ın d'si (7.4)
Ortalamalar arası farkın örneklem dağılımı (7.5)
Varyansların toplamı yasası (7.5)
Ortalamalar arası farkın standart hatası (7.5)
Ağırlıklı ortalama (7.5)
Birleşik varyans kestirimi (7.5)
*sd´* (7.6)
Varyansların homojenliği (7.7)
Heterojen varyanslar (7.7)
Behrens-Fisher problemi (7.7)
Welch-Satterthwaite denklemi (7.7)
Dayanıklılık (7.7)

## Alıştırmalar

7.1 Aşağıda ortalaması 4,5, standart sapması 2,7 olan dikdörtgensel dağılım gösteren 100 tesadüfî sayı verilmiştir. Bu veri kümesinin dağılımının grafiğini çiziniz.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 4 | 8 | 7 | 8 | 7 | 0 | 8 | 2 | 8 | 5 | 7 |
| 4 | 8 | 2 | 6 | 9 | 0 | 2 | 6 | 4 | 9 | 0 | 4 |
| 9 | 3 | 4 | 2 | 8 | 2 | 0 | 4 | 1 | 4 | 7 | 4 |
| 1 | 7 | 4 | 2 | 4 | 1 | 4 | 2 | 8 | 7 | 9 | 7 |
| 3 | 7 | 4 | 7 | 3 | 1 | 6 | 7 | 1 | 8 | 7 | 2 |
| 7 | 6 | 2 | 1 | 8 | 6 | 2 | 3 | 3 | 6 | 5 | 4 |
| 1 | 7 | 2 | 1 | 0 | 2 | 6 | 0 | 8 | 3 | 2 | 4 |
| 3 | 8 | 4 | 5 | 7 | 0 | 8 | 4 | 2 | 8 | 6 | 3 |
| 7 | 3 | 5 | 1 | | | | | | | | |

© Cengage Learning 2013

7.2 Alıştırma 7.1'deki veri kümesinden, her birinde beşer puan bulunan 50 örneklem tesadüfî olarak oluşturulmuş, her örneklemin ortalaması alınmış, bu ortalamalar aşağıda verilmiştir. Ortalamaların dağılımını çiziniz.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2,8 | 6,2 | 4,4 | 5,0 | 1,0 | 4,6 | 3,8 | 2,6 | 4,0 | 4,8 |
| 6,6 | 4,6 | 6,2 | 4,6 | 5,6 | 6,4 | 3,4 | 5,4 | 5,2 | 7,2 |
| 5,4 | 2,6 | 4,4 | 4,2 | 4,4 | 5,2 | 4,0 | 2,6 | 5,2 | 4,0 |
| 3,6 | 4,6 | 4,4 | 5,0 | 5,6 | 3,4 | 3,2 | 4,4 | 4,8 | 3,8 |
| 4,4 | 2,8 | 3,8 | 4,6 | 5,4 | 4,6 | 2,4 | 5,8 | 4,6 | 4,8 |

© Cengage Learning 2013

7.3 Alıştırma 7.1'de ve 7.2'de yer alan veri kümelerine ait ortalama ve standart sapmaları karşılaştırınız.

a. Böyle bir durumda merkezî limit teoremine göre çıkacak sonuç ne olacaktır?

b. Elde edilen sonuç beklentiyle örtüşüyor mu?

7.4 Alıştırma 7.2'de yer alan değerlerde, örneklemlerde yer alan değer sayısı 5 olup örneklemdeki eleman sayısı 5'ten fazla olursa sonuçta ne gibi farklılıklar söz konusu olur?

7.5 Alıştırma 7.2'de yer alan değerlerde örneklem sayısı 50 olarak tutulur ve her örneklemdeki değer 15'e çıkarılsa sonuçta ne gibi farklılıklar söz konusu olur?

7.6 Kruger ve Dunning (1999) yayımladıkları "Yeteneksiz ve bunun farkında değil" isimli çalışmalarında, mantıksal düşünme yeteneği kötü olan bireylerin bu yetenekleri hakkında bilgileri olmadığına yönelik bir hipotez test edilmiştir. Çalışmada her öğrencinin mantıksal düşünme yeteneği testinden yüzde kaç puan aldıklarını tahmin etmeleri istenmiştir. En düşük çeyrekte yer alan 11 birey kendilerini 68'inci yüzdelikte yetenekli olarak tahmin etmişlerdir. Bireylerin ham puanları şöyledir: 40, 58, 72, 73, 76, 78, 52, 72, 84, 70, 72. Bu bilgi bütün öğrencilerin ortalama üzerinde olduğuna dair bir örnek midir? Diğer bir deyişle, yüzde sıralamalarının ortalaması 50'den büyük müdür?

7.7 Geleneksel hipotez testleri için tek yönlü testlerin çok uygun olmadığını açıklamama rağmen yukarıda verilen örnek için tek yönlü test neden daha uygundur?

7.8 Bir önceki iki örnekte yer alan Kruger ve Dunning çalışmasından elde edilen sonuçlara göre 11 öğrencinin ortalaması 70, standart sapması 14,92'dir. Buna göre bireyler yeteneklerini olduğundan düşük tahmin etmişlerdir. Bu fark manidar mıdır?

7.9 Kişinin, performansını düşük ya da yüksek tahmin etmesi, düşük yeteneklinin kendini daha düşük, yüksek yeteneklinin ise kendisini daha yüksek algılaması ile ilgilidir. O halde bu açıklama niçin düşük puanlı bireylerin kendi puanları için çok yüksek tahmin yapmasını açıklamaz?

7.10 Alıştırma 7.8'de yer alan veriye ait %95 olasılıklı güven aralığını hesaplayınız.

7.11 Everitt, 1994 yılında yaptığı çalışmada, anoreksiya tedavisi için birkaç terapi yöntemi önermiştir. Çalışmanın bir bölümünde 29 kız bilişsel davranışçı terapi ile tedavi olmuşlardır ve kızların kiloları terapi öncesi ve sonrasında ölçülmüştür. Kızların terapi sonrasında kazandığı kilolar aşağıda verilmiştir. Veri terapi sonrası ölçülen ağırlık değerinden terapi öncesi ölçülen değer çıkarılarak elde edilmiştir. Bu nedene negatif değerler terapi sonrası kilo kaybı, pozitif değerler ise kilo alışı yaşandığını göstermektedir.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.7 | 0.7 | –0.1 | –0.7 | –3.5 | 14.9 | 3.5 | 17.1 | –7.6 | 1.6 | 11.7 |
| 6.1 | 1.1 | –4.0 | 20.9 | –9.1 | 2.1 | –1.4 | 1.4 | –0.3 | –3.7 | –0.8 |
| 2.4 | 12.6 | 1.9 | 3.9 | 0.1 | 15.4 | –0.7 | | | | |

a. Değerlerin dağılımı neye benzemektedir?

b. Grupta yer alan kızlar istatistiksel olarak manidar derecede kilo almışlar mıdır?

c. Sonuçların %95 düzeyde güven sınırlarını hesaplayınız.

d. Sonuçların etki büyüklüğü değerini hesaplayınız.

7.12 Alıştırma 7.11'de yer alan kilo artışlarının %95 olasılıklı güven sınırlarını hesaplayınız.

7.13 Katz, Lautenschlager, Blackburn ve Harris (1990) 28 öğrencinin SAT sınavında yer alan çoktan seçmeli maddeleri, ilgili olduğu paragrafları okumadan cevaplama performanslarını ölçmüşlerdir. Elde edilen veri ortalama puan 100 üzerinden 46,6 ve standart sapması 6,8'dir. Şansla doğru cevaplama beklentisi 20 doğru cevaptır.

a. Öğrencilerin, cevaplama performansları şans düzeyinin üstünde midir?

b. Öğrencilerin performansları şansın üstünde ise, elde edilen bu sonuç SAT sınavının lise performansını belirlemede geçerli olmadığı anlamına gelir mi?

7.14 Compas ve arkadaşları (1994), yaptıkları bir çalışma ile stres altındaki çocukların düşünülenden daha az kaygı ve depresyon belirtileri gösterdiklerini öne sürmüşlerdir. Ayrıca çalışmada bu çocukların "yalan söyleme ölçeğinden" (toplumsal olarak beklenen cevapları verme eğilimi ölçeği) elde edilen puanlarının beklenenden daha yüksek olduğu belirtilmiştir. Yalan söyleme ölçeğinin evren ortalaması 3,87'dir. Yapılan çalışmada yer alan stres altındaki 36 çocuğun ortalaması 4,39, standart sapması ise 2,61 bulunmuştur.

a. 38 öğrencinin yalan söyleme eğiliminin yüksek olup olmadığı nasıl test edilebilir?

b. Yokluk hipotezi ve alternatif hipotez ne olacaktır?

c. Veri kullanılarak nasıl bir sonuca varılabilir?

7.15 Alıştırma 7.14'te yer alan durum için %95 olasılıklı güven sınırlarını hesaplayınız. Elde edilen güven sınırları 7.14'te elde ettiğiniz sonuç ile uyumlu mudur?

7.16 Hoaglin, Mosteller ve Tukey (1983), yaptıkları çalışmada stresin bir göstergesi olarak kandaki beta endorfin seviyesini ölçmüşlerdir. 19 hastanın ameliyata girmeden 12 saat önce ve 10 dakika önce endorfin seviyelerini kaydetmişlerdir. Veri aşağıdaki gibidir:

| Kişi | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 12 saat | 10,0 | 6,5 | 8,0 | 12,0 | 5,0 | 11,5 | 5,0 | 3,5 | 7,5 | 5,8 |
| 10 dakika | 6,5 | 14,0 | 13,5 | 18,0 | 14,5 | 9,0 | 18,0 | 42,0 | 7,5 | 6,0 |

| Kişi | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|
| 12 saat | 4,7 | 8,0 | 7,0 | 17,0 | 8,8 | 17,0 | 15,0 | 4,4 | 2,0 |
| 10 dakika | 25,0 | 12,0 | 52,0 | 20,0 | 16,0 | 15,0 | 11,5 | 2,5 | 2,0 |

© Cengage Learning 2013

Yukarıdaki veriye dayanarak stresin endorfin seviyesi üzerindeki etkisi nedir?

7.17 Alıştırma 7.16'da yer alan her iki ölçüm kümesi için gerçek ortalama farklarına ait %95 olasılıkla güven aralığını bulunuz.

7.18 Alıştırma 7.16'da yer alan veri için etki büyüklüğünü hesaplayınız.

7.19 Hout, Duncan ve Sobel (1987), evli çiftlerin cinsel doyum oranları üzerine çalışmışlardır. 91 evli çifte "Cinsel ilişki benim ve eşim için zevklidir." cümlesine katılma dereceleri sorulmuş ve cevap olarak 1'den (hiçbir zaman ) 4'e (her zaman) kadar bir oran vermeleri istenmiştir. Elde edilen veri aşağıdaki gibidir: (Verinin fazla olduğunu biliyorum, fakat enteresan bir sorudur.)

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kocası | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Eşi | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| Kocası | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Eşi | 3 | 4 | 4 | 4 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| Kocası | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| Eşi | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 2 | 2 | 2 | 2 | 2 |
| Kocası | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 |

© Cengage Learning 2013

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eşi | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 1 | |
| Kocası | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | |
| Eşi | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | |
| Kocası | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Eşi | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Veri kümesi için eşleştirilmiş örneklem *t* testi ile işe başlayınız ve eşleştirilmiş örneklem *t* testinin neden uygun olduğunu açıklayınız.

7.20 Alıştırma 7.19'da yapılan analiz sonucunda çiftlerin cinsel olarak birbirleriyle uyumlu olup olmadıkları sonucuna varılabilir mi? Analiz sonuçlarına dayanarak hangi yorumlar yapılabilir ve hangi yorumlar yapılamaz?

7.21 Uygun bir bilgisayar programı kullanarak Alıştırma 7.19'da yer alan erkekler ve kadınların cinsel doyumları arasındaki ilişkiyi hesaplayınız. Elde edilen sonuç Alıştırma 7.19'da yapılan analizi nasıl etkiler? (Bölüm 9'a kadar korelasyon konusuna değinilmemiştir, bu nedenle sorunun çözümünde herhangi bir zorluk yaşanıyorsa Bölüm 9'a kısaca bir göz atılabilir. Korelasyon hesaplamaları için SPSS kullanılabilir.)

7.22 Alıştırma 7.19'da bulunan eşler arası cinsel doyum puanları ortalaması farkına ait %95 olasılıklı güven aralığını bulunuz ve elde edilen sonucu veriye dayanarak yorumlayınız.

7.23 Bazıları Alıştırma 7.19'da yer alan verinin sıralı (süreksiz) olduğu için *t* istatistiğinin uygulanmasına itiraz edebilirler. Bu itiraza karşı gelecek görüş ne olabilir? (Sorulan bu soru kolayca cevaplanabilecek özellikte değildir ve gerçekten böyle bir çelişkinin olması söz konusu olduğu için sorulmuştur.)

7.24 İlişkili örneklemin yer aldığı ancak değişkenlerden birinin diğerini etkileme olasılığının olduğu bir deney örneği veriniz.

7.25 Sullivan ve Bybee (1999), eşlerinden şiddet gören kadınlar için bir müdahale programı geliştirmişlerdir. Çalışmada deney ve kontrol grubu olmak üzere iki grup yer almıştır ve 10 hafta boyunca program uygulanmıştır, daha sonra programa katılan kadınlar üç sene izlenmiştir. 10 haftanın sonunda katılımcılara yaşam kalitesi ölçeği uygulanmıştır. 135 kişiden oluşan deney grubunun ortalaması 5,03; standart sapması ise 4,61 bulunmuştur. 130 kişiden oluşan kontrol grubunun, ortalaması 4,61; standart sapması 1,13 elde edilmiştir. Verilen bu değerlere göre uygulanan müdahale programı başarılı olmuş mudur?

7.26 Alıştırma 7.25'te yer alan gruplar arası ortalamalar farkının %95 olasılıklı güven aralığını hesaplayınız.

7.27 Alıştırma 7.25'te sözü edilen müdahale programının etkililiğini ölçmenin bir diğer yolu için müdahaleden önce ve sonra deney grubuna yaşam kalitesi ölçeği uygulanmıştır. Müdahaleden önce uygulanan ölçekten elde edilen değerlerin ortalaması 4,47; standart sapması 1,18; müdahaleden sonra uygulanan ölçekten elde edilen değerlerin ortalaması 5,03 ve standart sapması 1,01'dir. Her iki ölçüm için de örneklem büyüklüğü 135'tir. Verilen bu değerlere göre müdahalenin etkililiği hakkında ne söylenebilir? (Fark puanları verilmemiştir ancak fark puanlarının standart sapmasını 1,30 olarak varsayınız.)

7.28 Alıştırma 7.25'te bulunan kontrol grubunun müdahaleden önce ve sonraki ortalaması sırasıyla 4,32 ve 4,61; standart sapması ise sırasıyla 0,98 ve 1,13'tür. Kontrol grubunun örneklem büyüklüğü 130'dur. Kontrol grubu için burada verilen ve deney grubu için daha önce verilen veri kümelerini kullanarak her iki grup için ön test ve son test puanları arasındaki farkı inceleyiniz. (Kullanılması gerekirse fark puanlarının standart sapması 1,25'tir.)

7.29 Katz ve arkadaşlarının (1990) yaptıkları ve Alıştırma 7.13'te değinilen çalışmada SAT sınavında paragrafları okuyarak soruları cevaplayan 17 öğrenci ile paragrafları okumadan soruları cevaplayan 28 öğrencinin performans puanları karşılaştırılmıştır. İlk grubun ortalaması ve standart sapması sırası ile 69,6 ve 10,6; ikinci grubun ise 46,6 ve 6,8'dir.

a. Yokluk hipotezi nedir?
b. Alternatif hipotez nedir?
c. Uygun *t* testini uygulayınız.
d. Sonuçları yorumlayınız.

7.30 Birçok anne doğumdan sonra kısa bir dönem için depresyona girer. Annelerin doğum sonrası depresyonunu test eden bir çalışma deseni kurunuz ve bu bölümde açıklanan yöntemleri kullanarak depresyon ortalamasındaki artışın nasıl kestirilebileceğini açıklayınız.

7.31 Alıştırma 7.25'te Everitt'in yaptığı çalışma yer almıştır. Bu çalışmaya göre bilişsel davranış terapisi anoreksiya tedavisinde etkili olmuştur. Ancak bu süreç içerisinde kilo almanın bir sebebi de yaşlanma olabilir. Bu durumu kontrol etmenin bir yolu, bilişsel terapi alan bir deney grubu (n = 29) ile terapi almayan kontrol grubunun (n = 26) kilo alma düzeylerini karşılaştırmaktır. İki grubun kilo almalarına dayalı veri kümesi aşağıdaki gibidir.

| Kontrol | | Bilişsel Terapi | |
|---|---|---|---|
| –0,5 | 3,3 | 1,7 | –9,1 |
| –9,3 | 11,3 | 0,7 | 2,1 |
| –5,4 | 0,0 | –0,1 | –1,4 |
| 12,3 | –1,0 | –0,7 | 1,4 |
| –2,0 | –10,6 | –3,5 | –0,3 |
| –10,2 | –4,6 | 14,9 | –3,7 |
| –12,2 | –6,7 | 3,5 | –0,8 |
| 11,6 | 2,8 | 17,1 | 2,4 |
| –7,1 | 0,3 | –7,6 | 12,6 |
| 6,2 | 1,8 | 1,6 | 1,9 |
| –0,2 | 3,7 | 11,7 | 3,9 |
| –9,2 | 15,9 | 6,1 | 0,1 |
| 8,3 | –10,2 | 1,1 | 15,4 |
| | | –4,0 | –0,7 |
| | | 20,9 | |
| **Ortalama** | **–0,45** | | **3,01** |
| **Standart Sapma** | **7,99** | | **7,31** |
| **Varyans** | **63,82** | | **53,41** |

Grup ortalamalarını uygun *t* testini kullanarak karşılatırınız. Elde edilen sonuçları yorumlayınız.

7.32 Alıştırma 7.31'deki veriye ait %95 olasılıklı güven aralığını ve *d* değerini hesaplayınız. Söz konusu durum için şartlar uygunsa güven aralığının bulunmasında daha önce sözü edilen bilgisayar programını kullanınız.

7.33 Alıştırma 7.19'da eşlerin cinsel doyumlarına ait veriyi ve cinsel doyumlarını gördük. Bu verinin birbiriyle ilişkisi olmayan erkek ve kadınlardan toplandığını varsayınız. Böyle bir durumun analize etkisi ne olabilir?

7.34 Alıştırma 7.19'daki veride gözlemlerin bağımsız olduğunu düşünerek *t* testi analizi yapınız. Elde edilen sonucu yorumlayınız.

7.35 Alıştırma 7.34 ve 7.19 ile elde edilen sonuçlar arasındaki fark neden olması gerekenden daha büyüktür?

7.36 Alıştırma 7.31'deki Everitt'in anoreksiya üzerine yaptığı çalışmada tesadüfî atamanın rolü nedir ve hangi şartlar altında tesadüfî atama yapmak daha zordur?

7.37 Konusal Kavrama Testi'nde kişilere ne anlama geldiği belli olmayan resimler gösterilir ve kişilerden bu resimlerle ilgili bir hikâye anlatmaları istenir. Bu hikâyeler birçok şekilde puanlanabilir. Werner, Stabenau ve Pollin (1970) yaptıkları çalışmada, 20 normal, 20 şizofren çocuğun annelerine Konusal Kavrama Testi uygulamışlar ve olumlu anne-çocuk ilişkisini yansıtan hikâye sayısına göre (10 üzerinden) puanlamışlardır. Veri aşağıdaki gibidir:

© Cengage Learning 2013

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Normal | 8 | 4 | 6 | 3 | 1 | 4 | 4 | 6 | 4 | 2 |
| Şizofren | 2 | 1 | 1 | 3 | 2 | 7 | 2 | 1 | 3 | 1 |
| Normal | 2 | 1 | 1 | 4 | 3 | 3 | 2 | 6 | 3 | 4 |
| Şizofren | 0 | 2 | 4 | 2 | 3 | 3 | 0 | 1 | 2 | 2 |

© Cengage Learning 2013

a. Çalışmaya ait sıfır hipotezi nedir?

b. Hipotezi, uygun analiz yöntemiyle test edip sonucu yorumlayınız.

7.38 Alıştırma 7.37 için iki grubun varyans değerlerine bakmak neden daha akıllıcadır?

7.39 Alıştırma 7.37'de elde edilecek manidar sonuç, anne-çocuk ilişkisinin düşük olması durumunda çocukların şizofren olabileceği şeklinde yorumlanır. Böyle bir sonuç niçin "sorunlu" olarak tanımlanabilir?

7.40 Birçok araştırmada deneklerin istenilen yönde sonuçlar vermesine yönelik konular seçilir. (Araştırmacılar görmek istediklerini görürler.) Öğrencilerin denek olarak kullanıldığı düşünülsün. Bütün deneklere deneyden önce kafein alacakları bilgisi verilmiştir ancak yarısına kafeinin performansı arttırdığı, diğer yarısına ise düşürdüğü açıklanmıştır. Bağımlı değişken öğrencilerin 2 dakika içerisinde çözdükleri basit aritmetik problemleri sayısıdır. Veri kümesi aşağıdaki gibidir.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| İyi beklenti: | 19 | 15 | 22 | 13 | 18 | 15 | 20 | 25 | 22 |
| Zayıf beklenti: | 14 | 18 | 17 | 12 | 21 | 21 | 24 | 14 | |

Nasıl bir sonuca varırsınız?

7.41 Alıştırma 7.40'taki veride $\mu_1 - \mu_2$ için %95 olasılıklı güven aralığını bulunuz.

7.42 Karar verme düzeyini çalışan bir araştırmacı, 10 öğrenciden 10 dakika içinde çözebildiği kadar problemi çözmelerini istemiştir. Gruplardan birine (5 kişi) bu testin problem çözme yeteneklerini ölçtüğü, diğer gruba (5 kişi) zaman doldurma testi olduğu bilgisi verilmiştir. Veri aşağıdaki gibidir:

| | | | | | |
|---|---|---|---|---|---|
| Doğuştan yetenek: | 4 | 5 | 8 | 3 | 7 |
| Zaman doldurma: | 11 | 6 | 9 | 7 | 9 |

Çözülen problem sayısının ortalaması deney şartlarına göre değişmekte midir?

7.43 Başka bir araştırmacı Alıştırma 7.42'de bulunan deneyi yeniden uygulamıştır ve aynı sonuçları elde etmiştir. Ancak araştırmacı sonuçları 10 dakika içinde çözülen problem sayısı şeklinde değil, her soru için harcanan ortalama süre cinsinden verilmesinin daha uygun olduğunu düşünmüştür (örneğin, 10 dakikada 4 soru =10/4 =2,5 dakika/problem). Veri kümesi aşağıdaki gibidir:

| | | | | | |
|---|---|---|---|---|---|
| Doğuştan yetenek: | 2,50 | 2,00 | 1,25 | 3,33 | 1,43 |
| Zaman doldurma: | 0,91 | 1,67 | 1,11 | 1,43 | 1,11 |

Bu veriyi uygun bir *t* testi ile analiz ediniz ve sonucu yorumlayınız.

7.44 Alıştırma 7.42 ve 7.43'ün karşılaştırılması ile elde edilen sonuç nedir?

7.45 Daha önce de belirttiğim gibi Levene testi, gözlemler arasındaki mutlak farkın ve ortalamanın hesaplanması, ardından bu farka *t* testi uygulanmasından oluşur. Herhangi bir bilgisayar programı ile mutlak farkı hesaplamak ve *t* testi uygulamak oldukça kolaydır. Hem farkı hem de *t* testini hesaplayınız ve SPSS'in hangi yaklaşımı kullandığını belirleyiniz. (İpucu, bu örnekte $F = t^2$ ve SPSS ile hesaplanan *F* değeri yaklaşık olarak 6 ondalık basamağa kadar 0,391148'dir.)

7.46 Klinikte yapılan bir çalışmaya göre ebeveynlerinden birinin ölümüne şahit olan çocukların yetişkinliklerinde depresyon ve kaygı eğilimlerinde artış olduğu gözlenmiştir. Mireault ve Bond (1992) ebeveynlerinden biri ölmüş olan 140, her iki ebeveyni hayatta olan 182 ve ailesi boşanmış olan 59 lise öğrencisinden veri toplamışlardır. Veri kümesi Mireaut.dat isimli dosyada, açıklamaları ise "Bilgisayar Alıştırmaları" isimli ekte yer almaktadır.

a. Herhangi bir istatistik programı kullanarak ilk iki gruba ait Kısa Semptom Envanteri ile elde edilmiş depresyon, kaygı ve genel semptom indeksi puanlarını karşılaştırınız (Derogatis, 1982).

b. Yapılan üç *t* testi birbirinden bağımsız mıdır? (İpucu: Problemi çözebilmek için üçüncü gruba ait veri kümesini ihmal etmek ya da silmek gereklidir. Bu işlemin nasıl yapılacağı kullanılan bilgisayar programının kullanma kılavuzunda yer alır.)

7.47 Genel olarak kadınların kaygı ve depresyon belirtilerini erkeklere göre daha fazla gösterdiği raporlandırılmıştır. Mireault'un yaptığı çalışmada toplanan veri bu hipotezi destekler mi?

7.48 Mireault'un çalışmasında yer alan ayrı ayrı Grup 1 - Grup 2; Grup 1 - Grup 3; Grup 2 - Grup 3'ün genel semptom indeksi puanlarını t testi kullanarak karşılaştırınız. (Üç farklı grubun ortalamalarını karşılaştırmak için bu uygun bir yöntem değildir ancak daha uygun olan analiz yöntemi Bölüm 12'de açıklanacaktır.)

7.49 Alıştırma 7.25'teki eşleştirilmiş çiftlere ait verinin anlamlı etki büyüklüğünü hesaplayınız.

7.50 Alıştırma 7.31'deki iki bağımsız gruba ait verinin anlamlı etki büyüklüğünü hesaplayınız.

7.51 Bölüm 6 Alıştırma 6.34'te, Hout ve arkadaşlarının, evli çiftlerin cinsel doyumları üzerine yaptıkları çalışmadan elde edilen veri ile çapraz tablo oluşturulmuş ve ki-kare değeri elde edilmiştir. Alıştırma 7.19'da aynı veri farklı bir bakış açısı ile yeniden analiz edilmiştir. Deneklere "Cinsel ilişki benim ve eşim için eğlencelidir." tepkileri alınmıştır ve "asla, bazen, sık sık, her zaman" seçeneklerinden birini seçmek yerine 1'in "asla", 4'ün "her zaman" seçeneğine karşılık geldiği 4 dereceli ölçek haline dönüştürülmüştür. Böylece veri kategorik olmaktan çıkmış ve sürekli hale getirilmiştir.

a. Veriye  ait ölçek düzeyinin analiz ile ilişkisi nedir?

b. Bu alıştırma ve alıştırma 6.35'te farklı istatistiksel testler kullanılması bir yana bırakıldığında iki alıştırma, veri için farklı soruların cevabını aramaktadır. Bu farklı sorular nelerdir?

c. Kadınların 15'i eşleri cevapladığı halde cevaplamayı reddetmiş, erkeklerin de 8'i eşleri cevapladığı halde soruyu cevaplamadığı varsayılsın. Böyle bir durumda ne yapılabilir?

d. Yapılan t-testi analizi için ne düşünüyorsunuz? Alternatif olarak nasıl bir yolla analiz edilebilirdi?

7.52 Alıştırma 7.31'deki deneyin sonuçlarını açıklamak için gerekli olan bilgiyi içeren kısa bir paragraf yazınız. Yazılan bu paragraf araştırma makalesinde yazılması gereken sonuç bölümünün kısa versiyonu olacaktır.



Bölüm 8

# Güç

## Hedefler

Bir istatistiksel teste ait güç kavramını tanıtma ve çeşitli istatistiksel yöntemlerin gücünü nasıl hesaplayabileceğimizi gösterme.

## Konular


**8.1** Temel Güç Kavramı

**8.2** Bir Testin Gücünü Etkileyen Faktörler

**8.3** Gücün Geleneksel Yolla Hesaplanması

**8.4** Tek-Örneklem *t* İçin Güç Hesaplamaları

**8.5** Birbirinden Bağımsız İki Ortalama Arasındaki Fark İçin Güç Hesaplamaları

**8.6** Eşleştirilmiş-Örneklemler *t* Testi İçin Güç Hesaplamaları

**8.7** Güç Üzerine Olan Durumu Tersine Çevirme

**8.8** Daha Kompleks Desenlerde Güç Düşünceleri

**8.9** Hesaplamaları Basitleştirmek İçin G*Power'ın Kullanılması

**8.10** Geriye Dönük Güç

**8.11** Güç Analizinin Sonuçlarının Raporlanması

UYGULAMALI İSTATİSTİK ÇALIŞMALARIN ÇOĞU, deneysel sonuçların analizleri için gerçekleştirildiğinden, öncelikle I. Tip hata olasılığını ($\alpha$) azaltma (en azından kontrol etme) ile ilgilenmektedir. Deneyleri tasarlarken araştırmacılar II. Tip hata olarak bilinen çok önemli bir hata olasılığını ($\beta$) ihmal etme eğiliminde olurlar. I. Tip hata gerçekte olmayan bir farkı bulma sorunu ile ilgilenirken, II. Tip hata, aynı ölçüde ciddî bir problem olan ve gerçekte olduğu halde bulunamayan fark sorunu ile ilgilenir. Tipik bir deney için geçerli olan zaman ve para maliyetlerini düşündüğümüz zaman, araştırılan etki evrende var olsa bile araştırmanın başında bu farkı bulma olasılığının düşük olabileceğini bilememenin, araştırcıların bakış açılarının darlığından kaynaklandığını söyleyebiliriz.[1]

Araştırcıların II. Tip hatayı neden ihmal etme eğiliminde olduklarına dair oldukça iyi tarihsel nedenler vardır. Cohen (1988), $H_1$'e yeterince dikkat edilmeden sıfır hipotezinin doğru ya da yanlış olduğuna bakıldığına dair Fisher'ın yaptığı vurguyu ilk neden olarak belirtmektedir. Neyman-Pearson $H_1$'in önemine vurguda bulunmasına rağmen, Fisher'ın görüşü oldukça etkili olmuştur. Buna ek olarak son zamanlara kadar birçok ders kitabı bu probleme değinmekten sakınmış, güç konusuna değinen kitaplar ise konuyu ortalama bir okuyucunun kolayca anlayabileceği açıklıkta açıklamamışlardır. Fakat Cohen, bu problemi birkaç yayınında açıkça tartışmıştır.[2] Cohen (1988), bu probleme birçok yayınında oldukça ayrıntılı bir açıklama getirmiştir. Welkowitz, Ewen ve Cohen'ın (2000) giriş niteliğindeki kitabında bu problem, yakınlaştırma (aproximation) tekniği kullanılarak oldukça basit bir düzeyde açıklanmıştır. Eldeki ünitede bu yaklaşım benimsenmiştir. Kullanışlı yöntemler içeren oldukça iyi iki çalışma Cohen'da (1992a, 1992b) bulunur. Bu kaynakların hiçbirinde veya bu konuda ve doğrudan bu konu ile ilgili olmasa da Cohen'ın birçok konuda yazdığı mükemmel yayınları anlamada zorluk yaşamamalısınız.

**güç**

Bize hata yapmış olabileceğimizi düşündürdüğünden, II. Tip hatalar açısından yaklaşmak probleme negatif yönden bakmaktır; çünkü bu bize hata yapabileceğimizi düşündürmektedir. Daha pozitif bir yaklaşım, alternatif hipotez doğru iken yanlış olan $H_0$'ın doğru bir şekilde reddedilmesi olasılığı olarak tanımlanan güç açısından konuşmaktır. Böylece, **güç** = 1 – $\beta$'dır. Güçlü bir deney, yanlış bir $H_0$'ı reddetmek için daha iyi bir şansa sahip olandır.

Bu bölümde genellikle Welkowitz, Ewen ve Cohen'ın (2000) yaklaşımını ve bize bir testin gerçek gücü için iyi bir yaklaşım sunan bir yöntemi kullanacağız. Bu yaklaşım özellikle, gücün ,85 veya ,83 olup olmadığının değil ,80'e veya ,30'a yakın bir değerde olup olmadığının önemsendiği durumda kusursuzdur. Özellikle bu bakış açısından incelendiğinde, bu ünitede sunulan yaklaşımın mükemmel olduğu görülür. Bu konuda oldukça iyi bir yazılım vardır ve biz özellikle iyi bir programa burada kısaca değineceğiz, fakat burada kesinliklerine ilişkin kazanım yanıltıcıdır. Bununla şunu söylemek istiyorum, yazılımın bir bölümünden elde ettiğimiz bir cevap kesinlikten daha çok kesinliğin bir yansımasını verebileceğinden, güç analizlerinde yaptığımız parametre kestirimleri yeterince hata içerir. Burada kullanacağım yöntem, gücün hesaplanmasında yer alan kavramlara açıklık getirecektir ve eğer daha kesin cevaplar elde etmek isterseniz oldukça iyi ve ücretsiz olan yazılımlar indirebilirsiniz. Bu mükemmel yazılımlardan biri olan Faul ve Erdfelder tarafından geliştirilen G*Power programına internetten http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ adresinden ulaşabilirsiniz ve bu sitede Macintosh ve Windows programlarının ikisini de bulabilirsiniz. Bu ünitede önce el ile güç hesaplamalarını göstereceğim. Sonra G*Power programı kullanılarak elde edilen sonuçları ve programın sunduğu avantajlara yer vereceğim.

Bu yaklaşım birçok hipotezin test edilmesine kadar genelleştirilebilir olmasına rağmen, konuyu açıklama amaçlı olarak şu anda iki örneklem ortalamaları arasındaki farkı test etme ile ilgileneceğiz.

---

1 Son zamanlarda dergi editörleri güç ile ilgili bilgiler sorma konusunda daha girişkendirler. Fakat zannediyorum güç analizlerinin çoğu editor veya hakemlerin isteklerini karşılamak için deneyler tamamlandıktan sonra yapılmaktadır.

2 Farklı bir yaklaşım Murphy ve Myors (1998) tarafından getirilmiştir. Bu yazarlar tüm güç hesaplamalarını *F* dağılımına dayandırmıştır. *F* dağılımı bu kitabın tamamında görülür ve aslında bu kitapta yer verilen tüm istatistikler bir *F*'ye dönüştürülebilir. Murphy ve Myors'ın yaklaşımı incelenmeye değerdir ve bu ünitede bulduğumuz sonuçlara oldukça yakın sonuçlar verecektir.

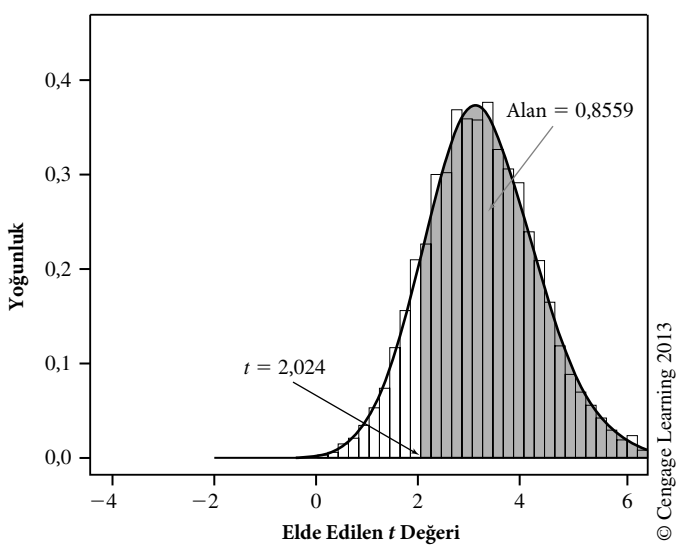
# 8.1 Temel Güç Kavramı

Güç kestirimlerinin hesaplanmasını açıklamadan önce, yeniden örnekleme kavramını kullanarak gücün altında yatan temel anlamı inceleyebiliriz. Önceki bölümde, Aronson ve arkadaşları (1998) tarafından klişe tehdidi üzerine yapılan çalışmadan söz edilmişti. Bu yazarlar, matematikteki performanslarının önemli olduğu vurgulandığında, beyaz erkek öğrencilerin başarılarının, Asyalı erkek öğrencilere matematikte daha başarılı oldukları hatırlatıldığında beyaz erkek öğrencilerden matematikte daha başarılı olduğunu göstermiştir. Bu önemli bir bulgudur ve bu çalışmayı tekrarlamayı denemek sıra dışı olmaz. Her grupta –bir kontrol grubu ve bir tehdit altındaki grup- 20 öğrencinin yer aldığı bir tekrar çalışması yapmayı planladığımızı düşünelim. Fakat bu çalışmayı tekrarlamak için para ve enerji harcamadan önce başarılı bir tekrarın olasılığı için fikir sahibi olmamız gerekir. İşte bu, gücün ne olduğu ile ilgilidir.

Güç hesaplamasına göz atmak için, kontrol ve tehdit altındaki gruplara ait evrenlerin ortalama ve standart sapma değerlerinin ne olması gerektiğine dair fikir sahibi olmaya ihtiyacımız vardır. Bu parametrelere yönelik en iyi tahminimiz, Aronson ve arkadaşlarının kendi örneklemleri için buldukları ortalamalar ve standart sapmalar olabilir. Bunlar büyük olasılıkla kesin değerler değil, fakat elimizdeki en iyi kestirimlerdir. Bundan dolayı Kontrol evreninin ortalamasının 9,64 ve standart sapmasının 3,17; Tehdit evreninin ortalamasının 6,58 ve standart sapmasının 3,03 olduğunu varsayabiliriz. Ayrıca evrenlerin normal olarak dağıldığını varsayacağız. Söylediğim gibi her grupta 20 katılımcının olmasını planlıyoruz.

Bunu modellemenin kolay bir yolu, ortalama ve standart sapmanın kontrol grubundakiyle aynı olduğu durumda eldeki puanların oluşturduğu bir evrenden 20 gözlem çekmektir. Benzer şekilde ortalaması ve standart sapması deney grubununkiyle aynı olan bir evrenden 20 gözlem çekeceğiz. Daha sonra bu veri için *t* istatistiğini hesaplayacağız, bunu bir kenarda saklayacağız ve aynı işlemi 9.999 kez tekrarlayacağız. Bu bize 10.000 *t* değeri verecektir. Ayrıca 38 *sd* ile $t_a(38)$ kritik değerinin 2,024 olduğunu biliyoruz. Dolayısıyla 10.000 *t* değerinden kaç tanesinin manidar olduğunu sorabiliriz (yani 2,024'e eşit veya ondan büyük olduğunu).

Böyle bir örnekleme çalışmasının sonucu uygun *t* dağılımı ile birlikte Şekil 8.1'de gösterilmiştir. Sonuçların %86'sı *t* = 2,024'ten büyük, %14'ü kritik değerden küçüktür. Sonuç olarak mevcut parametre kestirimleri ve örneklem büyüklüğü dikkate alındığında bu deneyin gücü 0,86'dır. Bu değer, kritik değeri geçen sonuçların yüzdesidir. Aslında bu değer, birçok pratik çalışma için makul düzeyde bir güç değeridir.[3]




**Şekil 8.1** Tekrar örnekleme çalışmasında elde edilen *t* değerleri dağılımı

3 Bu şekli çizmek için bir program ResampleForPower.R. Web sitesinde mevcuttur. Parametreleri ve örneklem büyüklüğünü değiştirebilirsiniz. Bu program ayrıca gerçek gücü de hesaplar ki bu değer 0,8601'dir.

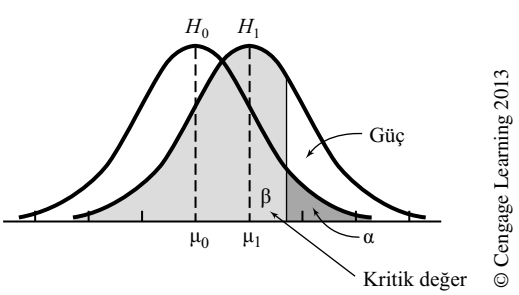
## 8.2 Bir Testin Gücünü Etkileyen Faktörler

Tekrarlı örneklemler çekmek suretiyle güce sezgisel bir yaklaşımla baktıktan sonra, şimdi biraz daha teorik yönden bakacağız. Bekleneceği gibi, güç çok değişkenli bir fonksiyondur. Güç, (1) $\alpha$, I. Tip hata olasılığının, (2) doğru alternatif hipotezin ($H_1$), (3) örneklem büyüklüğünün ve (4) kullanılacak özel testin yer aldığı bir fonksiyondur. Bağımsız örneklemlerin eşleştirilmiş örneklemlere karşı bağıl güçleri hariç tutulduğunda ve test varsayımlarının karşılandığı durumda son ilişkiyi düşünmeyeceğiz. Zira bu kitapta konuşulan işlemlerin büyük çoğunluğunun, üzerinde çalışılan soruyu cevaplamak için mevcut testler arasında en güçlüleri olduğu gösterilebilir. Bununla birlikte bir testin temel varsayımları ihlâl edildiğinde Bölüm 18'de yer alan parametrik olmayan testlerin ve özellikle tekrar örnekleme testlerinin genellikle daha güçlü olduğunu unutmamanız önemlidir.

### Kısa Bir Tekrar

Öncelikle Ünite 4'te konuşulanları hızlıca tekrar etmeye ihtiyacımız vardır. Şekil 8.2'deki iki dağılımı düşünün. Soldaki dağılım ($H_0$ ile etiketlenen), sıfır hipotezinin doğru olduğu durumda ($\mu = \mu_0$) ortalamanın örnekleme dağılımını gösterir. Sağdaki dağılım, eğer $H_0$ yanlış ve gerçek evren ortalaması $\mu_1$'e eşitse, bu durumda sahip olacağımız ortalamanın örnekleme dağılımını belirtir. Bu dağılımın yerleştirilmesi tamamen $\mu_1$ değerinin ne olabileceğine bağlıdır.

Tek kuyruklu test kullandığımızı varsaydığımızda $H_0$ dağılımının koyu dolgulu sağ kuyruğu, I. Tip hata olaslılığı olan a'yı göstermektedir (aksi durumda $\alpha$/2'yi gösterir). Bu alan, t'nin manidar değerleri ile sonuçlanan örneklem ortalamalarını içerir. İkinci dağılım ($H_1$), $H_0$'ın yanlış ve gerçek ortalamanın $\mu_1$ olduğu durumda ortalamanın örnekleme dağılımını göstermektedir. $H_0$'ın yanlış olduğu zaman bile örneklem ortalamalarının çoğu (ve dolayısıyla bunlara karşılık gelen t değerleri) yine de kritik değerin soluna düşecektir. Bu durum bizi yanlış bir $H_0$'ı reddetmek için başarısız kılacak ve II. Tip hataya götürecektir. Bunu önceki gösterimde gördük. Bu hatanın olasılığı, Şekil 8.2'de hafif dolgulanmış alanla gösterilmiş ve $\beta$ ile etiketlenmiştir. $H_0$ yanlış olduğunda ve test istatistiği kritik değerin sağına düştüğünde yanlış bir $H_0$'ı doğru olarak reddedeceğiz. Bunu yapmanın olasılığı güç olarak adlandırılır ve $H_1$ dağılımında dolgusuz alan olarak gösterilmiştir. Şekil 8.1'de $H_0$ dağılımını göstermemiş olmama rağmen Şekil 8.2, Şekil 8.1'e oldukça benzemektedir.




**Şekil 8.2** $H_0$ ve $H_1$ altında $\bar{X}$'nın örnekleme dağılımı

## $\alpha$'nın Bir Fonksiyonu Olarak Güç

Şekil 8.2'nin yardımıyla gücün, $\alpha$'nın neden bir fonksiyonu olduğunu söylemek kolaydır. Eğer $\alpha$'yı artırmak istersek kesim noktamız sola doğru hareket eder. Böylece I. Tip hata olasılığında artışa karşılık gelmesine rağmen eş zamanlı olarak $\beta$ azalır ve güç artar.

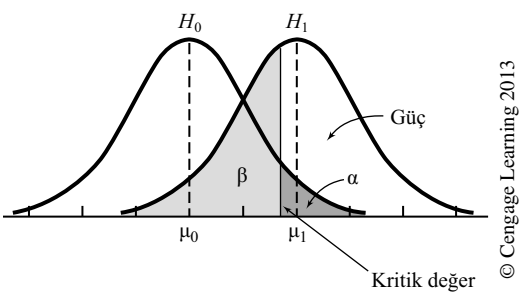
## $H_1$'in Bir Fonksiyonu Olarak Güç

Gücün doğru alternatif hipotezin bir fonksiyonu olduğu gerçeği [daha açık bir ifadeyle ($\mu_0 - \mu_1$), yani $\mu_0$ ($H_0$ altındaki ortalama) ve $\mu_1$ ($H_1$ altındaki ortalama) arasındaki fark] Şekil 8.2 ve 8.3'ün karşılaştırılması ile gösterilmiştir. Şekil 8.3'te $\mu_0$ ve $\mu_1$ arasındaki fark artmıştır ve bu durum II. Tip hata olasılığının hâlâ büyük olmasına rağmen güçte önemli bir artış olarak sonuçlanmaktadır. Bu şaşırtıcı değildir, çünkü bir fark bulma şansı gerçek farkın ne kadar büyük olduğuna bağlıdır.

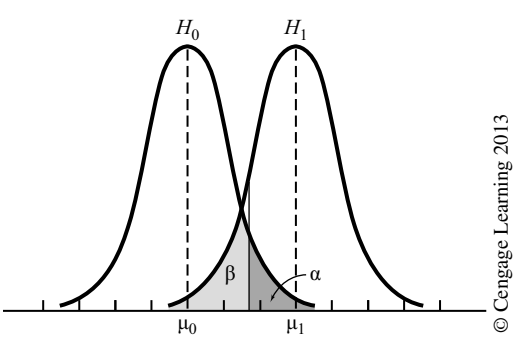
## $n$ ve $\sigma^2$'nin Bir Fonksiyonu Olarak Güç

Güç ve örneklem büyüklüğü arasındaki ilişki (ve güç ile $\sigma^2$ arasındaki) biraz gizlidir. Çünkü ortalamalar veya ortalamalar arasındaki farklarla ve ortalamanın örneklem dağılımı ile ilgilenmekteyiz. Ortalamanın örneklem dağılımının varyansının $n$ arttıkça veya $\sigma^2$ azaldıkça azaldığını biliyoruz çünkü $\sigma^2_{\bar{X}} = \sigma^2/n$. Şekil 8.4, Şekil 8.3 ile kıyaslandığında $n$'yi artırdıkça veya $\sigma^2$'yi azalttıkça iki örneklem dağılımına ($H_0$ ve $H_1$) ne olduğunu göstermektedir. Şekil 8.4 ayrıca $\sigma^2_{\bar{X}}$ azaldıkça iki dağılım arasındaki örtüşmenin azaldığını belirtmektedir. Bu durum güçte artış ile sonuçlanmaktadır. Dikkat edilirse iki ortalama ($\mu_0$ ve $\mu_1$) Şekil 8.3'e göre değişmeden aynı kalmaktadır.

Eğer bir uygulayıcı bir testin gücü ile ilgileniyorsa, güce hükmeden ve kolaylıkla yönetilebilen bu değişkenlerle ilgilenmelidir. $n$, $\sigma^2$ veya farktan ($\mu_0 - \mu_1$) ve daha kolay yönetilebildiğinden ve $\alpha$ üzerinde yapılacak değişiklik I. Tip hata olasılığını artırma yönünde olumsuz etkide bulunacağından, güç tartışmaları genellikle örneklem büyüklüğü değişiminin etkileri üzerine olmaktadır. Bununla birlikte McClelland (1997), deneysel tasarımdaki değişimlerin bir deneyin gücünü artırabildiğini ifade etmektedir.




**Şekil 8.3** $\mu_0 - \mu_1$ farkının $\beta$ üzerine etkisi




**Şekil 8.4** Ortalamanın standart hatasındaki azalmanın $\beta$ üzerine etkisi

## 8.3 Gücün Geleneksel Yolla Hesaplanması

Şekil 8.2'den 8.4'e kadar gördüğümüz gibi, güç $H_0$ ve $H_1$ altındaki örneklem dağılımları arasındaki örtüşmenin derecesine bağlıdır. Ayrıca bu örtüşme, $\mu_0$ ile $\mu_1$ arasındaki uzaklığın ve standart hatanın bir fonksiyonudur. $H_0$'ın ne ölçüde yanlış olduğunun bir ölçüsü, $\mu_1$'den $\mu_0$'a olan uzaklığın kaç standart hata büyüklüğünde olacağıdır. Belli bir n değerine karşılık gelen gücü hesaplamak veya belli bir güç değeri için gerekli olan n değerini bulmak istediğimizde yukarıda açıklanan ölçümün örneklem büyüklüğünü içermesi (standart hatanın hesaplanması sırasında) bir sorun olabilir. Bu nedenle uzaklık ölçümüzü veya **etki büyüklüğümüzü** (*d*), işaretini ihmal ederek ve *n*'yi daha sonra katarak

**etki büyüklüğü *(d)***

$$d = \frac{\mu_1 - \mu_0}{\sigma}$$

olarak alacağız.

Böylece d, iki evrenin standart sapması kullanılarak bulunan $\mu_1$ ve $\mu_0$ arasındaki farkın ne ölçüde olduğunun, ana evrenin standart sapması cinsinden bir göstergesi olmaktadır. (İncelediğimiz örnek için $d$ = (9,64 - 6,58)/3,10 = ,987'dir.) Bu sonuç H1 altındaki ortalamaların yaklaşık bir standart sapma kadar farklılaştığını söylemektedir. (Burada birleşik standart sapmayı kullanıyorum.) *d*'nin *n*'den bağımsız olarak sadece $\mu_1$, $\mu_0$ ve $\sigma$ kullanılarak hesaplandığını görüyoruz. Ünitede 7'de etki büyüklüğünü iki ortalama arasındaki standartlaştırılmış fark olarak açıklamıştık. Ortalamalardan birinin sıfır hipotezi altındaki ortalama olmasına rağmen, buradaki de aslında aynı ölçümdür. İki evrenin ortalamasını karşılaştırma konusuna geldiğimizde bu noktaya tekrar temas edeceğim.

### Etki Büyüklüğünün Kestirilmesi

İlerideki hesaplamalar için bu bir temel oluşturacağından ilk iş *d*'yi hesaplamaktır. Bu hesaplama üç yol ile yapılabilir. Bunlar en yararlısından başlayarak aşağıda açıklanmaktadır:

1. *Önceki araştırmalar.* Önceki araştırmalar esas alınarak d için en azından kabaca bir tahmin yapabiliriz. Böylece diğer çalışmaların örneklem ortalamaları ve varyanslarına bakabilir ve $\mu_1$ - $\mu_0$ ile $\sigma$ için bilinçli bir tahmin yapabiliriz. Özellikle kabaca bir yaklaşımın hiçbir yaklaşımın olmadığı duruma göre daha iyi olduğunu kavradığımızda, pratikte bu işin görüldüğü gibi zor olmadığı anlaşılacaktır.
2. *Ne büyüklükte bir farkın önemli olduğuna dair bireysel değerlendirme*. Pek çok durumda bir araştırmacı, $\mu_1$ ile $\mu_0$ arasında en az 10 puanlık bir fark yakalamakla ilgileniyor olabilir. Bu araştırmacı aslında bundan daha az farkların önemli veya faydalı bir anlamı olmadığını, daha büyük farkların anlamlı olduğunu söylemektedir. (Bu durum özellikle biomedikal araştırmalarında yaygındır. Örneğin, böyle bir araştırmada kolesterolü belli bir değere indirmeye uğraştığımızda küçük değişimlerin bizim için önemi yoktur. Benzer bir durum ilâçları karşılaştırmak istediğimizde de ortaya çıkar. Önceden belirlenen bir miktar kadar yeni ilâç eski ilâçtan daha iyi değilse yeni ilâçla ilgilenmeyiz.) Burada bize $\mu_1$- $\mu_0$ değeri doğrudan verilmiştir. $\mu_1$ ve $\mu_0$'ın gerçek değerlerini bilmeye ihtiyacımız yoktur. Geriye sadece $\sigma$'yı hesaplamak kalmaktadır. Örnek olarak bir araştırmacı Yüksek Lisans Kayıt Sınavı puanlarını normalden 40 puan artıracak bir işlemle ilgilendiğini söyleyebilir. Bu test için standart sapmanın 100 olduğunu biliyoruz. Böylece d = 40/100 = ,40 elde ederiz. Eğer bu araştırmacı standart hatanın onda dördü kadar puanları artırmak istediğini söyleseydi, bize *d*'yi doğrudan vermiş olacaktı.
3. *Özel Uzlaşımları Kullanma*. Gerekli parametreleri hesaplayacak hiçbir yolun olmadığı bir durumla karşılaştığımızda son çare olarak Cohen (1988) tarafından önerilen uzlaşımlara başvurabiliriz. Cohen, *d* için yaklaşık olarak ve gelişigüzel üç düzey belirlemiştir:

| Etki Büyüklüğü | *d* | Örtüşmenin Yüzdesi |
|---|---|---|
| Küçük | ,20 | ,92 |
| Orta | ,50 | ,80 |
| Büyük | ,80 | ,69 |

Böylece, *ihtiyaç duyduğunda* araştırmacı, sadece küçük, orta veya büyük etki olup olmayacağına karar verir ve buna göre *d*'yi tayin edebilir. Bununla birlikte bu çözüm, sadece diğer seçenekler uygulanabilir olmadığı durumda seçilmelidir. Tablonun en sağ tarafı Örtüşmenin Yüzdesi olarak adlandırılmış olup Şekil 8.2'de gösterilen iki dağılımın ne dereceye kadar örtüştüğünü belirtir.[4] Böylece örneğin $d = 0{,}50$ olduğunda iki dağılımın %80'i örtüşür (Cohen, 1988). Bu, uygulanan yeni bir yöntemin ne büyüklükte bir fark oluşturacağını düşünmenin farklı bir yoludur.

Cohen, orta etkiyi zeki bir gözlemciye görünebilen; küçük etkiyi gerçek fakat gözle zor algılanabilen; büyük etkiyi ise küçük etki ile orta etki arasındaki uzaklık kadar orta etkinin üzerinde yer alan olarak tanımlar. Cohen (1969), bu kılavuz ilkeleri, etki büyüklüğünü hesaplamak için hiçbir yolun olmadığı durumlar için geliştirmiştir. Fakat zaman geçtikçe ve birçok araştırmacının güç analizi yaparken (büyük olasılıkla bu analizleri yapmanın zor olduğunu düşündüklerinden) başarısızlığa uğramasından dolayı cesareti kırıldıkça, bu uzlaşımların daha fazla kullanılabilirliğini sağlamaya çalışmıştır (bakınız Cohen, 1992a). Bununla birlikte Texas A&M'den Bruce Thompson bu konuda önemli bir noktaya vurgu yapmıştır. O, sonuçlanan test istatistiğinin olasılık değerine odaklanmak yerine elde edilen farkları *d* cinsinden açıklamayı yeğlemiştir. Bunu şu şekilde ifade etmiştir: "Sonuç olarak şu vurgulanmalıdır ki eğer biz ,05 ve ,01 gibi *p* değeri kesim noktalarına düşüncesizce danışmak yerine yine aynı düşüncesizlikle Cohen'ın temel kuralına onun güçlü uyarılarına aykırı olarak başvurursak sadece yeni bir metrikte düşüncesiz olmayı seçeriz (açıklama eklenmiştir)" (Thompson, 2000). Bu husus, *d* için herhangi keyfi bir uzlaşımın güç hesaplanmasında veya elde ettiğiniz farkın büyüklüğü ile okuyucuları etkileme amacı için kullanılıp kullanılmadığına bakılmaksızın da geçerlidir. Lenth (2001), Cohen'ınki gibi uzlaşımların kullanılmasının tehlikeli olduğunu ikna edici bir şekilde tartışmaktadır. Sadece d'nin pay ve paydasının oranına değil, *d*'nin pay ve paydasındaki değerlerine yoğunlaşmaya ihtiyacımız vardır. Lenth'in argümanı gerçekten araştırmacının kararlarında daha sorumlu olmasını sağlamak için bir girişimdir ve zannediyorum ki Cohen'ın samimiyetle buna karar vermiş olabileceğine dair şüphe duymaktayım.

Deneysel işlem gerçekleştirilmeden önce, araştırmacıdan beklediği farkı açıklamasının istenmesi sizde tuhaf bir izlenim bırakabilir. Birçok kişi "Deneysel işlemin nasıl sonuçlanacağını bilmiyorum. Ben sadece bir fark olup olmayacağını merak ediyorum." şeklinde cevap verebilir. Birçok araştırmacı bu şekilde konuşurken (ve ben de erdemli bir istisna değilim) bu ifadenin geçerliğini sorgulamalısınız. Biz deneylerimizde en azından belli belirsiz bir şekilde ne olacağını gerçekten bilmiyor muyuz? Eğer bilmiyorsak neden bu deneyleri yapıyoruz? "Neler olabileceğini merak ederim, eğer deney..." ifadesi nadiren de olsa meşru görünse bile genellikle "Bilmiyorum." ifadesi "Bunun ilerisini düşünmedim." olarak tercüme edilir.

## Etki Büyüklüğü ve *n*'yi Yeniden Birleştirmek

*n*'yi ayrıca ele almayı kolaylaştırmak için önceden örneklem büyüklüğünü etki büyüklüğünden ayırmaya karar vermiştik. Şimdi etki büyüklüğü ile örneklem büyüklüğünü birleştirecek bir yönteme ihtiyacımız var. Bu birleşimi göstermek için $\delta$**(delta)** = $d[f(\mathrm{n})]$[5] istatistiğini kullanırız. Burada n'nin özel fonksiyonu [yani f(n)] her bireysel test için farklı tanımlanacaktır. Bu sistemin elverişli yanı, bize güç hesaplamaları için tüm isatistiksel işlemlerde aynı d tablosunu kullanmaya imkân tanımasıdır.

**$\delta$(delta)**

---

4 James Grice ve Paul Barrett'a örtüşme yüzdesinin düzeltilmiş değerlerini sağladığı için teşekkür etmek isterim.

5 Matematik fobisi olan bir öğrenciyi "*f(n)*"yi okumaktan daha hızlı sıkacak tek şey, sonraki cümlede "*g(n)*"nin gelmesidir. Bu ifadeler, henüz belli olmayan bir şekilde *d*'nin *n*'ye bağlı olduğunu göstermektedir.

## 8.4 Tek-Örneklem *t* İçin Güç Hesaplamaları

İlk olarak tek örneklem t testi için güç hesaplamalarını inceleyeceğiz. Önceki bölümde $\delta$'nın *d*'ye ve *n*'nin bir fonksiyonuna bağlı olduğunu gördük. Tek-örneklem t için bu fonksiyon $\sqrt{n}$ olacak ve $\delta = d\sqrt{n}$ olarak tanımlanacaktır. Burada tanımlanan $\delta$ bilindiğinde Ekler Bölümü'ndeki Güç başlıklı tablo kullanılarak testimizin gücünü hemen belirleyebiliriz.

Bir klinik psikoloğun, psikolojik problemleri için tedavi arayan insanların IQ'larının genel evrenden daha yüksek olduğu hipotezini test etmek istediğini varsayalım. Araştırmacı tesadüfî olarak seçilen 25 hastanın IQ'larını kullanmak ve genel evrenin ortalaması ile hastaların seçildiği evrenin ortalaması arasında 5 puanlık bir farkı yakalamak amacıyla gerekli olan gücü bulmak için basit hesaplama kullanmak istemektedir. Böylece $\mu_1 = 105$, $\mu_0 = 100$ ve $\sigma = 15$ olur. Biliyoruz ki

$$d = \frac{105 - 100}{15} = 0,33$$

buradan

$$\delta = d\sqrt{n} = 0,33\sqrt{25} = 0,33(5)$$
$$= 1,65$$

bulunur.

Klinik uzmanları, örneklem ortalamalarının, ortalamanın üzerinde olmasını beklemesine rağmen, beklenmeyen durumlara karşı koruma amaçlı olmak üzere $\alpha = 0,05$ düzeyinde çift-kuyruklu test kullanmayı planlamaktadır. Ekteki Güç tablosundan $\alpha = 0,05$ manidarlık düzeyi ve $\delta = 1,65$ (çift kuyruklu) için güç 0,36 ile 0,40 arasında bulunur. Buradan doğrusal dönüşüm yoluyla gücün 0,38 olduğunu söyleyebiliriz. Bu şu anlama gelmektedir: Eğer $H_0$ yanlış ve μ1 gerçekten 105 ise klinik uzmanı, kendi örneklem ortalaması ve $H_0$ ile belirtilen ortalama arasında "istatistiksel olarak manidar " bir fark bulmayı bekleyebilir. G*Power yazılımını kullandığımızda ,359 sonucunu elde edeceğiz. Bu iki yaklaşımın büyük ölçüde aynı sonucu verdiğini görebilirsiniz. 0,38 oranı oldukça cesaret kırıcıdır; çünkü bu oran, eğer doğru ortalama gerçekten 105 ise klinik uzmanının %62 oranında II. Tip hata yapacağı anlamına gelir.

Araştırmacımızın deneysel işleme başlamadan önce güç sorununu incelemesi için her şey bitmiş değildir. Gücü artırmayı sağlayacak değişiklikleri yapmak için hâlâ bir şansı vardır. Örneğin, $\alpha$'yı 0,10 yaparak gücü yaklaşık 0,50'ye kadar çıkarabilir. Fakat bu işlem büyük olasılıkla başarısızdır. (Örneğin hakemler 0,05'ten daha büyük $\alpha$ değerlerini görmekten hoşlanmazlar.)

### Gerekli Örneklem Büyüklüğünü Kestirme

Alternatif olarak araştırmacı örneklem büyüklüğünü artırabilir, böylece güç artmış olur. Araştırmacımız ne büyüklükte bir n'ye ihtiyaç duyar? Cevap, araştırmacının hangi düzeyde güç istediğine bağlıdır. Araştırmacının gücü 0,80 istediğini varsayalım. Ekteki Güç tablosundan güç = 0,80 ve $\alpha = 0,05$ için $\delta = 2,80$ olmalıdır. Böylece $\delta$ ve *d*'ye sahip olduğumuzdan kolaylıkla n'yi aşağıdaki gibi hesaplayabiliriz:

$$\delta = d\sqrt{n}$$
$$n = \left(\frac{\delta}{d}\right)^2 = \left(\frac{2,80}{0,33}\right)^2 = 8,48^2$$
$$= 71,91$$

Hastalar genellikle tam sayı olarak ifade edildiğinden, bu değeri 72'ye yuvarlayacağız. Böylece eğer araştırmacı, $d = 0{,}33$ olduğunda H0'ı reddetmek için %80 oranında bir şansa sahip olmak istiyorsa (yani $\mu_1 = 105$ iken), tesadüfî olarak seçilen 72 hastanın IQ'larını kullanmak zorundadır. Bu sayı, araştırmacının kolaylıkla test edebileceğinden daha fazla hasta anlamına gelmesine rağmen buna alternatif, sadece daha düşük güç düzeyi belirlemek veya başka bir klinik psikologdan yardım alarak veri tabanına katkıda bulunmasını sağlamaktır. Neden gücü 0,80 seçtiğimizi merak edebilirsiniz. Çünkü bu güç değeri ile hâlâ %20 oranında II. Tip hata yapma olasılığına sahibiz. Bu elbette bir sorundur. Örneğin, gücün 0,95 olmasını isteyelim. Birkaç basit hesaplama ile örneklem büyüklüğünün $n = 119$ olduğu bulunur. Güç = 0,99 için yaklaşık 162 hastaya ihtiyacınız olacaktır. Bu değerler, bu özel deney durumu veya araştırmacının kaynakları düşünüldüğünde makul görünmemektedir. Güçteki artışın genellikle n'deki artış ile sağlandığını hatırlayın. Yüksek güç değerlerinde ise maliyet çok yüksek olabilir. Eğer veri kümesini Nüfus Bürosu tarafından sağlanan kayıtlardan alıyorsanız bu durum mezun olan lise öğrencileri ile çalışmaktan çok daha farklıdır. 0,80 güç değeri II. Tip hatayı, I. Tip hatanın dört katı yapmaktadır. Bu durum hataların bağıl önemini yansıtan bir durum olarak kullanılabilir. Ayrıca şunu vurgulamalıyım, Etik Kurulları fazla olduğunu düşündüğü örneklem büyüklüklerine izin vermez.

## Merkezî Olmayan Parametreler

**merkezî olmayan parametre**

$\delta$ istatistiğimiz birçok ders kitabında **merkezî olmayan parametre (moncentralits parameter)** olarak adlandırılır. Bu kavram oldukça basittir ve dikkate alınmaya değerdir. (Bazı bilgisayar yazılımları bir merkezî olmayan parametre girmenizi isteyecektir.) Öncelikle biliyoruz ki

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

$\mu$, X'lerin örneklendiği dağılımın gerçek ortalaması olduğu sürece herhangi bir sıfır hipotezinin doğru veya yanlışlığına bakılmaksızın t, sıfır etrafında dağılır. Eğer $H_0$, hipotezi $H_0$, $\mu = \mu_0$ ($\mu$'nun bazı özel değerleri) biçiminde ifade edildiyse ve $H_0$ doğruysa,

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

yine sıfır etrafında dağılacaktır. Eğer $H_0$ yanlış ve $\mu \neq \mu_0$ ise

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

sıfır etrafında dağılmayacaktır, çünkü $\mu_0$'ı çıkarırken yanlış evren ortalamasını çıkarmış oluyoruz. Aslında *t* dağılımı aşağıdaki noktada ortalanacaktır:

$$\delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$$

Bu, dağılımın ortalamasını 0'dan $\delta$'ya kaydırır ve merkezî olmama derecesi olarak adlandırılır; $\delta$ merkezî olmayan bir parametredir ve genellikle mop (ncp) ile gösterilir. ($\mu_1 = \mu_0$ olduğunda $\delta$ nedir?) Merkezi olmayan parametre, sıfır hipotezinin nasıl yanlış olduğunu açıklamanın sadece bir yoludur.

Güç sorusu, "$H_0$ altında *t*'nin kritik değerinden daha büyük ve merkezî olmayan bir dağılım değerini nasıl buluruz?" sorusuna dönüşmektedir. Diğer bir ifadeyle $H_0$'ın yanlış olmasından dolayı t'nin normal değerinden daha büyük değerler beklenmesine rağmen şans etkisi ile ara sıra küçük değerler elde edeceğiz. $\pm t_{.025}$ arasında gerçekleşen bu değerlerin oranı, II. Tip hatanın olasılığı olan $\beta$'dır. Bildiğimiz gibi $\beta$'yı kullanarak güce ulaşabiliriz. Güç = $1 - \beta$.

Cohen'ın katkısı merkezî olmayan parametreyi ($\delta$), örneklem büyüklüğü ve etki büyüklüğü olarak iki parçaya ayırmasında görülebilir. Bir parça ($d$) sadece evren parametrelerine bağlıyken, diğer parça örneklem büyüklüğüne bağlıdır. Böylece Cohen, üzerinde sınırlı işlem yapabileceğimiz parametrik faktörleri ($\mu_1$, $\mu_0$ ve $\sigma$), üzerinde daha fazla kontrolümüz olabilen örneklem özelliklerinden ($n$) ayırmıştır. Bu ayırma, ilgili teoride temel bir değişiklik oluşturmamasına rağmen, kavramın anlaşılmasını ve kullanılmasını kolaylaştırmaktadır.

## 8.5 Birbirinden Bağımsız İki Ortalama Arasındaki Fark İçin Güç Hesaplamaları

Bağımsız iki ortalama arasındaki farkı test etmek istediğimizde güç işlemleri, tek ortalama için kullandığımız işlemlere çok benzerdir. Kesim 8.3'te d'yi $H_1$ altındaki $\mu$ ve $H_0$ altındaki $\mu$ arasındaki farkı $\sigma$'ya bölerek elde ettik. İki bağımsız ortalama arasındaki farkı test ederken, ortalamalar arasındaki fark üzerinde çalışıyor olacağımıza rağmen temelde yine aynı şeyi yapacağız. Böylece, $H_1$ altında iki evren ortalaması arasındaki farkı ($\mu_1 - \mu_2$), $H_0$ altındaki farktan ($\mu_1 - \mu_2$) çıkarıp $\sigma$'ya bölmek istemekteyiz. ($\sigma_1^2 = \sigma_2^2 = \sigma^2$ varsaydığımızı hatırlayın.) Bununla birlikte klâsik uygulamaların çoğunda $H_0$ altındaki ($\mu_1 - \mu_2$) farkı sıfırdır; dolayısıyla bu terim formülümüzden düşer. Böylece,

$$d = \frac{(\mu_1 - \mu_2) - (0)}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma}$$

olur. Yukarıdaki eşitlikte pay, $H_1$ altında olması beklenen farkı ve payda evrenlerin standart sapmasını gösterir. Buradaki $d$'nin Bölüm 7'de gördüğünüz $d$ ile aynı olduğunu fark etmelisiniz. Bölüm 7'de $d$, Cohen'ın d'si veya bazen Hedges'in $g$'si olarak da adlandırılmıştı. Tek fark, örneklem ortalamaları yerine burada $d$'nin evren ortalamaları açısından ifade edilmesidir.

İki örneklem durumunda eşit $n$'ler içeren deneyler ile farklı $n$'ler içeren deneyleri ayırt etmeliyiz. Bu iki durumu ayrı ayrı ele alacağız.

### Eşit Örneklem Büyüklükleri

İki yöntem arasındaki farkı test etmek istediğimizi ve evren ortalamaları farkının yaklaşık 5 puan olmasını beklediğimizi veya en az 5 puanlık bir farkla ilgilendiğimizi varsayalım. Ayrıca geçmişteki veri kümelerinden $\sigma$'nın yaklaşık 10 olduğunu varsayalım. Buna göre,

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{5}{10} = 0{,}50$$

olur. Böylece, iki ortalamanın arasında bir standart sapmanın yarısı kadar bir fark bekliyoruz. Cohen (1988), bunu orta büyüklükte etki olarak adlandırmaktadır. İlk olarak iki grubun her birinde 25 gözlemin yer aldığı bir deneyin gücünü inceleyeceğiz.

İki örneklem durumunda merkezi olmama parametresini ($\delta$) aşağıdaki gibi tanımlayacağız:

$$\delta = d\sqrt{\frac{n}{2}}$$

Burada $n$ = *bir örneklemdeki* gözlem sayısıdır (bütünde $2n$ gözlem vardır). Böylece,

$$\delta = (0{,}50)\sqrt{\frac{25}{2}} = 0{,}50\sqrt{12{,}5} = 0{,}50(3{,}54)$$
$$= 1{,}77$$

buluruz.

Ek Güç'e bakıldığında çift kuyruklu testte $\alpha$ = ,05 ve $d$ = 1,77 için güç = ,43 bulunur. Böylece eğer araştırmacımız bu deneyi gerçekte 25 kişi ile yürütürse ve eğer $\delta$'nın kestirimi doğru ise, araştırmacımız $H_0$'ı gerçekte 0,43 olasılıkla reddedecektir; eğer $H_0$ yanlışsa beklentisi daha yüksektir (ve II. Tip hata yapma olasılığı 0,57).

"Güç = 0,80 için kaç kişiye ihtiyacımız olur?" sorusuna dönelim. Ek Güç'ten çift kuyruklu testte $\delta$ = 2,80 olduğu görülür.

$$\delta = d\sqrt{\frac{n}{2}}$$

$$\frac{\delta}{d} = \sqrt{\frac{n}{2}}$$

$$\left(\frac{\delta}{d}\right)^2 = \frac{n}{2}$$

$$n = 2\left(\frac{\delta}{d}\right)^2$$

$$= 2\left(\frac{2,80}{0,50}\right)^2 = 2(5,6)^2$$

$$= 62,72$$

Burada $n$, örneklem başına kişi sayısını temsil eder. Dolayısıyla güç = ,80 için örneklem başına 63 kişiye, toplamda ise 126 kişiye ihtiyacımız vardır.

## Eşit Olmayan Örneklem Büyüklükleri

Şimdiye kadar sadece $n_1 = n_2 = n$ durumu ile ilgilendik. Bununla birlikte deneyler genellikle farklı büyüklüklerde iki örneklem içerir. Elbette bu durum $\delta$'yı hesaplarken bir zorluk oluşturur, çünkü n için bir değere ihtiyacımız vardır. Hangi değeri kullanabiliriz?

Yeterince büyük ve birbirine çok yakın örneklem büyüklüklerinde $n_1$ ve $n_2$'den küçük olan $n$'i kabul ederek muhafazakar bir yaklaşım elde edebiliriz. Bununla birlikte eğer örneklem büyüklükleri küçük veya iki $n$ oldukça farklı ise bu yaklaşım tatmin edici değildir. Bu gibi durumlar için daha kesin bir çözüme ihtiyacımız vardır.

Mantıklı görünen (fakat doğru olmayan) bir yaklaşım, n'yi n1 ve n2'nin aritmetik ortalaması olarak almaktır. Bu yöntem, iki örneklemi eşit olarak ağırlıklandırır. Bununla birlikte, biliyoruz ki örneklemlerin varyansı n ile orantılı olmayıp 1/n ile orantılıdır. Bu ilişkiyi dikkate alan ölçüm, aritmetik ortalama olmayıp harmonik ortalamadır. k sayının ($X_1$, $X_2$, ...., $X_k$) **harmonik ortalaması** ($\overline{X}_h$) aşağıdaki gibi tanımlanır:

**harmonik ortalama** ($\overline{X}_h$)

$$\overline{X}_h = \frac{k}{\sum \frac{1}{X_i}}$$

Böylece iki örneklem büyüklüğü için ($n_1$ ve $n_2$),

$$\overline{n}_h = \frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2n_1n_2}{n_1 + n_2}$$

olur. Öyleyse $\delta$'nın hesaplanmasında $\overline{n}_h$'ı kullanabiliriz.

Bu ünitenin başında Aronson ve arkadaşları (1998) tarafından klişe söz tehditi üzerine yapılan çalışmadan söz etmiştik. Burada anılan çalışmaya geri döneceğiz, fakat doğrudan hesaplamaya odaklanacağız. Aronson, kontrol ve deney gruplarının ortalamalarını sırasıyla 9,58 ve 6,55 olarak bulmuştur. Bu bulgu, benim Bölüm 7'de oluşturduğum örnek veri-

den önemsiz derecede farklıdır. Bu ortalamaların birleşik standart hatası yaklaşık 3,10 idi. Aronson'un evren ortalamaları ve standart sapma kestirimlerinin temelde doğru olduğunu varsayacağız. (Bu değerler elbette bazı tesadüfî hatalardan etkilenir, fakat bu değerler, bu parametreler için elimizdeki en iyi tahminlerdir). Böylece;

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{9{,}58 - 6{,}55}{3{,}10} = \frac{3{,}03}{3{,}10} = 0{,}98$$

buluruz. Bu çalışmayı araştırma yöntemleri dersinde tekrarlamak istiyorum fakat akılsızca bir risk almak ve "çalışmalıydı" demek istemiyorum. Sınıfımda çok öğrenci var, fakat bunların yaklaşık 30'u erkektir ve laboratuvar gruplarına eşit olarak dağılmamışlardır. Bu deneyi yürütmek için seçtiğim yoldan dolayı Kontrol grubunda 18, Deney grubunda ise 12 erkek öğrenci bekleyebileceğimi varsayın. Bu durumda etkin örneklem büyüklüğünü ($\delta$'yı hesaplarken kullanılan örneklem büyüklüğü) aşağıdaki gibi hesaplayacağız:

$$\bar{n}_h = \frac{2(18)(12)}{18 + 12} = \frac{432}{30} = 14{,}40$$

**etkin örneklem büyüklüğü**

**Etkin örneklem büyüklüğünün** iki bireysel örneklem büyüklüğünün aritmetik ortalamasından daha az olduğunu görüyoruz. Diğer bir ifadeyle bu çalışma, grup başına 14,4 kişi (toplamda 28,8 kişi) ile yapılan deneyin sahip olacağı güç ile aynı güce sahiptir. Yani örneklem büyüklüklerinin aynı olduğu bir deneyde 28,8 kişi ile sahip olunan gücün eşiti farklı örneklem büyüklükleri olduğunda 30 kişi gerektirir.

Devam edersek,

$$\delta = d\sqrt{\frac{\bar{n}_h}{2}} = 0{,}98\sqrt{\frac{14{,}4}{2}} = 0{,}98\sqrt{7{,}2}$$
$$= 2{,}63$$

elde ederiz. $\delta = 2{,}63$ için, güç = 0,75; $\alpha = {,}05$ (çift kuyruklu).

Bu durumda güç, güven oluşturmak için biraz düşük kalmaktadır. Bir laboratuvar örneğinde beklenildiği gibi, risk alıp çalışmayı yapabilirim, fakat deney başarısız olduğunda problemli olabilir.

Buna bir çare olarak biraz daha öğrenci toplanabilir. Dersimdeki 30 erkek öğrenciyi kullanacağım fakat başka bir dersten katılmaya gönüllü 20 kişi daha bulabilirim. İki farklı sınıfı birleştirerek öğrencilerime kötü deneysel tasarım öğretiminin riski altında (Bunun bir problem olabileceğini söylemek en azından benim için bir özür olur ve mazeretimi sağlayabilir.) bu öğrencileri ekleyeceğim ve örneklem büyüklüklerinin 28 ve 22 olmasını bekleyeceğim.

Bu örneklem büyüklükleri ile $\bar{n}_h = 24{,}64$ olur. O halde, aşağıdaki elde edilir:

$$\delta = d\sqrt{\frac{\bar{n}_h}{2}} = 0{,}98\sqrt{\frac{24{,}64}{2}} = 0{,}98\sqrt{12{,}32}$$
$$= 3{,}44$$

Ek Güç'ten gücün yaklaşık olarak 0,93'e eşit olduğunu görüyoruz ki bu bizim amacımız için yeterlidir.

Örneklem büyüklüklerim eşit değildi fakat çok ciddi farklı da değildi. Birbirinden oldukça farklı örneklem büyüklüklerine sahip olduğumuz zaman ve bu durum kaçınılmazsa, küçük grup büyük grupla kıyaslandığında olabildiğince büyük olmalıdır. Örneklem büyüklüklerini eşit yapmak için denekleri çıkarmamalısınız. Bu, gücü azaltmak demektir.[6]

6 McClelland'ın (1997) ortaya koyduğu güçlü argümana göre ikiden fazla grubumuz varsa ve bağımsız değişken sıralama ölçeğinde ise güç, bağımsız değişkenin aşırı düzeylerine orantısız çok miktarda denekler atanarak artırılabilir.

## 8.6 Eşleştirilmiş-Örneklemler *t* Testi İçin Güç Hesaplamaları

Eşleştirilmiş iki örneklem arasındaki farkı test etmek istediğimizde, eldeki problem biraz daha zorlaşmakta ve ek bir parametrenin kullanılması gerekmektedir. Bu nedenle böyle durumlar için güç analizi çoğunlukla pratik değildir. Bununla birlikte bu soruna genel çözüm, deneysel tasarımın önemli bir ilkesini ve böylece derinlemesine incelemenin haklılığını gösterir.

Eşleştirilmiş örneklem *t* testinde *d* aşağıdaki gibi tanımlanır:

$$d = \frac{\mu_1 - \mu_2}{\sigma_{X_1 - X_2}}$$

Yukarıdaki eşitlikte $\mu_1 - \mu_2$ gözlemlere ait iki evrenin ortalamalarındaki beklenen farkı gösterir (fark puanlarının beklenen ortalaması). Buradaki problem $\sigma_{x_1 - x_2}$ standart sapmasının $X_1$ ve $X_2$'ye ait evrenlerin standart sapmalarının değil, bu evrenlerden çekilen fark puanlarının standart sapması olmasından kaynaklanır. $\sigma_{x_1}$ ve $\sigma_{X_2}$ için zekice bir tahmin yapabilmemize rağmen $\sigma_{x_1 - x_2}$ için muhtemelen bir fikrimiz yoktur.

Bununla birlikte her şey kaybedilmiş değildir. Birkaç varsayım altında $\sigma_{x_1 - x_2}$'yi hesaplamak mümkündür. Varyans toplam yasası (Bölüm 7, s. 207'de açıklanmıştı.) iki değişkenin toplamı veya farkı için varyans,

$$\sigma^2_{X_1 \pm X_2} = \sigma^2_{X_1} + \sigma^2_{X_2} \pm 2\rho\sigma_{X_1}\sigma_{X_2}$$

eşitlikleriyle ifade edilmişti.

İki değişkenin farkı için varyansların homojenliğiyle ilgili genel varsayımı ele alırsak aşağıdaki eşitlikleri yazabiliriz:

$$\sigma^2_{X_1 - X_2} = 2\sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1 - \rho)$$
$$\sigma_{X_1 - X_2} = \sigma\sqrt{2(1 - \rho)}$$

Yukarıdaki eşitliklerde $\rho$ (ro), evrende $X_1$ ve $X_2$ arasındaki evren korelasyonu olup 1 ile –1 arasında değerler alır. Hemen hemen tüm durumlar için bu korelasyon pozitif olup muhtemelen eşleştirilmiş örneklem *t*'nin de bu durumlar içinde olmasını isteriz.

Bir an için $\rho$'yu kestirebildiğimizi düşünelim, geri kalan işlemler tek örneklem t testi için yapılanların aynısıdır. d ve $\delta$'yı

$$d = \frac{\mu_1 - \mu_2}{\sigma_{X_1 - X_2}}$$

ve

$$\delta = d\sqrt{n}$$

eşitlikleriyle hesaplarız. Daha sonra $\sigma_{X_1 - X_2}$'yi $\sigma\sqrt{2(1 - \rho)}$ olarak hesaplarız ve $\delta$'nın değerini tablodan buluruz.

Örnek olarak Aronson'ın klişe tehdidi çalışmasını kullanmak istediğimi varsayın. Fakat bu sefer çalışmayı eşleştirilmiş örneklem tasarımı ile yapacağım. Elimde 30 erkek öğrenci vardır. Önce Asyalı öğrencilerin daha başarılı olduklarına dair hiçbir şey söylemeden testi uygulayabilirim. Sonraki haftanın laboratuvarında tehdit durumunu açıklayarak testi tekrar uygulayayım. (Taşınma etkisi ve diğer olumsuz durumları en aza indirmek için bu çalışmanın nasıl geliştirilebileceğini düşünerek daha iyisini yapabilirsiniz). Tehdit açıklamasının yapıldığı öğretimde puanların düşük olacağını beklediğimizi varsayalım. Fakat bu test ilk hafta aynı kişilere uygulandığından düşüş 9,58'den sadece 7,55'e olacaktır. Standart sapmanın 3,10'da aynı kalacağını varsayalım. Ortalamalar arasındaki farkın standart hatasını bulmak için iki sınav puanları arasındaki korelasyona ihtiyacımız vardır.

Fakat burada şanslıyız. Çünkü Aronson'ın matematik soruları Yükseklisans Giriş Sınavı'na yönelik bir deneme sınavından alınmıştır ve aradığımız korelasyon bu sınavın test-tekrar test güvenirliği ile hesaplanmıştır. Bu sınavın güvenirliğinin 0,92 civarında olacağına dair oldukça iyi bir fikre sahibiz. O hâlde,

$$\sigma_{X_1 - X_2} = \sigma\sqrt{2(1 - \rho)} = 3.10\sqrt{2(1 - .92)} = 3.1\sqrt{2(.08)}$$
$$= 1.24$$
$$d = \frac{\mu_1 - \mu_2}{\sigma_{X_1 - X_2}} = \frac{9.58 - 7.55}{1.24} = 1.64$$
$$\delta = d\sqrt{n} = 1.64\sqrt{30} = 8.97$$
$$\text{Güç} = ,99$$

İlk laboratuvar alıştırmamdan daha küçük bir etki büyüklüğüne sahip olduğuma dikkat ediniz. Çünkü dürüst olmaya çalıştım ve deneysel işlemlerden dolayı ortalamalardaki farkın düşük olabileceğini kestirebildim. Bununla birlikte eşleştirilmiş örneklem tasarımından gelen ek güçten dolayı, yeni gücüm orijinal örneğimdekinden daha büyüktür.

Diğer taraftan daha az güvenilir ve $\rho = 0{,}40$ olan bir test kullandığımızı varsayın. $\sigma$'nın değişmediğini ve ortalamalar arasında 2,03 birimlik bir fark beklediğimizi varsayalım. O zaman,

$$\sigma_{X_1 - X_2} = 3.10\sqrt{2(1 - .40)} = 3.10\sqrt{2(.60)} = 3.10\sqrt{1.2} = 3.40$$
$$d = \frac{\mu_1 - \mu_2}{\sigma_{X_1 - X_2}} = \frac{2.03}{3.40} = 0.60$$
$$\delta = 0.60\sqrt{30} = 3.29$$
$$\text{Güç} = ,91$$

$\rho$ düştükçe gücün de düştüğünü görüyoruz. (Güç hâlâ tatmin edici düzeyde fakat olması gerektiğinden çok daha küçüktür). $\rho = 0$ olduğunda iki değişkenimiz ilişkili değildir. Böylece eşleştirilmiş örneklem durumu neredeyse bağımsız örneklem durumuna benzemektedir. Burada önemli nokta şudur: Pratik amaçlar için düşünüldüğünde eşleştirilmiş örneklem durumunda minimum güç, $\rho = 0$ olduğunda gerçekleşir ve böylece bağımsız örneklemlere sahip oluruz. Böylece eşleştirilmiş örneklemler kullanılmasının oldukça uzak olduğu durumlarda bile ($X_1$ ve $X_2$ arasında pozitif korelasyon beklediğimiz zaman) eşleştirilmiş örneklem tasarımı buna karşılık gelen bağımsız gruplar tasarımından çok daha güçlüdür. Bu durum, eşleştirilmiş örneklemler kullanan tasarımların başlıca avantajlarından birini göstermekte olup sizi bu hesaplamalara götürmemin temel nedenini de ortaya koymaktadır.

Gücü hesaplamak için bir yaklaştırma işlemi kullandığımızı hatırlayın. Esasen örneklemlerin yeterince büyük olduğunu varsaydığımızda t dağılımının z dağılımına benzediğini varsayabiliriz. Eğer durum böyle değilse, bir eşleştirilmiş örneklem t'nin, buna karşılık gelen bağımsız gruplar *t*'ye ait *sd*'nin yarısı kadar *sd*'ye sahip olacağı gerçeğini dikkate almak zorundayız. Ayrıca $\rho = 0$ olduğunda iki tasarımın gücü tamamen eşit olmayacaktır.

## 8.7 Güç Üzerine Olan Durumu Tersine Çevirme

Burası gücü kullanarak farklı ve çok önemli bir noktaya temas etmek için iyi bir yerdir. Sık sık araştırmalar yaparız, sonucun alfa = ,05 düzeyinde manidar olduğunu görürüz ve oldukça heyecanlanırız. Fakat bazı kişiler p değerinin her şey olmadığını ve her şeyi bize söylemediğini ifade etmektedir. Bu kişilerden bazıları güven sınırlarının da en az farkın manidar olmasını bilmek kadar bilgi verici olduğunu bize hatırlatmakta ve çok daha fazlasını söylemektedir. Bu konuda okuduğum en iyi yazılardan biri Cumming'dir (2008).

Cumming'in vurguladığı noktayı göstermek için Aronson ve arkadaşlarının (1998) klişe tehtidi çalışmasına geri döneceğiz. Tehdit grubunun ortalamasını 6,93 (6,58 yerine) olarak değiştirdiğimizi varsayınız. Böylece ortalamalar farkı 3,06'dan 2,71'e düşer. Diğer tüm istatistikleri dokunmadan bırakıyorum. Şimdi *t*'yi hesaplarsak $t(21) = 2{,}092$ buluruz. *t*'nin kritik değeri $t_{21;\,0{,}05} = 2{,}08$'dir. Dolayısıyla sıfır hipotezini $\alpha = {,}05$ düzeyinde reddedebiliriz. (Gerçek olasılık 0,04875'tir.) Bu sonuç ile Aronson ve arkadaşlarının klişe tehdidinin manidar bir etki yaptığı sonucuna ulaşabilir ve doğru olabiliriz. Ayrıca etki büyüklüğü ölçüsü $d = 0{,}879$ olup iki durum arasındaki standart sapma farkının neredeyse 9/10'unu göstermektedir.

Fakat şimdi ilerlediğimi, Aronson ve arkadaşlarının çalışmasını tekrarlamak istediğimi ayrıca ilgili parametreler için en iyi tahminlerim olarak Aronson ve arkadaşlarının sonuçlarını aldığımı varsayın. Bu mantıklı olmayan bir şey değildir. Peki benim deneyimin gücü ne olacak? Az bir hesaplama, güç = 0,520 olacağını gösterir. Bu, manidar bir fark elde etme şansımın yaklaşık 50/50 olduğu anlamına gelmektedir. Bu, çok etkileyici değildir. Buradan ne sonuç çıkarmalıyız? İlk sonuç, sonraki deneyde ne olacağı ile ilgili p değerinin çok iyi bir gösterge olamayacağı şeklindedir. İkinci sonuç, güven aralığının, ki bu durumda $0{,}364 < \mu_1 - \mu_2 \leq 5{,}742$ olur, bunun daha fazla bilgi verici olduğunu ve durumlar arasındaki farkın doğru tarafı üzerinde sıkı bir tutamağımızın olmadığını gösterir.

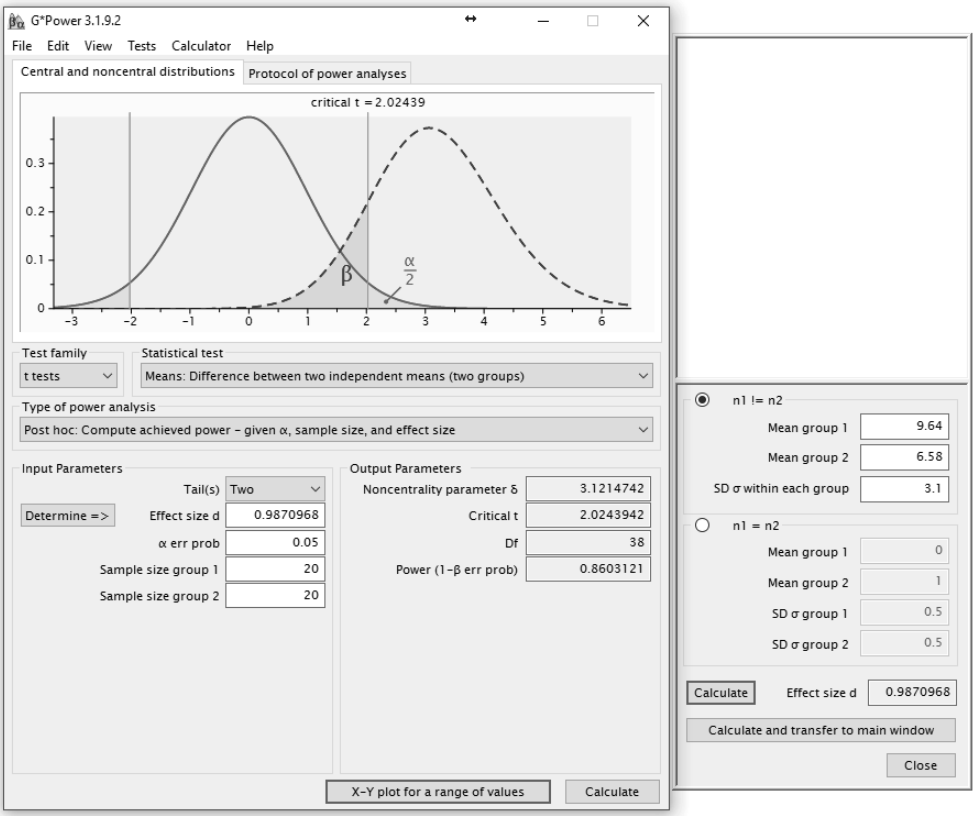
## 8.8 Daha Kompleks Desenlerde Güç Düşünceleri

Bu ünitede tartışmaları çoğunlukla şu ana kadar görmüş olduğumuz istatistiksel işlemlerle kısıtladım. Fakat burada tartıştıklarımızdan daha kompleks çok sayıda desen vardır. Özellikle tek yönlü varyans analizi, iki bağımsız gruptan daha fazla grup içeren ileri bir durumu ve faktöriyel varyans analizi bir bağımsız değişkenden daha fazla olan durumları gösterir. Bu durumların ikisinde de, *t* testinde kullandığımız hesaplama işlemlerinin makul derecede ileri seviyesini uygulayabiliriz. Uygun bölümlerde bu hesaplamaları konuşacağım fakat birçok durumda bu hesaplamaları yapmak için G*Power gibi programları kullanmak yerinde olacaktır. Güç hesaplamalarının arkasındaki teorik konuların çoğunu görmüş olmak iyi bir şey olup aslında öğreneceğimiz şeylerin çoğu şu ana kadar bildiklerimizin genişletilmiş halidir.

## 8.9 Hesaplamaları Basitleştirmek İçin G*Power'ın Kullanılması

G*Power programı birkaç yıldır vardır ve son zamanlarda yeni bir versiyonu geliştirilmiştir.[7] Yeni versiyonun kullanımı biraz daha karmaşıktır fakat mükemmeldir ve çabalamaya değerdir. Size programı indirmeyi ve üzerinde çalışmayı tavsiye ediyorum. G*Power'da nasıl yol alacağınızın her zaman açık olmadığını kabul ediyorum (çok sayıda seçenek vardır) fakat cevabını bildiğiniz (en azından yaklaşık olarak) bir örneği alarak program üzerinde çalışabilirsiniz. (Ben sabırsız biriyim. Bu nedenle doğru cevabı elde edene kadar farklı şeyler denerim. Programı kullanabilmek için yardım dosyalarını okumak daha mantıklı olabilir.)[7]

7 Bu yazının yazıldığı zamanda önceki versiyon, sitelerinde hâlâ mevcuttu. Bu versiyon ile başlayabilirsiniz çünkü bu versiyon daha az seçenek sunduğundan dolayı kullanımı daha kolay olabilir.



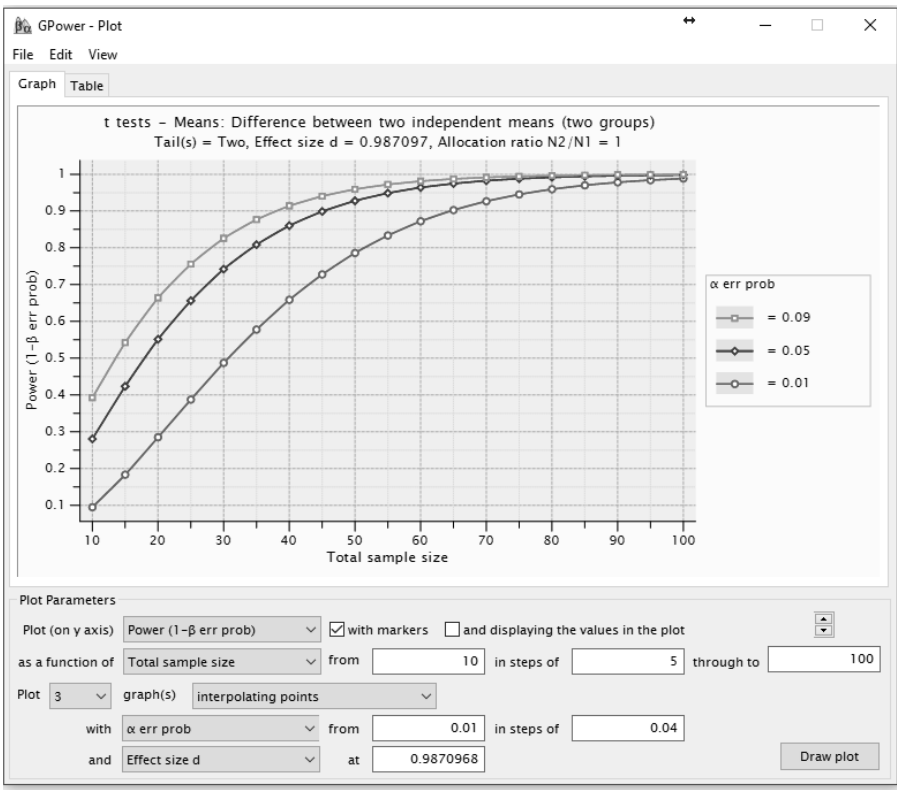

**Şekil 8.5** G*Power'ın ana ekranı (versiyon 3.1.2)

Franz Faul, Edgar Erdfelder, Albert-Georg Lang and Axel'in nazik izinleriyle Buchner/G*Power'dan alınmıştır.

G*Power yazarları tarafından oluşturulan bir yardım sayfası http://www.psycho.uni-duesseldorf.de/abteilungen /aap/gpower3/download-and-register/Dokumente/GPower3-BRM-Paper.pdf adresinde bulunmakta ve programın kullanımına yönelik mükemmel bir tartışma sunulmaktadır.

Yazılımın kullanılmasını göstermek için eşit olmayan örneklem büyüklüklerine başvurarak Bölüm 8.4'teki örneği tekrar oluşturacağım. İlk başladığınızda sizinki biraz farklı görünebilir olmasına rağmen Şekil 8.5 G*Power'ın açılış ekranını göstermektedir. (Ben versiyon 3.1.2'yi kullanıyorum.) Şimdilik ekranın yukarısındaki grafik çizimini düşünmeyin, çünkü zaten büyük olasılıkla bunları elde edemeyeceksiniz. Bu nedenle bir "Test Family" ve "Statistical test" seçebileceğiniz kutulara gidin. "Test family" olarak "*t* tests" ve "Statistical test" olarak "Means: Difference between two independent means (two groups)" seçin. Bunların altında yer alan "Post hoc: Compute achieved power-given a sample size, and effect size." ifadesini seçin. Bu yazılımı ben yazıyor olsaydım "Post hoc" ifadesini kullanmazdım, çünkü bu ifade ne yaptığınızı mutlaka yansıtmak zorunda değildir. (Post hoc gücü sonraki bölümde konuşacağım. Bu seçenek aslında "öncül" gücü hesaplayacaktır. Bu güç, eğer ortalamalar ve standart sapma kestirimleriniz doğruysa ve eğer girdiğiniz örneklem büyüklüğünü kullanırsanız deneyden önce sahip olacağınız güçtür.)

Şimdi çift kuyruklu test istediğinizi belirtmeye, üzerinde çalıştığınız alfa düzeyini (örneğin ,05) ve kullanmayı planladığınız örneklem büyüklüğünü girmeye ihtiyacınız var. Sonrasında kestirilen etki büyüklüğünü (d) eklemelisiniz. Eğer el ile hesapladıysanız bu değeri girin. Eğer böyle değilse "Determine ⇒" düğmesini tıklayın. Sağda bir diyalog kutusu açılacaktır. Beklenen ortalamaları ve standart sapmayı girin ve "calculate and transfer to main window" düğmesini tıklayın. Son olarak ana pencereye geri dönün, örneklem büyüklüğünü girin ve "Calculate" düğmesine basın. Pencerenin üzerindeki dağılımlar mucizevi bir şekilde görülecektir.




Franz Faul, Edgar Erdfelder, Albert-Georg Lang and Axel'in nazik izinleriyle Buchner/G*Power'dan alınmıştır.

**Şekil 8.6** Örneklem büyüklüğü ve alfanın fonksiyonu olarak güç

Bunlar Şekil 8.2'yi andırmaktadır. Ayrıca programın merkezî olmayan parametreyi (d), mevcut serbestlik derecesi altında ihtiyacınız olan *t*'nin kritik değerini ve son olarak bizim durumuz için 0,860 olan gücü hesapladığını göreceksiniz. Bu, bizim bulduğumuz sonuç ile uyumludur.

Örneklem büyüklüğü ve $\alpha$ düzeyi ile gücün nasıl arttığını *X* - *Y* grafiğinin çizimini isteyerek görebilirsiniz. Bu işi kendiniz yapın. Şekil 8.6'da örnek bir çıktı gösterilmiştir. Bu şekle göre gücün yüksek düzeylerinin büyük etkiler veya büyük örneklemler gerektirdiği açıktır. Etki büyüklüğündeki değişim ile gerekli olan örneklem büyüklüğü değişiminin nasıl olacağını gösteren grafikleri oluşturabilmelisiniz. Bunu size bırakıyorum.

## 8.10 Geriye Dönük Güç

**öncül güç**

Genel olarak yukarıdaki tartışma deney başlamadan önce hesaplanan güç olan **öncül güç (priori power)** üzerine odaklanmıştır. Bu güç, evrenimize ait parametreleri temsil edeceğine inandığımız ortalamalar, varyanslar, korelasyonlar, oranlar vb.nin mâkul kestirimine dayanır. Bu, istatistiksel gücü ele aldığımızda genellikle düşündüğümüz şeydir.

**geriye dönük güç**

Son yıllarda **geriye dönük güce** (veya post hoc power) olan ilgi giderek artmaktadır. Bizim amaçlarımız doğrultusunda geriye dönük güç, deney tamamlandıktan sonra bu deneyin

sonuçlarına dayanarak hesaplanan güç olarak tanımlanacaktır. (G*Power örneğinde "post hoc güç" ifadesini kullanmaya itiraz etmemin nedeni buydu. Orada deney yapılmadan önce gücü hesaplıyorduk.) Örneğin, geriye dönük güç "Eğer evren ortalamaları ve varyansları, bu deneyde bulunan değerlere eşitse sonuçta ortaya çıkan güç nedir?" sorusunu sorar.

Geriye dönük gücü hesaplamamız gerektiğine dair geçerli bir neden, gelecek araştırmaların desenlenmesine yardım etmektir. Bir deneyi yeni tamamladığımızı ve onu farklı örneklem büyüklüğü ve farklı demografik yapıda katılımcılar ile tekrar yapmak istediğimizi varsayınız. Henüz elde ettiğimiz sonuçları alabilir, onları evren ortalamaları ve standart sapmalarının doğru bir göstergesi olarak kabul edip bu değerleri kestirilen etki büyüklüğünü hesaplamak için kullanabiliriz. Böylece güç kestirimleri yapmak için bu etki büyüklüğünü kullanabiliriz. Bu işlemler, Aronson'ın çalışmasını tekrarlarken temelde ne yaptığımızı gösterir. Geriye dönük gücün bu şekilde kullanılması, ki bu aslında sonraki deneyimizin öncül gücüdür, nispeten tartışmasızdır. SAS ve SPSS gibi istatistiksel paketler bu hesaplamaları sizin için yapacaktır. G*Power'a yapması için verdiğim komutlar da buydu.

Tartışmalı olan durum ve olması gereken, geriye dönük güç hesaplamalarının *elde edilen sonuçların açıklanması* olarak kullanılmasıdır. Literatürdeki yaygın tavsiye, eğer çalışma manidar değilse fakat yüksek geriye dönük güce sahipse, sıfır hipotezinin kabul edilmesi doğrultusundadır. Bu görüş, eğer yüksek güce sahip olsaydınız yanlış bir sıfır hipotezini reddetme olasılığınızın yüksek olacağı argümanına dayanır. Böylece manidar olmama durumu, sıfır hipotezinin doğru veya doğruya oldukça yakın olduğunu gösterir. Bu oldukça inandırıcıdır. Fakat Hoenig ve Heisey'nin (2001) işaret ettiği gibi, burada yanlış bir sayıltı vardır. Sıfır hipotezini reddedememek ve hâl böyle olunca yüksek geriye dönük güce sahip olmak mümkün değildir. Aslında p'nin tam olarak ,05 olduğu bir sonuçta geriye dönük güç gerçekte ,50 olacaktır[8] ve $p > ,05$ için geriye dönük güç azalacaktır. Sıfır hipotezini $\alpha = ,05$ düzeyinde reddetmek için güçlükle başarısız olup ,80 güce sahip olan örnek bir çalışma oluşturmak bile mümkün değildir. Böyle bir şey olamaz!

Bazen ileri sürülen görüşlerden biri, geriye dönük gücün elde edilen p değerine göre size daha fazla bilgi verici olacağı şeklindedir. Bu görüş, önceki paragraftakinin bir türevi şeklindedir. Bununla birlikte verilen etki büyüklüğü ve örneklem büyüklüğü için p ve geriye dönük güç arasında 1:1 ilişki olduğunu göstermek kolaydır. Biri diğerinden türetilebilir. Dolayısıyla geriye dönük güç, manidar olmayan sonuçları açıklama yönünden ek bir bilgi sunmamaktadır.

Hoenig ve Heisey'nin (2001) ileri sürdüğü gibi, enerjimizi, bulgularımızın ortaya ne koyduğu hakkında daha fazla şey öğrenmeye çalışmak için geriye dönük gücü hesaplamaya odaklanmak yerine parametrelerin güven limitlerini veya etki büyüklüklerini hesaplamak için kullanmak daha iyi olacaktır. Örneğin, eğer iki bağımsız grup için $t(48) = 1,90$, $p = 0,063$ olacak şekilde elimizde bir t testi varsa, sıfır hipotezini reddetmekte başarısız olacağız. Geriye dönük gücü hesapladığımızda 0,46 ve $\mu_1 - \mu_2$ için %95 olasılıklı güven aralığını hesapladığımızda $-1,10 \leq \mu_1 - \mu_2 \leq 39,1$ buluruz. Güven aralığı, üzerinde çalıştığımız konu hakkında gücün sadece 0,46 olduğu gerçeğinden daha fazlasını söylemektedir. (Fark nispeten büyük ve manidar olduğunda bile güven aralığı, evren ortalamaları arasındaki farkın büyüklüğü ile ilgili çok iyi bir fikir olmadığını gösterir.)

Geçmişe dönük güç, meta-analizdeki gibi ya da gelecekteki çalışmamızı plânlarken; literatürdeki çalışmaları değerlendirdiğimizde, faydalı bir araç olabilir. Ancak geriye dönük güç, manidar olmayan sonuçlarımızı açıklamak için yararlı bir araç olmayacaktır.

---

8 Aslında bunu Bölüm 8.1 ve 8.5'te gördük. Orada Aronson ve arkadaşlarının değiştirilmiş tekrar çalışmasına ait bulgular için gücü hesaplamıştım.

## 8.11 Güç Analizinin Sonuçlarını Raporlama

Basılan bir çalışmada yürüttüğümüz deneyin gücü hakkında genellikle çok fazla şey söylemeliyiz. Bu belki de birkaç yıldır güç hesaplaması yapmamamızdan kaynaklanan ertelemedir. Bununla birlikte Yöntem bölümünüze deneyinizin gücünü açıklayan birkaç cümle eklemek yararlı olabilir. Örneğin, izlediğiniz yöntemi açıkladıktan sonra aşağıdaki gibi açıklamalarda bulunabilirsiniz:

Jones ve diğerlerinin (referansları listele) çalışmalarını temel alarak ortalama farkımızın yaklaşık 8 puan, her grup için standart sapmanın ise yaklaşık 5 puan olması gerektiğini tahmin ettik. Bu bize kestirilen etki büyüklüğünü *(d)* 8/11 = ,73 olarak verir. Güç kestirimini ,80 olarak amaçladığımızdan bu güç düzeyine kestirilen etki büyüklüğümüz ile ulaşmak için iki grubun her biri için 30 katılımcı kullandık.

## Anahtar Terimler

Güç (Giriş)
Etki büyüklüğü (*d*) (8.3)
$\delta$ (delta) (8.3)
Merkezî olmayan parametre (8.4)
Harmonik ortalama ($\overline{X}_h$) (8.5)
Etkin örneklem büyüklüğü (8.5)
Öncül güç (8.10)
Geriye dönük güç (8.10)
Post hoc güç (8.10)

## Alıştırmalar

8.1 Akran baskısının etkisi üzerine yapılan birçok araştırma göstermiştir ki akran baskısı ölçeği için ortalama etki puanı 520, standart sapması ise 80'dir. Bir araştırmacı durumlarda küçük bir değişiklik yaparak ortalamanın 500 olacağı şekilde puanların oluşacağını göstermek istemektedir. Bunu yapmak için bir t testi uygulayarak kendi örneklem ortalamasını evren ortalaması olan 520 ile karşılaştırmayı planlıyor.

a. Sorudaki etki büyüklüğü nedir?
b. Araştırmacının örneklem büyüklüğü 100 ise $\delta$'nın değeri kaçtır?
c. Testin gücü nedir?

8.2 Alıştırma 8.1'de açıklanan durumun grafiğini Şekil 8.1'deki gibi çiziniz.

8.3 Alıştırma 8.1'de gücü ,70; ,80 ve ,90'a çıkarmak için ne büyüklükte örneklemlere ihtiyaç duyulur?

8.4 İkinci bir araştırmacı oldukça farklı bir değişiklik yaparak ortalama etki puanını 520'den 550'ye çıkarabileceğini göstermeyi düşünmektedir.
a. Sorudaki etki büyüklüğü nedir?
b. Araştırmacının örneklem büyüklüğü 100 ise $\delta$'nın değeri nedir?
c. Testin gücü nedir?

8.5 Alıştırma 8.4'te açıklanan durumun grafiğini Şekil 8.2'deki gibi çiziniz.

8.6 Üçüncü bir araştırmacının Alıştırma 8.2 ve 8.5'teki durumların ikisini de yürüttüğünü varsayınız. Araştırmacı iki deneysel değişim arasında bir fark bulmak için birleşik deneyin gücünü bulmak istiyor.

a. Sorudaki etki büyüklüğü nedir?
b. İki grubun da örneklem büyüklüğü 50 ise *d*'nin değeri nedir?
c. Testin gücü nedir?

8.7 Bir fizyolojik psikoloji laboratuvarı birkaç yıldır tavşanlarda kaçınma davranışı üzerine çalışmakta ve konu ile ilgili birçok yayın yapmaktadır. Bu araştırma ile özel bir görev için ortalama tepki gecikme süresinin 5,8 saniye, standart sapmanın ise 2 saniye olduğu açıkça görülmektedir (yüzlerce tavşana dayanarak). Şimdi araştırmacılar tavşan amigdalalerinin bazı bölgelerinde doku yapısı değişikliği oluşturarak bu hayvanlarda en zayıf kaçınma durumunu göstermek istemektedir (yani tavşanlar cezalandırılmış bir tepkiyi daha çabuk tekrarlayacaklar). Araştırmacılar gecikme süresinin 1 saniye kadar azalmasını beklemekte ve tek örneklem *t* testi kullanmayı planlamaktadırlar ($\mu_0 = 5,8$).

a. Başarı şansının en az 50:50 olması için kaç deneğe ihtiyaç duyarlar?

b. Başarı şansının en az 80:20 olması için kaç deneğe ihtiyaç duyarlar?

8.8 Alıştırma 8.7'deki araştırma grubu tek grup ile çalışmak ve bunu $\mu_0 = 5,8$ ile karşılaştırmak yerine iki grup (biri doku değişikliği yapılmış, diğeri doku değişikliği yapılmamış) ile çalışmaya karar vermişlerdir. Araştırmacılar hâlâ aynı derecede fark beklemektedir.

a. Güç = 0,60 olması için araştırmacılar toplam olarak kaç deneğe ihtiyaç duyarlar?

b. Güç = 0,90 olması için araştırmacılar toplam olarak kaç deneğe ihtiyaç duyarlar?

8.9 Bir araştırma görevlisi önceden güç hesaplamaları yapmadan Alıştırma 8.8'de açıklanan deneyi yapmaktadır. Her grupta 20 denek ile çalışmak istemiştir, fakat kaza ile bir raf kafesin devrilmesi sonucunda 5 deneği iptal etmek zorunda kalmıştır. Bu deneyin gücü nedir?

8.10 Düşük ve normal ağırlıkta doğup şu anda bir yaşına giren bebeklerin bilişsel gelişimlerini karşılaştıran bir çalışma yürütüyoruz. Geliştirdiğimiz bir ölçeği kullanarak iki grubun örneklem ortalamalarını sırasıyla 25 ve 30, birleşik standart sapmasını 8 bulduk. Bu deneyi her grupta 20 denek olacak şekilde tekrar yapmak istediğimizi varsayınız. Gerçek ortalamalar ve standart sapmaların tam olarak kestirildiğini varsayarsak tekrar çalışmamızda manidar bir fark bulmayı sağlayacak öncül olasılık nedir?

8.11 Alıştırma 8.10'daki orijinal veriye *t* testi uygulayın. Bu soruya cevabınız, Alıştırma 8.10'da verdiğiniz cevabın tersine ne gösterir?

8.12 İki doktora öğrencisi tezlerini yeni tamamlamışlardır. Her biri, iki bağımsız grup için *t* testi kullanmıştır. Biri grup başına 10 denek kullanarak manidar bir *t* bulmuştur. Diğeri grup başına 45 denek kullanarak aynı büyüklükte manidar bir *t* bulmuştur. Hangi sonuç sizi daha çok etkiler?

8.13 Alıştırma 8.12'deki cevabınızı savunmak için bir grafik çizin (Şekil 8.2'ye benzer biçimde).

8.14 Toplam 30 deneğin yer aldığı, örneklem büyüklükleri birbirine oldukça yaklaştıkça, gücün artacağını gösteren basit bir iki grup örneği hazırlayın.

8.15 Sorunlu bir doktora adayı tezini başarılı bir şekilde savunabilmek için kendini manidar sonuçlar bulmak zorunda hissetmektedir. Bu aday normal bir grup ile önceden suç işleyenlerden oluşan bir grup arasında kendi geliştirdiği ölçek ile ölçülen sosyal sorumlulukları açısından bir fark olduğunu göstermek istemektedir. Bununla birlikte doktora adayının bir problemi vardır. Adayın elinde normal grubun gerçek ortalamasının 38 olduğunu gösteren veri vardır ve adayda bu deneklerden 50 kişi bulunmaktadır. Aday, geçmişte suç işleyenler olarak sınıflanan 100 lise mezununa veya geçmişte suç işlemiş ve liseyi terk etmiş 25 deneğe ulaşabilir durumdadır. Aday, lise mezunlarının ortalamanın yaklaşık 35 olduğu bir evrenden geldiğini, okulu terk eden grubun ise ortalamasının yaklaşık 30 olduğu bir evrenden geldiğini zannetmektedir. Aday, bu gruplardan sadece birini kullanabilir. Hangisini kullanmalıdır?

8.16 Bölüm 8.5'te bulunan sonuçlara tekrar ulaşmak için G*Power veya benzer bir yazılım kullanın.

8.17 Kadınları kapsayacak şekilde Aronson'ın çalışmasını (Bölüm 8.5'te açıklanmıştır.) genişletelim. (Kadınlar maalesef erkekler kadar matematik becerileri için güçlü bir yatırıma sahip değillerdir. Muhtemelen onlar başkalarından daha iyi yapmak için de şartlandırılmamaktadırlar.) Kadınlar için grubun ortalamasını 8,5, kontrol grubunun ortlamasını 8 beklemekteyiz. Ayrıca 3,10 olan kestirilmiş standart sapmanın mâkul olduğunu ve örneklem büyüklüğünün 25 olacağını varsayınız. Bayanlarda klişe tehtidinin bir etkisini göstermek için bu deneyin gücünü hesaplayınız.

8.18 Bir ortalama için, $\alpha$ = ,05 manidarlık düzeyinde ve tek kuyruklu bir sıfır hipotezini test etmek istediğimizi varsayınız. Ayrıca tüm gerekli varsayımların karşılandığını kabul ediniz. Doğru bir $H_0$'ı reddetme olasılığının yanlış bir $H_0$'ı reddetme olasılığından çok daha fazla olduğu bir durum olabilir mi? (Diğer bir ifadeyle, güç $\alpha$'dan küçük olabilir mi?)

8.19 $\sigma$ = 15 ve $n$ = 25 için $H_1$: $\mu_0 > 100$'e karşılık $H_0$: $\mu_0 = 100$'ü test ediyoruz. $H_1$ altındaki ortalamanın hangi değeri, II. Tip hatanın olasılığına eşit güç ile sonuçlanır? (İpucu: İki dağılımı çizmeye çalışın, hangi alanları eşitlemeye çalışıyorsunuz?)

## Tartışma Soruları

8.20 Prentice ve Miller'ın (1992) sunduğu ilginç bir düşünceye göre, öğrencilerin çoğu hangi konuda çalışıyor olurlarsa olsunlar, etki büyüklüğünü artırmak için elinden gelenin en iyisini yapmaya çalışırlarken (örneğin, gruplar arasındaki farkı artırarak), bazı araştırmalar etkiyi azaltmaya ve hâlâ bir fark bulmaya odaklanır. (Örneğin, kişilerin kendi gruplarına ait bireylerin lehinde olduğu bilinmesine rağmen tesadüfî atama yoluyla gruplar oluşturursanız bile etkinin var olduğu gösterilmiştir.) Buna göre Prentice ve Miller'ın ifadesi şu şekildedir: "Açıkladığımız çalışmalarda araştırmacılar işlemsel gücü en aza indirmişlerdir ve böyle yaparak altta yatan sürecin gücünü gösterebilmede başarılı olmuşlardır."

a. Bu, size durumu açıklayan doğru bir ifade olarak gelmekte midir? Diğer bir ifadeyle, araştırmacıların deneylerini minimum güç ile yürüttüklerine katılıyor musunuz?

b. Bu yaklaşım, psikolojideki birçok araştırma için mantıklı görünüyor mu?

c. Büyük etkiler bulmak daima önemli midir? Oldukça küçük etkiler bulmak ne zaman önemli olabilir?

8.21 İki bağımsız grup ile Aronson'ın klişe tehdidi çalışmasına dayandırılan hipotetik çalışmada bir laboratuvar grubundaki tüm erkek öğrenciler aynı koşulda testi alıyorlar. Daha sonra başka laboratuvardaki erkek öğrenciler başka bir koşulda testi alıyorlar.

a. Bu yaklaşımda yanlış olan nedir?

b. Hangi alternatifleri önerebilirsiniz?

c. Bu laboratuvarlarda benim ihmal ettiğim birçok kadın vardır. Eğer onları da kullanmış olsaydım ne olabileceği hakkında ne düşünürsünüz?

8.22 Eşleştirilmiş örneklem $t$ testi kullanmak için Aronson'ın çalışmasının değiştirilmiş hâlinde daima öncelikle kontrol koşullarını, bunu izleyen sonraki haftada ise deney koşullarını verdim.

a. Bu, koşulların sırasının rastgele belirlenmesinden neden daha iyi bir yaklaşım olabilir?

b. Eğer her hafta tamamen aynı testi verirsem ilk uygulamadan kaynaklı bazı hatırlamalar olabilir. Bu problemin üstesinden nasıl gelebilirim?

8.23 Alıştırma 8.21 ve 8.22'nin neden bir istatistik metninde yer aldığını düşünüyorsunuz?

8.24 Bir farkın $\alpha$ = ,05 düzeyinde istatistiksel olarak zar zor manidar olan bir örnek oluşturunuz. (İpucu: $t$ için kritik değeri bulunuz, $\mu_1$, $\mu_2$, $n_1$ ve $n_2$'yi oluşturun ve sonra s'nin gerekli değeri için çözünüz.) Şimdi bu deneyin geriye dönük gücünü hesaplayınız.



Bölüm 9

# Korelasyon ve Regresyon

## Hedefler

Korelasyon ve regresyonu tanıtma, değişkenler arasındaki ilişkileri açıklamaya başlama.

## Konular


9.1 Dağılımın Grafiği
9.2 Yaşam Temposu ile Kalp Hastalığı Arasındaki İlişki
9.3 Stres ve Sağlık Arasındaki İlişki
9.4 Kovaryans
9.5 Pearson Momentler Çarpımı Korelasyon Katsayısı (*r*)
9.6 Regresyon Doğrusu
9.7 Veriye Uygun Bir Doğrunun Çizilmesinin Diğer Yolları
9.8 Yordamanın Doğruluğu
9.9 Regresyon ve Korelasyon Altındaki Varsayımlar
9.10 $\hat{Y}$ 'nin Güven Sınırları
9.11 Test Alma Becerilerinin Rolünü Gösteren Bir Bilgisayar Örneği
9.12 Hipotez Testi
9.13 Son Bir Örnek
9.14 Korelasyon ve Regresyonda Varsayımların Rolü
9.15 Korelasyonu Etkileyen Faktörler
9.16 Pearson'ın *r*'si İçin Güç Hesaplaması

ilişkiler
farklar

BÖLÜM 7'DE örneklem ortalamaları arasındaki farka ait hipotez testleriyle ilgilendik. Bu bölümde değişkenler arasındaki ilişkiye dair soruları incelemeye başlayacağız. **İlişkiler** ve **farklar** arasında çok fazla ayrım yapılmamasına rağmen (eğer işlemler farklı ortalamalara sahipse, ortalamalar işlemlerle ilişkilidir) ayrım araştırmacının ilgisi ve araştırma yapısı açısından kullanışlıdır. Ortalamalar arasındaki farkla ilgileneceğimiz zaman, araştırma genellikle bağımsız değişkenin (örnek: İşlem A ve İşlem B) nicel veya nitel birkaç düzeyinden oluşur ve araştırmacı bağımlı değişkenin bir işlemden diğerine farklılaştığının gösterimiyle ilgilenir. Ancak, ilişki ile ilgileneceğimiz zaman, bağımsız değişken ($X$) genellikle birçok nicel düzeye sahiptir ve araştırmacı bağımsız değişkenin, bağımlı değişkenin bir *fonksiyonu* olduğunu göstermekle ilgilenebilir.

korelasyon
regresyon

Bu bölüm birbiriyle ilişkili iki konu ile ilgilenecektir: **korelasyon** ve **regresyon.** İstatistikçiler genellikle bu iki teknik arasında ayrım yapmaktadırlar. Fark, pratikte çok sık incelenmese de, üzerinde kısaca durulacak kadar önemlidir. Basit korelasyon ve regresyon problemlerinde veri, $N$ tür nesneye ait, söz konusu iki değişkenin her birinden birer gözlem olacak şekilde, iki gözlem içerir. Eğer bir labirentteki farelerin koşma hızı ($Y$) ile bazı ölçütlere erişmek için deneme sayıları ($X$) (ikisi de öğrenmenin genel ölçümüdür) arasındaki korelasyonla ilgileniliyorsa, her birinden fare için koşma hızı puanı ve ölçüte ulaşmak için denemeler puanı elde edilmelidir. Benzer şekilde koşma hızının ($Y$), pekiştireç olarak alınan yiyecek kapsülüne ($X$) bağlı regresyon ile ilgileniliyorsa her bir farenin koşma hızı ve aldığı yem kapsülü sayısına ait değerleri olmalıdır. Bu iki durum arasındaki farklılık korelasyon ve regresyon arasındaki istatistiksel ayrımı gösterir. Her iki durumda, $Y$ (koşma hızı) araştırmacının kontrolünün dışındaki **tesadüfi değişken**dir. Bir deneme yapılıp hız ölçümünü gerçekleştirinceye kadar farenin koşma hızının ne olacağı bilinmemektedir. Önceki durumda $X$ tesadüfi değişkendir, çünkü ölçüte ilişkin yapılan deneme sayıları araştırmacının kontrolünün ötesinde hayvanın ne kadar hızlı öğrendiğine de bağlıdır. Başka bir ifadeyle araştırmanın tekrarları, hem $Y$ hem $X$ değişkeninin farklı değerlerini vermektedir. Buna rağmen yiyecek kapsülü örneğinde $X$ **sabit değişken**dir**.** Kapsüllerin sayısı araştırmacı tarafından belirlenmekte (örneğin; 0, 1, 2 veya 3 kapsül) ve tekrarlar boyunca sabit kalmaktadır.

desadüfi değişken

sabit değişken

Birçok istatistikçiye göre $X$'in sabit veya araştırmacı tarafından veri toplanmadan önce belirlenmesi durumunda *regresyon* ifadesi kullanımı kalıplaşmıştır. Bu durumlarda $X$ örneklem hatası içermez ve araştırmanın tekrarlı ölçümleri aynı $X$ kümesi değerlerini içerir. *Korelasyon* kelimesi ise $X$ ve $Y$ değerlerinin her ikisinin birden tesadüfî değişken olarak seçildiği durumları tanımlamada kullanılır. Bu durumda, hem $X$ hem de $Y$'ler bir tekrardan diğerine değişecektir ve böylece her iki değişken de örneklem hatası içerecektir. Bu ayrım kısaca **doğrusal regresyon modeli** ve **iki değişkenli normal model** arasındaki temel ayrımdır. Bu iki model arasındaki ayrımı, Bölüm 9.9'da ayrıntılı olarak ele alacağız.

doğrusal regresyon modeli
iki değişkenli normal model

İki model arasındaki ayrım, istatistiksel temele uygun olmasına rağmen, pratikte kırılmaya eğilimlidir. Her iki değişkenin tesadüfi olduğunda regresyonun (korelasyondan daha çok) hedeflendiği durumların örneklerini göreceğiz. Daha faydacı bir ayrım araştırmacının ilgisine bağlıdır. Araştırmada, $X$ bilgisine dayanarak $Y$'yi **yordama** amaçlanmışsa regresyondan, fakat iki değişkenin arasındaki ilişkinin derecesini elde etmek amaçlanmışsa korelasyondan söz ederiz. Bu ayrıma mantıklı itirazların artmasına karşın, pratikte, bu iki tekniğin farklı kullanılmasının avantajları söz konusudur.

yordama

Fakat regresyon sadece "yordama" ile sınırlı değildir. Aslında davranışçı bilim adamlarının kullanış amaçlarından sadece küçük bir parçası olabilir. Örneğin, Bölüm 15'te söz edilecek bir örnekte, her bir eyaletin eğitim harcaması ile öğrenci-öğretmen oranı ve öğ-

renci performansları arasındaki ilişkiye bakılacaktır. Burada özellikle belirli bir eyalet için kendine özgü harcamaları ve öğrenci-öğretmen oranını dikkate alarak öğrenci başarısının yordanmasına bakacağız. Daha çok bu yordayıcı değişkenler arasındaki ilişki ve hep birlikte başarıyı nasıl açıkladıkları ile ilgileneceğiz. İlişkileri anlamak yerine sonuçların yordanması regresyonun en temel hedefidir.

Korelasyon ve regresyon arasındaki ayrımı yaptıktan sonra, iki teknik birbiriyle çok yakından ilişkili olduğundan birlikte kullanmaya devam edeceğiz. Genel bir problem, bir değişkeni diğerinin bilgisinden tahmin etmek (regresyon) ve ilişkinin derecesini elde etmek (korelasyon) için aralarındaki denklemi geliştirmede olur. Şu an belirtilecek tek kısıtlama *X* ve *Y* arasındaki ilişkinin doğrusal olmasıdır. Eğrisel ilişkiler burada dikkate alınmayacaktır, ancak Bölüm 15'te birbiriyle ilişkili tekniklerle nasıl ele alındığını görecegiz.

## 9.1 Dağılımın Grafiği

**dağılımın grafiği**

İki değişken arasındaki ilişkiyi açıklamak amacıyla değişkenlere ait ölçümleri elde ettiğimizde, bu ilişki hakkında fikir sahibi olmak için en yararlı tekniklerden biri **dağılımın grafiği**dir (dağılım diyagramı). Dağılım grafiğinde, çalışmadaki her bir denek iki boyutlu uzayda bir nokta ile temsil edilir. Bu noktanın koordinatları ($X_i$ ve $Y_i$), bireyin (veya nesnenin) *X* ve *Y* değişkenlerindeki bağıl değerleridir. Şekil 9.1'de üç dağılım grafiği örneği yer almaktadır. Bunlar gerçek veri kümelerine aittir.

**yordayıcı**
**ölçüt**

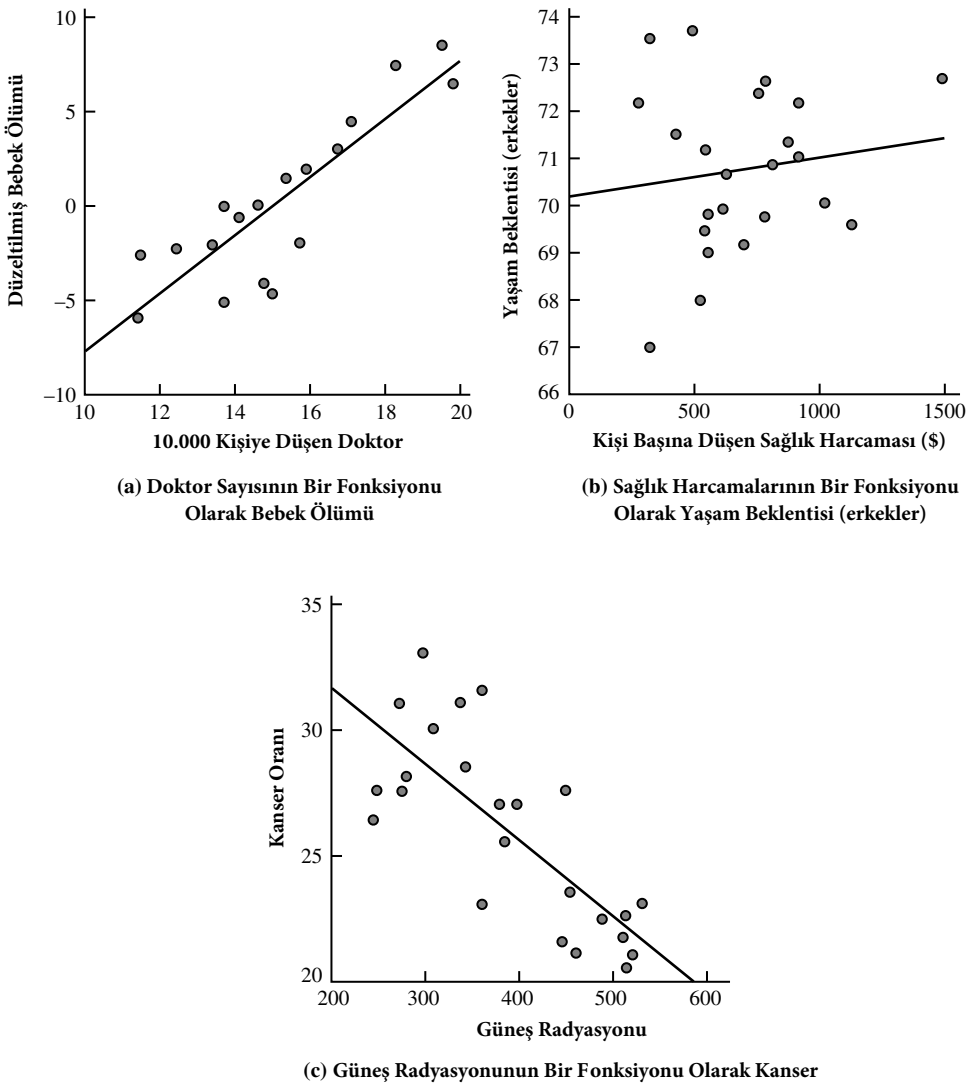
Bir dağılım grafiğinde **yordayıcı** değişken genellikle apsis veya *X*-ekseni ve **ölçüt** değişkeni ordinat veya *Y*-ekseni üzerinde temsil edilir. Çalışmanın nihaî amacı bir değişkeni diğerinin bilgisinden yordamak veya açıklamaksa ayırt etmek kolaydır; ölçüt değişken yordanacak olan, yordayıcı değişken ise yordamakta kullanılandır. Problem sadece korelasyon katsayısını elde etmek ise, ayrım kolaydır (kansere yakalanma sıklığı, sigara içme miktarına bağlı olacak ve sıklık değeri ordinat üzerinde yer alacaktır) veya tersi olabilir (ne koşma hızı ne de ölçüt için yapılan deneme sayıları açıkça birbirine bağımlı pozisyonda değildirler). Ayrımın açık olmadığı durumlarda hangi değişkenin *X* olarak etiketleneceği önemsizdir.

Şekil 9.1'deki üç dağılım grafiğini ele alalım. Şekil 9.1a, St. Leger, Cochrane ve Moore (1978) tarafından raporlanmış gayri safi millî hasılaya göre düzeltilmiş bebek ölümleri ve 10.000 kişiye düşen doktor sayıları arasındaki ilişkiye ait verinin grafiğidir.[1] Burada, bebek *ölümlerinin* artması ile doktor sayısının artması sürpriz sonucuna dikkat edilmelidir. Bu kesinlikle beklenmedik bir sonuçtur; fakat hemen hemen kesinlikle şansa bağlı değildir. Bu veriye bakılarak ve bölümün geri kalanını okuyarak bu sürpriz sonucun muhtemel açıklamalarını düşünebilirsiniz. Ohio Universitesi'nde görevli Justin Fuller bazı ilginç öneriler sunmuştur. Ancak bunların büyük bir ihtimalle raporlama problemi olması muhtemeldir - daha çok doktor daha iyi raporlama demektir. Aynı zamanda doğum uzmanlarının, doğumdan hemen sonra ölenlerin canlı doğum sayılarını arttırmaları ve doğumdan hemen sonra ölenlerin ölü sayılmaları, ancak canlı doğmadıkları takdirde sayılmadıkları düşünülmektedir.

Şekil 9.1a-9.1.c'de birleştirilmiş doğrular "veriye en iyi uyumu" gösteren doğruları temsil eder. Bu bölümün büyük bir kısmının konusu doğruların nasıl belirleneceğidir. Şekillerin her birine ait ilişkileri açıklığa kavuşturmak için doğrular ekledim. Bu doğrular *Y*'nin *X* yardımıyla yordandığı (kısaltılmışı "*X*' e bağlı "*Y*" ) **regresyon doğruları** olarak adlandırılır ve her örneklem veya gözlemde $X_i$'nin verilen değeri için $Y_i$'nin en iyi yordananını temsil eder. Verilen herhangi belli bir *X* değeri için, regresyon doğrusunda rastlayan

**regresyon doğruları**

1 Ölüm oranı için negatif değerler, gayri safi millî hasılaya göre düzeltilmiş olan ölüm oranlarından gelmektedir. Düzeltmeden sonra ölüm oranı negatif olabilir.




**(a) Doktor Sayısının Bir Fonksiyonu Olarak Bebek Ölümü**

**(b) Sağlık Harcamalarının Bir Fonksiyonu Olarak Yaşam Beklentisi (erkekler)**

**(c) Güneş Radyasyonunun Bir Fonksiyonu Olarak Kanser**

© Cengage Learning 2013

**Şekil 9.1** Üç dağılım grafiği

noktanın yüksekliği *Y*'nin en iyi yordananını temsil eder ($\widehat{Y}$ diye gösterilir ve "*Y* şapka" diye okunur). Diğer bir deyişle $X_i$'den regresyon doğrusuna dik bir çizgi çizeriz ve yatay olarak *Y* eksenine taşırız ve $\widehat{Y}_i$ değerini okuruz. *Y*'nin üzerindeki şapka, *Y*'nin yordandığını gösterir.

**korelasyon (*r*)**

Regresyon doğrusu etrafındaki noktalar kümesinin derecesi (diğer bir ifadeyle *Y*'nin gerçek değeri ile yordanan değerinin uyuşma derecesi) *X* ve *Y* arasındaki **korelasyonla (r)** ilişkilidir. Korelasyon katsayısının ranjı 1 ile -1 arasında değişir. Şekil 9.1a'daki noktalar kümesi doğruya, iki değişken arasında güçlü bir ilişki olduğunu gösterecek şekilde çok yakındır. Eğer noktalar tam olarak doğru üzerine düşerse korelasyon +1,00 olur. Örnekte olduğu gibi korelasyonun değeri ,81'dir ve davranış bilimleri için gerçek değişkenler arasında yüksek bir ilişkiyi temsil eder.

Şekil 9.1.b'de ortalama ömür (erkekler için) ile 23 gelişmiş ülke (çoğunlukla Avrupa) için kişi başına düşen sağlık harcaması arasındaki ilişkiye ait verinin grafiğini çizdim. Bu veri Cochrane, St. Leger ve Moore'da (1978) bulunmaktadır. Sağlık harcamalarının uluslararası düzeyde tartışıldığı dönemde bu veri bizi duraklatmaktadır. Eğer bir ülkenin sağlığı, ortalama ömür ile ölçülseydi (kuşkusuz sadece ve kesinlikle çok iyi ölçüm değil), sağlık hizmetlerine harcanan toplam para miktarının sağlık kalitesinin sonuçları ile ilişkisinin olmadığı görülürdü (farklı ülkelerin de kendi harcamalarını benzer şekilde paylaştırdığı varsayılacaktır). (Onbinlerce dolar harcanan bir organ nakli *bir bireyin* ömür beklentisini birkaç yıl uzatabilir ancak ülkenin ortalama ömründe bir etki etmeyecektir). Ancak, çocuklarda önleme hizmetleri için harcanan benzer miktardaki para, sonunda gözle görülür bir etkiye neden olabilir - bu nedenle örneğin bu metine dahil edilmesi öncelikle psikologlar içindir. En uzun ortalama yaşam ömrüne sahip iki ülke (İzlanda ve Japonya), sağlık hizmetlerine, en kısa ortalama yaşam ömrüne sahip olan ülke (Portekiz) ile yaklaşık aynı miktarda para harcamıştır. Amerika Birleşik Devletleri, harcamalarda en yüksek ikinci orana sahiptir, ancak ortalama yaşam ömründe neredeyse en alt sıralarda yer almaktadır. Şekil 9.1b söz konusu iki değişken arasında belirgin bir ilişkinin olmadığı durumu temsil etmektedir. Eğer iki değişken arasında kesinlikle bir ilişki yok ise korelasyon 0,0 olur. Bu örnekte korelasyon sadece ,14'tür ve bu değer bile 0,0'dan pek farklı olmadığını göstermektedir.

Son olarak Şekil 9.1c'de *Newsweek*'de (1991) yayımlanmış göğüs kanseri ile güneş ışınları arasındaki ilişki üzerine bir makaleden alınan veri görülmektedir. Güneşi sevenler için, ekstra güneş ışığının en azından biraz yararı olduğunu öğrenmek cesaret vericidir. Burada güneş radyasyonu miktarı arttıkça, göğüs kanserinden ölümlerin görülme sıklığının azaldığına dikkat edilmelidir. (Son yıllarda, bazı kanser türleri oranındaki azalmanın güneş ışınları sayesinde artan, vücudun D vitamini üretimi ile ilişkili olduğunun düşünüldüğüne dair çok sayıda araştırma yapılmıştır.[2] Farklı yollarda veriyi tanımlayan başarılı bir makale Gardland ve diğerlerinin (2006) çalışmasında bulunabilir). Bu örnek negatif ilişkiye iyi bir örnektir ve korelasyon -,76'dır.

Buradaki korelasyon katsayısı işaretinin, ilişkinin yönünü belirtmekten başka bir anlamının olmadığı unutulmamalıdır. Korelasyon değerinin ,75 ve -,75 olması aynı ilişki derecesini belirtir. Sadece ilişkinin yönü farklıdır. Şekil 9.1a ve 9.1c, korelasyon işaretleri dışında neredeyse aynı büyüklüğü göstermektedir (,81'e karşı -,76).

## 9.2 Yaşam Temposu ile Kalp Hastalığı Arasındaki İlişki

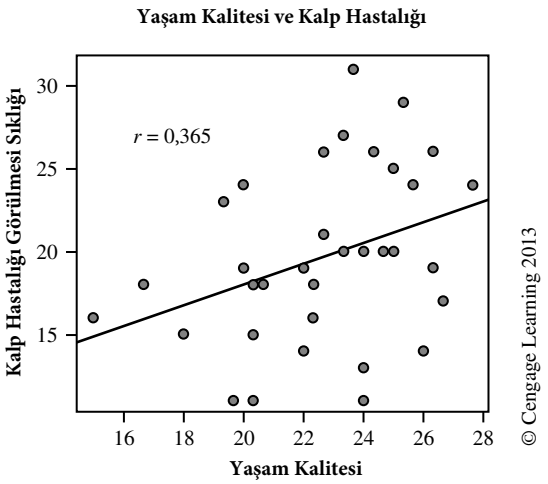
Şekil 9.1'de ya çok güçlü ilişkili (pozitif veya negatif) ya da birbirinden neredeyse bağımsız değişkenlerin örnekleri gösterilmiştir. Şimdi korelasyonun yüksek olmadığı ancak yine de manidar düzeyde 0'dan büyük olduğu bir örneğe döneceğiz. Ayrıca, davranış bilimleri adamlarının sık sık yaptığı çalışma türlerine de daha benzerdir.

Daha hızlı yaşam temposuna sahip bireylerin kalp krizi ve ölümcül hastalığın diğer formlarına daha müsait olduğuna dair ortak bir kanı vardır. ("A tipi kişilik" tartışması akla gelmelidir). Levine (1990) "yaşam temposu" ve yaşa göre düzeltilmiş iskemik kalp hastalığından ölüm oranlarına ilişkin veriyi yayımlamıştır. Çalışmasında farklı coğrafik konum ve büyüklüklere sahip 36 şehirden veri toplamıştır. İş "yaşam temposu" ölçümüne geldiğinde o çok yaratıcı olmuştur. Bankadaki kasiyerin 20 doları bozmada harcadığı zamanı, normal bir insanın 60 feet uzaklığı ne kadar zamanda yürüdüğünü ve insanların hangi hızla konuştuğunu gizlice kaydetmek için bir kronometre kullanmıştır. Levine ayrıca her bir

2 Yakın tarihli bir çalışma (Lappe, Davies, Travers-Gustafson ve Heaney (2006)) D vitamini seviyesi ile çeşitli kanser türlerinin düşük oranı arasındaki bir ilişkiyi göstermiştir.

şehirden yaşa göre düzeltilmiş iskemik kalp hastalığından ölüm oranlarını da kaydetmiştir. Aşağıdaki veri kümesinde "yaşam temposu" üç ölçümün ortalaması olarak ele alınmıştır (ölçüm birimleri keyfîdir). Her üç değişken Web'de yer alan veri kümesinde mevcuttur.) Burada bağımlı iki değişkenin olduğu bir örnek söz konusudur fakat bir tanesinin yordayıcı değişken olduğu açıktır (Yaşam temposu $X$ (yatay) eksende ve Kalp hastalığı $Y$ (dikey) eksende yer almıştır).

İlgili Verinin Grafiği Şekil 9.2'de çizilmiştir.




**Şekil 9.2** Yaşam temposu ve yaşa göre düzeltilmiş kalp hastalığı

**Tablo 9.1** Birleşik Devletler'in 36 şehrinde yaşam temposu ve kalp hastalığından ölüm oranları

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hız *(X)* | 27,67 | 25,33 | 23,67 | 26,33 | 26,33 | 25,00 | 26,67 | 26,33 | 24,33 | 25,67 |
| Kalp *(Y)* | 24 | 29 | 31 | 26 | 26 | 20 | 17 | 19 | 26 | 24 |
| Hız *(X)* | 22,67 | 25,00 | 26,00 | 24,00 | 26,33 | 20,00 | 24,67 | 24,00 | 24,00 | 20,67 |
| Kalp *(Y)* | 26 | 25 | 14 | 11 | 19 | 24 | 20 | 13 | 20 | 18 |
| Hız *(X)* | 22,33 | 22,00 | 19,33 | 19,67 | 23,33 | 22,33 | 20,33 | 23,33 | 20,33 | 22,67 |
| Kalp *(Y)* | 16 | 19 | 23 | 11 | 27 | 18 | 15 | 20 | 18 | 21 |
| Hız *(X)* | 20,33 | 22,00 | 20,00 | 18,00 | 16,67 | 15,00 | | | | |
| Kalp *(Y)* | 11 | 14 | 19 | 15 | 18 | 16 | | | | |
| $\Sigma X = 822{,}333$ | | | $\Sigma Y = 713$ | | | $\Sigma XY = 16{,}487.67$ | | | | |
| $\Sigma X^2 = 19{,}102.33$ | | | $\Sigma Y^2 = 15{,}073$ | | | $N = 36$ | | | | |
| $X = 22{,}84$ | | | $\bar{Y} = 19.81$ | | | $\text{cov}_{XY} = 5.74$ | | | | |
| $s_X = 3{,}015$ | | | $s_Y = 5.214$ | | | | | | | |

Şekilden de görebilceğiniz gibi, yaşa göre düzeltilmiş kalp hastalığından ölüm oranları yaşam temposunun hızlı olduğu şehirlerde -burada bireyler daha hızlı konuşmakta, daha hızlı yürümekte ve basit görevleri daha hızlı oranda gerçekleştirmekte- daha yüksek olma eğilimindedir. Örüntü önceki örneklerdeki kadar açık değildir fakat, birçok psikolojik değişkende görülebilen örüntülere benzerdir.

Şekil 9.2'nin incelenmesinde, yaşam temposu ve kalp hastalığı arasında dikkate değer pozitif bir ilişki olduğunu görebilirsiniz - yaşam temposu arttıkça kalp hastalığından ölüm oranı artar ve tam tersi de doğrudur. Burada **doğrusal ilişki** söz konusudur çünkü en iyi uyumu gösteren eğri doğrudur. (Bu veriye en iyi -veya neredeyse en iyi- bir doğru uyarsa doğrusal bir ilişki olduğu söylenir. Eğer en iyi uyum eğrisi doğru değilse **eğrisel ilişki**den söz edilebilir.) Burada ilişkiyi daha açık hale getirmek için doğruyu çizdim. Şekil 9.2'deki grafiğe bakınız. Eğer sadece en yüksek ve en düşük yaşam temposuna sahip bireylere bakarsanız, ölüm oranlarının en yüksek yaşam temposuna sahip bireylerde neredeyse iki katı olduğunu görürsünüz.

doğrusal ilişki

eğrisel ilişki

## 9.3 Stres ve Sağlık Arasındaki İlişki

Psikologlar uzun yıllardır stres ve sağlık arasındaki ilişki ile ilgilenmişlerdir ve stresin bireylerin fiziksel ve psikolojik sağlıklarının her ikisine de olumsuz etkisi olduğunu gösteren kanıtlar biriktirmişlerdir. Levine'in (1990) yaşam temposu üzerinde yaptığı çalışma buna iyi bir örnektir. Başka bir örneği ele alıp hesaplamalar ve kavramlar çerçevesinden bakacağız. Wagner, Compas ve Howell (1988) birinci sınıf üniversite öğrencilerinin, stres ve ruh sağlıkları arasındaki ilişkiyi incelemişlerdir. Son zamanlardaki yaşam durumlarının algılanan önemini ve çekiciliğini ölçmek için geliştirdikleri ölçeği kullanarak olumsuz olayların bildirilen sıklığı ve cevaplayıcının her bir olayın etkisine ilişkin subjektif tahminine göre ağırlıklandırılmış bir ölçüm gerçekleştirmişlerdir. Bu ölçümler bireylerin algılanan sosyal ve çevresel stresin ölçümü olarak kabul edilmiştir. Öğrencilerden ayrıca Hopkins Semptom kontrol listesinde yer alan 57 psikolojik semptomun varlığını ve yokluğunu değerlendirerek tamamlamaları istenmiştir. Stres ve semptom ölçümleri için gövde-ve-yaprak diyagramı ile Q-Q grafikleri Tablo 9.2'de görülmektedir.

Bu değişkenler arasındaki ilişkiyi düşünmeden önce, değişkenler üzerinde ayrı ayrı çalışmak ihtiyacındayız. Stres değişkeni için gövde-ve-yaprak diyagramı, dağılımın tek modlu ve hafif pozitif çarpık olduğunu göstermektedir. Birkaç uç değer hariç, bu değişkenle ilgili rahatsız edici bir durum gözlenmemektedir. Ancak Semptomların dağılımı (gösterilmedi) kesinlikle çarpıktır. Semptomlar keyfî bir ölçekte olduğundan, logaritmik dönüşümü almakta bir sorun yoktur. $\log_e$ Semptomları,[3] ölçeğin aşağılardan çok üst kısımlarından çekeceğiz ve dağılımı daha çok normal yapma eğiliminde olacağız. Bu yeni değişken lnSemptom olarak etiketlenecektir çünkü matematik ve istatistikte çoğu işlemde $\log_e$ göstermek için "ln" kullanılır. Tablo 9.2'de görülen Q-Q grafikleri her iki değişkenin normale yakın dağıldığını göstermektedir. Her bir değişkende makul bir değişkenlik olduğu unutulmamalıdır. Bu değişkenlik önemlidir; çünkü farklı stres puanlarının semptomlardaki farklılıkla ilgili olduğunu göstermek istersek, bu farklılıklara sahip olmak birinci sırada önemlidir.

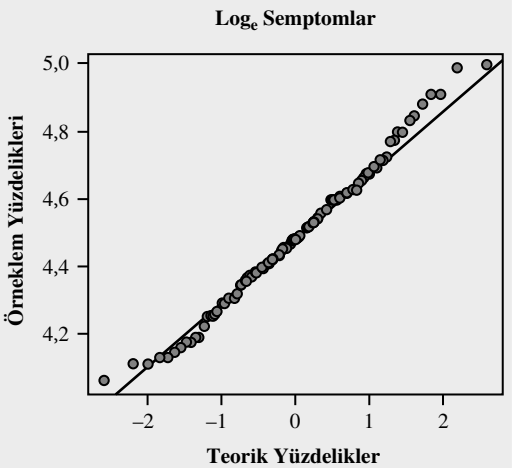
3 Log'lar herhangi bir tabanda kullanılabilir ancak istatistikte *e* tabanındaki doğal log'lar kullanılır. Taban seçiminin sonuçlarda önemli bir etkisi olmayacaktır.

**Tablo 9.2** Stres ve ruh sağlığı arasındaki ilişkiye ait verinin açıklaması

**Semptomlar**

Ondalık basamak virgülü, rakamların en solundaki basamaktan hemen öncedir.

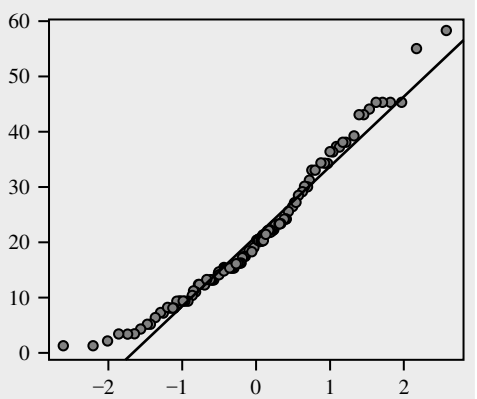
| | |
|---|---|
| 40 | 6 |
| 41 | 11334 |
| 41 | 67799 |
| 42 | 2 |
| 42 | 5556899 |
| 43 | 0000244 |
| 43 | 66677888999 |
| 44 | 111222334 |
| 44 | 555577888899 |
| 45 | 0111223344 |
| 45 | 55667 |
| 46 | 00001112222224 |
| 46 | 567799 |
| 47 | 112 |
| 47 | 67 |
| 48 | 0034 |
| 48 | 8 |
| 49 | 11 |
| 49 | 89 |




**Stres**

Ondalık basamak virgülü, rakamların en solundaki basamaktan hemen öncedir.

| | |
|---|---|
| 0 | 1123334 |
| 0 | 5567788899999 |
| 1 | 011222233333444 |
| 1 | 555555566667778889 |
| 2 | 0000011222223333444 |
| 2 | 56777899 |
| 3 | 0013334444 |
| 3 | 66778889 |
| 4 | 334 |
| 4 | 5555 |
| 5 | |
| 5 | 58 |




© Cengage Learning 2013

## 9.4 Kovaryans

**kovaryans (kov$_{XY}$ veya s$_{XY}$)**

Tablo 9.3'te yer alan veride hesaplanması incelenen korelasyon katsayısı[4], **kovaryans** (**kov$_{XY}$** veya **s$_{XY}$**) olarak adlandırılan istatistiğe dayanır. Kovaryans iki değişkenin birlikte değişkenliğini yansıtan bir değerdir.

Kovaryans matematiksel olarak aşağıdaki ifadeyle tanımlanır:

$$kov_{XY} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{N - 1}$$

4 Tüm veri kümesinin bir kopyası bu kitabın internet sitesinde Table-9.3.dat adlı dosyadadır.

**Tablo 9.3** 10 Temsili katılımcının stres ve semptom veri kümesi

| Katılımcı | Stres ($X$) | Semptomlar ($Y$) |
|---|---|---|
| 1 | 30 | 4,60 |
| 2 | 27 | 4,54 |
| 3 | 9 | 4,38 |
| 4 | 20 | 4,25 |
| 5 | 3 | 4,61 |
| 6 | 15 | 4,69 |
| 7 | 5 | 4,13 |
| 8 | 10 | 4,39 |
| 9 | 23 | 4,30 |
| 10 | 34 | 4,80 |
| ... | ... | ... |
| $\Sigma X =$ 2278 | | $\Sigma Y =$ 479,668 |
| $\Sigma X^2 =$ 65,038 | | $\Sigma Y^2 =$ 2154,635 |
| $\overline{X} =$ 21,290 | | $\overline{Y} =$ 4,483 |
| $s_X =$ 12,492 | | $s_Y =$ 0,202 |
| | $\Sigma XY =$ 10353,66 | |
| | N = 107 | |

© Cengage Learning 2013

Bu eşitlikten kovaryansın, form yönünden varyansın benzeri olduğu görülmektedir. Eşitlikteki bütün $Y$'ler yerine, $X$ yazılırsa $s_X^2$ ; $X$ 'ler yerine $Y$ yazılırsa $s_Y^2$ elde edilir.

Stres ve InSemptomlar veri kümeleri için, yüksek stres puanlarının yüksek semptom puanları ile eşleşeceğini bekleriz. Böylece, çok problemli ve stresli katılımcılar için hem $(X - \overline{X})$ ve $(Y - \overline{Y})$ hem de bunların çarpımları pozitif olacaktır. Düşük stres ve az sayıda problem bildiren katılımcılar için, $(X - \overline{X})$ ve $(Y - \overline{Y})$'nin her ikisi de negatif, ancak çarpımları yine pozitif olacaktır. Bu yüzden $(X - \overline{X})$.$(Y\text{-}\overline{Y})$'nin toplamları büyük ve pozitif olup, büyük pozitif kovaryans verecektir.

Bunun tersi durumda, güçlü negatif ilişki beklenir. Burada, $(X - \overline{X})$'nın büyük pozitif değerleri, büyük negatif $(Y - \overline{Y})$ değeriyle eşlecektir ve tam tersi de geçerlidir. Böylece, ortalamadan sapmaların çarpımlarının toplamı büyük ve negatif olacak, bu durum negatif güçlü ilişkiyi gösterecektir.

Son olarak $X$ ile $Y$ arasında hiçbir ilişkinin olmadığı durumları düşünelim. Bu durumda $(X - \overline{X})$'in pozitif değeri bazen $(Y - \overline{Y})$'nin pozitif, bazen de negatif değeri ile eşleşecektir. Ortalamadan sapmaların çarpımlarının yarısı yaklaşık pozitif, yarısı da negatif olacak ve sıfıra yakın toplam oluşturacaktır. Bu durum da, değişkenler arasında ilişkinin olmadığını gösterecektir.

Verilen bir veri kümesi için, $X$ ve $Y$, mükemmel (tam) korelasyon gösterebilir; böyle bir durumda $kov_{XY}$'nin değerinin pozitif maksimum, $X$ ve $Y$'nin mükemmel (tam) negatif ilişki gösterdiği durumda da $kov_{XY}$'nin değeri maksimum negatif olur. Eğer iki değişken mükemmel ilişkisiz ise ($r = 0{,}00$) $kov_{XY}$ sıfır değerini alır. Bu durum ispatlanabilir.

Kovaryansın hesaplanması için bir basit ifade aşağıdaki gibidir:

$$kov_{XY} = \frac{\Sigma (X - \overline{X})(Y - \overline{Y})}{N - 1} = \frac{\Sigma XY - \dfrac{\Sigma X \Sigma Y}{N}}{N - 1}$$

Tablo 9.2'de kısaltılmış formu verilen verinin tümü için kovaryansın hesaplanması aşağıdaki gibidir.

$$kov_{XY} = \frac{10353,66 - \dfrac{(2278)(479,668)}{107}}{106} = \frac{10353,66 - 10211,997}{106} = 1,336$$

## 9.5 Pearson Momentler Çarpımı Korelasyon Katsayısı *(r)*

Kovaryans için açıkladıklarımız, kovaryansın iki değişken arasındaki ilişki derecesinin ölçümü olarak kullanılabileceğini tavsiye eder. Fakat $kov_{XY}$'nin mutlak değerinin *X* ve *Y*'nin standart sapmasının fonksiyonu olmasından dolayı, ortaya güçlük çıkar. Bu nedenle örneğin $kov_{XY}$ = 1,336 değeri; standart sapmanın küçük olduğu durumlarda büyük korelasyon derecesini; standart sapmanın büyük olduğu durumlarda da küçük korelasyon derecesini yansıtabilir. Bu zorluğun üstesinden gelmek için kovaryans değeri değişkenlerin standart sapmalarına bölünür ve korelasyon tahmini yapılır. Bu aşağıdaki gibi tanımlanır:

$$r = \frac{cov_{XY}}{s_X s_Y}$$

$kov_{XY}$'nin maksimum değerinin $\pm s_X s_Y$ olduğu gösterilebildiğinden *r*'nin sınırları ±1,00'dir. Bu nedenle *r*, kovaryansın maksimum derecesine ulaşmasının ölçüsü olarak da yorumlanır.

Tablo 9.3 ve izleyen hesaplamalardan, $s_X$ = 12,492, $s_Y$ = 0,202 ve $kov_{XY}$ = 1,336 olduğu bilinmektedir. *X* ve *Y* arasındaki korelasyonun değeri,

$$r = \frac{cov_{XY}}{s_X s_Y}$$

$$r = \frac{1,336}{(12,492)(0,202)} = ,529$$

olur.

Bu katsayı dikkatli yorumlanmalıdır, sahip olmadığı anlamlar yüklenmemelidir. Özellikle *r* = ,53, stres ve semptomlar arasındaki ilişkinin %53'ü şeklinde *yorumlanmamalıdır*. Korelasyon katsayısı basitçe ölçekte +1 ve –1 arasında bir noktadır ve bu sınırlardan herhangi birine ne kadar yakınsa, iki değişken arasındaki ilişki de o kadar güçlüdür. Daha belirli bir yorum için kısaca tartışılacak $r^2$ değerinden söz edebiliriz. Korelasyonun işaretinin yalnızca ilişkinin yönünü ve muhtemelen ölçeğin keyfî belirlenen doğasını yansıttığını tekrar belirtmekte fayda vardır. "Doğru madde sayısı" değişkeni "yanlış madde sayısına" değiştirilirse, korelasyonun işareti değişecek fakat mutlak değerinde herhangi bir etkiye neden olmayacaktır.

### Düzeltilmiş *r*

**ρ(ro)**

**evren için korelasyon katsayısı**

Hesapladığımız ve normal olarak rapor ettiğimiz korelasyon, **ρ(ro)** ile gösterilen **evren için korelasyon katsayısı**nın yansız bir kestiricisi değildir. Neden böyle olduğunu görmek için tesadüfî olarak seçilmiş örneğin (23,18) ve (40,66) olan iki nokta çiftini ele alalım. Bu noktalar grafikte seçilip bir doğru ile birleştirildiğinde, doğru mükemmel uyum göstere-

cektir, çünkü ilkokulda da öğrenildiği gibi, iki noktadan bir doğru geçer. Doğru mükemmel uyum gösterdiği için, noktalar tesadüfi seçilmesine rağmen korelasyon 1,00 olacaktır. Açıkçası korelasyonun 1,00 olması, noktaların çizildiği evrende korelasyonun 1,00 ya da ona yakın olduğu anlamına gelmez. Gözlemlerin sayısı az olduğunda, örnekleme korelasyonu evrendeki korelasyon katsayısının yanlı bir kestiricisi olacaktır. Bunu düzeltmek için **düzeltilmiş korelasyon katsayısını ($r_{\text{düzeltilmiş}}$)** hesaplayabiliriz.

**düzeltilmiş korelasyon katsayısı ($r_{\text{düzeltilmiş}}$)**

$$r_{\text{düzeltilmiş}} \sqrt{1 - \frac{(1 - r^2)(N - 1)}{N - 2}}$$

Bu, bağıl olarak evrene ait korelasyon katsayısının yansız bir kestiricisidir.

Örneğimizde, örneklem büyüklüğü oldukça büyük alınmıştır ($N$ = 107). Bu nedenle $r$ ve $r_{\text{düzeltilmiş}}$ arasında çok büyük bir fark beklenmemelidir.

$$r_{\text{düzeltilmiş}} = \sqrt{1 - \frac{(1 - ,529^2)(106)}{105}} = ,522$$

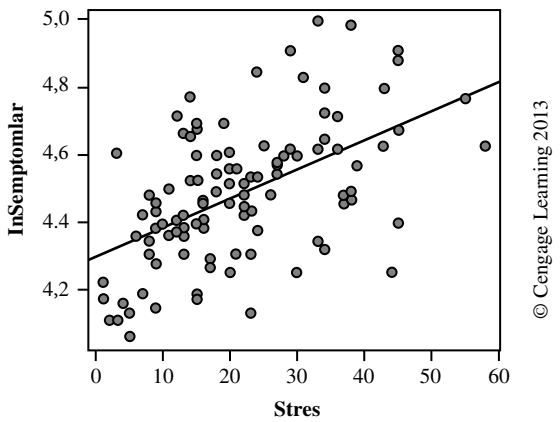
olup $r$ = ,529 değerine oldukça yakındır. Çok küçük örneklemlerde bu uyum söz konusu olmayacaktır.

Bölüm 15'te *Y*'nin çoklu yordayıcılarını içeren çoklu regresyon ele alındığı zaman, düzeltilmiş korelasyon için olan bu eşitliğin geçerli olmaya devam ettiğini göreceğiz. Tek fark $p$ yordayıcı sayısını belirtmek üzere payda da $N$-$p$-1 olacaktır. (Yukarıdaki eşitlikte yer alan $N$-2 buradan gelir.)

Düzeltilmiş $r$ ile örneklem varyansını hesaplama yöntemi arasında benzerlik kurabiliriz. Daha önce açıkladığım gibi, varyansı hesaplarken evren varyansının yansız kestirimini elde etmek için sapmaların karesinin toplamını $N$ – 1'e böleriz. Bu düzeltilmiş $r$'nin hesaplanmasındaki işlem ile kıyaslanabilir. Burada garip olan, hiç kimse evren varyansının yansız kestiriminden başka bir şey raporlamayı düşünmezken, evren korelasyon katsayısının yanlı kestirimiyle ilgili bir şey düşünmemektedir. Neden tutarsız davranıldığı konusunda bir fikrim yoktur, yalnızca yapılmaktadır. Düzeltilmiş değerin burada ele alınmasının tek sebebi, birçok bilgisayar yazılımının her iki istatistiği de sunması ve öğrencilerin çoğunlukla aradaki farkı ve hangisinin dikkate alınacağını merak etmesidir. Pratikte tüm amaçlar için istenilen, düzeltilmemiş korelasyon katsayısıdır.

## 9.6 Regresyon Doğrusu

Stres ve psikolojik semptomlar arasında mâkul derecede pozitif ilişki olduğu görülmüştür ($r$ = ,529). İki değişken arasındaki dağılımın grafiğine ve strese ($X$) bağlı olarak semptomların ($Y$) yordanabileceği regresyon doğrusuna bakarak ilişkinin nasıl olduğu hakkında daha iyi bir fikir elde edebiliriz. *X*'e bağlı olarak *Y*'nin yordanmasında en iyi uyumu gösteren doğrunun çizildiği dağılım grafiği Şekil 9.3'te gösterilmiştir. Bu doğrunun nereden geldiği kısaca açıklanacaktır, ancak öncelikle dikkat edilmesi gereken ilk şey, semptom puanlarının logaritmalarının (loglarının) stres puanlarının artışı ile doğrusal olarak arttığıdır. Korelasyon katsayısı ilişkinin olduğunu belirtir, ancak bunun ne anlama geldiği grafik olarak gösterildiği zaman daha kolay anlaşılır. Düşük stres puanlarından yüksek puanlara doğru çıkıldıkça regresyon eğrisi (doğrusu) etrafındaki noktaların dağılımının derecesinin aynı olduğuna dikkat edilmelidir, fakat yaklaşık ,50 ortalama korelasyon değeri ile dağılım oldukça geniştir. Buradaki yöntemlerin dayandığı varsayımları ele aldığımızda dağılımı daha ayrıntılı olarak inceleyeceğiz.




**Şekil 9.3** Stresin bir fonksiyonu olarak $Log_e$ (Semptomlar) grafiği

Lise derslerinden de hatırlayacağınız gibi, doğrunun denklemi $Y = bX + a$ formundadır. Bu denklemi amacımız doğrultusunda;

$$\hat{Y} = bX + a$$

ifadesiyle yazacağız. Burada;

$\hat{Y}$ = $Y$'nin yordanan değeri

**eğim**

$b$ = regresyon doğrusunun **eğim**i ($X$'teki bir birim değişime karşı gelen $Y$'deki değişim miktarı.)

**sabit**

$a$ = **sabit** ($x = 0$ olduğunda $Y$'nin değeri)

$X$ = yordayıcı değişkenin değeridir.

Buradaki amacımız, doğrusal (lineer) fonksiyona en iyi uyacak $a$ ve $b$ değerlerinin bulunması olacaktır. Diğer bir ifade ile eldeki verimizin, regresyon doğrusunun (farklı $X$ değerleri için $\hat{Y}$ değerleri) elde edilen gerçek $Y$ değerine mümkün olduğunca yakın olmasını sağlayacak olan $a$ ve $b$ değerlerinin elde edilmesi için kullanılması olacaktır. Fakat "en iyi uyum gösteren" ifadesi nasıl açıklanır? Bunun mantıklı bir yolu, **yordama hataları** –($Y - \hat{Y}$) sapmaları– açısından açıklamaktır. Çünkü, $\hat{Y}$ semptom (lnSempton) denklemimizin belli bir stres düzeyi için *kestireceği* değişken ve $Y$ *elde edilen* değer olduğu için ($Y - \hat{Y}$) genellikle **artık** olarak adlandırılan yordama hatasıdır. Bu hataları en aza indirecek doğruyu ($\hat{Y}$ değerleri kümesini) elde etmek istiyoruz. Sadece hataların *toplamını* en aza indirgeyemeyiz, çünkü sonsuz sayıda doğru vardır –($\overline{X}$, $\overline{Y}$) noktasından geçen herhangi bir doğru– bu toplam her zaman sıfır olacaktır. (Bazıları hedefin daha üzerinde, diğerleri de altında yer alacaktır.) Bunun yerine, hataların karesinin toplamını en aza indirgeyecek doğru aranacaktır –bu $\sum(Y - \hat{Y})^2$'ni en aza indirger. (Bölüm 2'de varyans tartışılırken benzer bir durum ifade edilmiştir. Bölüm 2'de ortalamadan sapmalar tartışılırken, burada regresyon doğrusundan sapmalar (kayan ya da değişen ortalamalar) tartışılmıştır. Yordama ve varyans hataları kavramlarının, görülebileceği gibi ortak yanları mevcuttur.)[5]

**yordama hataları**

**artık**

$a$ ve $b$'nin ideal değerleri, $\sum(Y - \hat{Y})^2$ toplamını en aza indirecek $a$ ve $b$ değerlerinin bulunmasıyla elde edilir. Çözüm zor değildir ve isteyenler bu kitabın önceki baskılarında

5 İlgilenenler için Rousseeuw ve Leroy (1987) en aza indirebilecek alternatif kriterin iyi bir tartışmasını sunmaktadır.

**normal denklemler**

veya Draper and Smith'te (1981, s. 13) bulabilirler. Bu problemin çözümü **"normal denklemler"** ile sağlanır.

$$a = \overline{Y} - b\overline{X}$$

$$b = \frac{kov_{XY}}{s_X^2}$$

$\sum(Y - \hat{Y})^2$ değerini en küçük yapacak $a$ ve $b$[6] değerleri için eşitlikleri biliyoruz Bunu belirtmek için, $X$'ten $Y$'yi kestirirken hataları en aza indirmek için tasarlandığını göstermek amacıyla, bazen sabitler $a_{y \cdot x}$ ve $b_{y \cdot x}$ olarak ifade edilir. Eğer bir karışıklık olmayacaksa alt indisler genellikle ihmal edilir. (Eğer amaç $Y$'ye bağlı olarak $X$'i kestirmek ise bu durumda bir önceki eşitlikte $X$ ve $Y$ yer değiştirir.)

Regresyon katsayısının hesaplanmasına örnek olarak Tablo 9.3'te yer alan veriyi düşünün. Tablodan $\overline{X} = 21{,}290$, $\overline{Y} = 4{,}483$ ve $s_X = 12{,}492$ olduğu bilinmektedir. Ayrıca $\text{kov}_{XY} = 1{,}336$ olduğu bilinmektedir. Böylece,

$$b = \frac{\text{cov}_{XY}}{s_X^2} = \frac{1{,}336}{12{,}492^2} = 0{,}0086$$

$$a = \overline{Y} - b\overline{X} = 4{,}483 - (0{,}0086)(21{,}290) = 4{,}300$$

$$\hat{Y} = bX + a = 0{,}0086(X) + 4{,}300$$

$Y$'nin $X$'e bağlı regresyon doğrusunun grafiğini Şekil 9.3'te daha önce görmüştük. Bu eşitlik, o doğrunun denklemidir.[7]

Regresyon doğrusunun nasıl çizileceği burada açıklanacaktır. Doğruyu çizmek için $X$'in herhangi iki değeri alınır, her ikisi için $\hat{Y}$ değeri hesaplanır, bu koordinatlar şekil üzerinde işaretlenir ve bir doğru ile birleştirilir. Eldeki veri için doğrunun denklemi şu şekildedir:

$$\hat{Y}_i = (0{,}0086)(X_i) + 4{,}300$$

$X_i = 0$ için,

$$\hat{Y}_i = (0{,}0086)(0) + 4{,}300 = 4{,}300$$

$X_i = 50$ için,

$$\hat{Y}_i = (0{,}0086)(50) + 4{,}300 = 4{,}730$$

Şekil 9.3'te gördüğümüz gibi doğru ($X = 0$, $Y = 4{,}300$) ve ($X = 50$, $Y = 4{,}730$) noktalarından geçer. Ayrıca regresyon doğrusu (0, $a$) ve ($\overline{X}$, $\overline{Y}$) noktalarından geçer. Bu, doğruluğun kontrolü için kullanılır.

Eğer iki regresyon doğrusu hesaplanırsa, ($X$'e bağlı $Y$ ve $Y$'ye bağlı $X$), iki doğrunun çakışık olmadığı görülecektir. Bu doğrular ($\overline{X}$, $\overline{Y}$) noktasında kesişecek, fakat farklı eğimlere sahip olacaklardır. Farklı doğrular olduğu gerçeği, farklı amaçlar için tasarlandığı gerçeğini yansıtır, biri $\sum(Y - \hat{Y})^2$ değerini, diğeri $\sum(X - \hat{X})^2$ değerini en aza indirir. Bir değişkende ortalama değere sahip bir bireyin diğer değişkende de ortalama değer alması beklendiğinden doğrular ($\overline{X}$, $\overline{Y}$) noktasında kesişir, fakat doğrular sadece iki değişken arasındaki korelasyon ±1,00 olduğunda çakışır.

---

6 $b$ için alternatif bir formül $b = r(s_y/s_x)$ olarak yazılabilir. Bu regresyon doğrusu ile korelasyon katsayısı arasındaki ilişkiyi açıkça gösterir. $s_y = s_x$ olduğunda $b$, $r$'ye eşitir. Bu, değişkenler standartlaştırıldığında, her iki değişkenin de standart sapması 1 olur.

7 Gary McClelland veriye ait noktaları girerek, regresyon doğrusu üzerindeki etkilerini görebileceğiniz harika bir java uygulaması hazırlamıştır. Bu uygulama http://www.uvm.edu/~dhowell/fundamentals7/SeeingStatisticsApplets/CorrelationPoints.html adresinde bulunabilir.

## Regresyonun Yorumları

Belirli durumlarda regresyon doğrusu kendi başına kullanışlıdır. Örneğin, üniversite kabul yönetimi, öğrencilerin lise ortalama başarı puanlarına dayalı olarak üniversite performansını yordayan bir eşitlik ile ilgilenebilir (Bölüm 15'te tartışıldığı gibi, yönetim birden fazla yordayıcı ile ilgilenmek istemesine rağmen). Benzer şekilde bir nöropsikolog, hastanın bir veya daha fazla yordayıcı değişkene dayalı olarak cevap oranının yordanması ile ilgilenebilir. Eğer gerçek oran beklenenin altında ise, bu hastanın sağlığı hakkında endişelenmeye başlayabiliriz (Crawford, Garthwaite, Howell ve Venneri, 2003). Bu örnekler biraz sıra dışıdır. Psikolojide regresyonun birçok uygulamasında, gerçeğin yordanması ile çoğu zaman ilgilenilmez. Her ne kadar aile bütçesi ve eğitim başarısı arasındaki ilişkinin bilinmesi ile ilgilenilse de, herhangi bir çocuğun eğitim başarısının kestiriminde aile bütçesi ölçümünü kullanmak pek mümkün değildir. Bireysel yordamalardan çok, genel ilkeler ile daha çok ilgileniriz. Gerçekte regresyon eşitliği, belirli durumlar için yordamada kullanılmasa da genel ilkeler hakkında fikir verir. (Bu bölümün sonunda çarpıcı bir örnek göreceksiniz.)

## Sabit Terim

Sabit terim, $X$'in değeri sıfır olduğunda $\hat{Y}$ değerindeki artış olarak tanımlanmıştı. Bu hal, sadece bazı durumlar için geçerlidir, diğer durumlarda değil; öncelikle, $X = 0$ değeri civarında veya $X$'in tanımlanan değerlerinde $Y$ eksenini kesim noktasının tahmininde kullanılır. Eğer, örneğin, bir grup fazla kilolu insanın benlik saygıları ($Y$) ve verdikleri kiloları ($X$) (aradaki ilişkinin doğrusal olduğunu varsayarak) dikkate alırsak, sabit terim bize, hangi düzeydeki benlik saygısı için bireyin 0 pound vermesi (ağırlığının aynı kalması) gerektiğini belirtir. Sabit terimin bundan başka bir yorumu yoktur, daha çok matematikseldir. Eğer yetişkinlerin benlik saygıları ($Y$) ile kiloları ($X$) arasında ilişkiye bakmak istiyorsak kilosundaki artış sıfır olduğunda benlik saygısının ne olacağını sormak gereksiz bir sorudur. Mevcut veri kümesinden böyle bir çıkarım anlamsızdır. (Bu durumda, sıfırdan farklı bir sabit terim bize geniş bir kilo aralığında -0 ile 300 pound aralığında- doğrusallığın olmayabileceğini tavsiye eder, fakar biz zaten uç değerlerde doğrusallığın olmayabileceği ile ilgilenmeyiz.)

Birçok durumda, her $X$ değerinden, $X$ değerlerinin ortalamasını çıkararak veriyi merkezlendirmek faydalıdır. Bunu yaparsanız, 0 olan bir $X$ değeri, $X$'in ortalaması ve sabit terim burada $X$ kendi değerlerinin ortalamasında iken $Y$'nin kestirilen değeri olur. Bu fikir göz ardı edilmemelidir. $X = 0$'ın anlamlı bir yorumunun olabilmesi için bir değişkeni dönüştürme fikri sadece regresyon problemlerinde değil, istatistikte geniş uygulanabilirliğe sahiptir. Bir değişkeni bazı noktalara merkezlemenin eğim ya da korelasyon katsayısı üzerine bir etkisi yoktur.

## Eğim

Eğim, $X$'teki bir birimlik değişim için $\hat{Y}$'daki değişim olarak tanımlanmıştı. Aslında, $Y$'nin yordanan *değişim oranının* bir ölçüsüdür. Tanıma göre eğim, çoğunlukla anlamlı bir ölçümdür. Eğer eğitim süresi ile gelirin regresyonuna bakılırsa, eğim bize her eklenen (sonraki) okul yılına ait, gelirdeki değişimin ne kadar ilişkili olduğunu söyleyecektir. Benzer şekilde; bir mühendis, galon başına mil olarak yakıt ekonomisinin (mpg) otomobilin ağır-

lığı ile ilişkisinin eğiminin 0,01 olduğunu biliyorsa ve mpg ile ağırlık arasında nedensel bir ilişki olduğunu varsayarsa, azaltabileceği her pound başına, yakıt ekonomisini 0,01 mpg arttırabileceğini bilir. Böylece üretici 30 poundluk yedek lâstik yerine 20 poundluk yedekle değiştirirse, otomobilin 0,1 mpg kazanacağı kestirilebilir.

## Standartlaştırılmış Regresyon Katsayıları

**standartlaştırılmış regresyon katsayısı**

**$\beta$ (beta)**

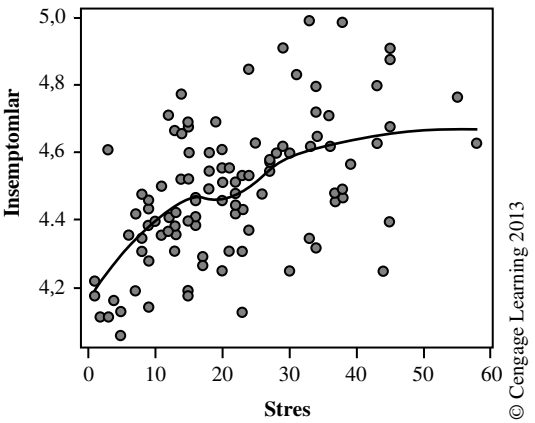
Standartlaştırılmış (her değişkenin ortalaması sıfır ve standart kayması bir) veri ile nadiren çalıştığımız halde, verideki her değişkenin standartlaştırılmış halleri için $b$'nin neyi temsil edeceğinin bilinmesi düşünülmeye değerdir. $X$ veya $Y$ değişkenindeki bir birimlik fark bir standart sapma farkı temsil edecektir. Böylece, eğer standartlaştırılmış veride eğim 0,75 ise, $X$'teki bir standart sapmalık artış, $\hat{Y}$'nın değerinde 3 çeyrek standart sapmalık artışa neden olacaktır. Standartlaştırılmış veri için eğim konuşulduğunda, standartlaştırılmamış veri için kullanılan $b$ katsayısından ayırmak için, $b$ yerine **standartlaştırılmış regresyon katsayısı** olan $\beta$ (**beta**)'yı kullanırız. Standartlaştırılmış değişkenler düşüncesine Bölüm 15'te çoklu regresyonu tartışırken döneceğiz. Hoş olan şey, standartlaştırılmış katsayıları veriyi standartlaştırmadan hesaplayabileceğimizdir. (Değişkenler standartlaştırıldığında sabit değer ne olacaktır?)

## Korelasyon ve Beta

Standart değişkenlerde eğim hakkında gördüğümüz her şey korelasyon katsayısı için de doğrudan uygulanabilir. $b$ değeri $\text{kov}_{XY}/s_x^2$ iken $r$ değerinin $\text{kov}_{XY}/s_X s_Y$ olarak tanımlandığını hatırlayınız. Eğer veri standartlaştırılmış ise $s_x = s_Y = s_x^2 = 1$ olup eğim ve korelasyon katsayıları birbirine eşit olacaktır. Bu nedenle tek bir yordayıcı değişkenin olduğu durumda değişkenler standartlaştırılmış ise, korelasyon katsayısının bir yorumu eğimin alacağı değere eşit olacağıdır. Örneğin $r$ = ,80'nin türetilmiş yorumu, $X$'de bir standart sapmalık değişimin $Y$'de ortalama 8/10 kadar bir standart sapmalık değişim ile ilişkili olduğunu gösterir. Bazı durumlara benzer yorumlar kullanışlı olarak uygulanabilir.

## Dikkat İçin Bir Not

b ve r'nin yorumu hakkında söylenenler biraz dikkatle ele alınmalıdır. Aile gelirindeki bir birimlik değişim ile birlikte akademik başarıdaki 0,75 birimlik değişim, Mary Smith için "aile gelirindeki artış otomatik olarak akademik başarıyı da artıracaktır" anlamında yorumlanamaz. Diğer bir deyişle, burada sebep ve sonuçtan söz edilmemektedir. Gelir değişkeni yüksek olan bireylerin neden-sonuç ilişkisi olmadan ya da aile geliri artırılsa idi ne olacağı iması olmadan başarı puanlarının da yüksek olacağı söylenebilir. Aile geliri (korelasyon durumunda) birçok diğer değişkenle ilişkilidir (eğitime karşı tutum, evdeki kitap sayısı, ortam çeşitliliğine erişim) ve sadece gelir değiştiğinde tüm bu değişkenlerde değişim beklemek için bir neden yoktur. Yoksulluğu gidermenin, bireylerin yaşamında çok büyük değişimlere yol açabileceğini savunan birisi genellikle sebep-sonuç tuzağına takılır. Yoksulluğu gidermek kesinlikle anlamlı ve önemli bir hedefdir, ancak gelir ve eğitim başarısı arasındaki korelasyon tümüyle bu konunun dışında olabilir.




**Şekil 9.4** Stresin fonksiyonu olarak lnSemptomların düzeltilmiş regresyon doğrusu eklenmiş dağılımının grafiği

## 9.7 Veriye Uygun Bir Doğrunun Çizilmesinin Diğer Yolları

**grafiği düzgünleştiriciler**

**splines**

**loess**

Veri kümesine uygun bir doğru grafiğinin çizilmesi çok yaygın ve veride ne olup bittiğinin anlaşılması için faydalı bir yol iken, başka alternatifler de vardır. İlişkinin bir noktaya kadar eğrisel olduğunu varsayınız - belki bir miktar artacak ve sonra düşecektir. Bu durumda eğrisel bir çizgi veriye en iyi uyumu gösterebilir. Veriye bir çizgi uydurmanın birçok yolu vardır ve bunların birçoğu **grafiği düzgünleştiriciler** başlığı altında incelenir. Birçok düzgünleştirme yöntemi **splines** ve **loess** başlıkları altında yer alır ve daha uzmanlaşmış yayınlarda ele alınır. Genellikle düzgünleştirme (smoothing) yordayıcının hedef değerine yakın *Y* değerlerinin ortalaması ile elde edilir. Başka bir ifade ile grafik üzerinde gidilerek elde edilen doğrular çizilerek elde edilir (bakınız Everitt, 2005). Şekil 9.4'te düzgünleştirilmiş bir grafiğe örnek görülmektedir. Bu grafik R kullanılarak oluşturulmuştur, fakat benzer bir grafik SPSS kullanılarak ve istenilen dağılım grafiği tanımlanıp, Fit panel tıklanılarak üretilebilir. Düzgünleştirilmiş doğruyu kullanmanın bir avantajı da ilişkinin tüm formu hakkında daha iyi fikir vermesidir. Eldeki veride verilen değişkenlik miktarı için, düzgünleştirilmiş grafiğin doğru grafiğinden daha iyi bir uyum verip vermediğini söylemek zordur, fakat semptomların stres düzeyi ile arttığını varsaymak mantıklıdır, ancak bu artış bir noktadan sonra düzelmeye başlayabilir.

## 9.8 Yordamanın Doğruluğu

Bir veri kümesine bir regresyon doğrusunun uydurulabilir olması, problemin çözüldüğü anlamına gelmez. Aksine yeni başlanmıştır. Önemli nokta veri boyunca düz bir doğrunun çizilip çizilmediği değil, bu doğrunun veriye mantıklı uyum sağlayıp sağlamadığıdır diğer bir deyişle sarfedilen gayretin bu işe yarayıp yaramadığıdır.

Yordama hatası tartışmasının başlangıcında *X*'in değeri bilinmeksizin *Y*'nin yordanmasının istendiği durumun ele alınması öğretici olacaktır.

### Bir Hata Ölçüsü Olarak Standart Sapma

Önceden de söz edildiği gibi, Şekil 9.3'te grafiği çizilen veri, stresli yaşam durumları sayısının (*X*) fonksiyonu olarak öğrencilerde gözlenen semptomların sayılarının log(*Y*) değelerini gösterir. Size özellikle gözlenen semptom sayısını tahmin görevinin verildiğini

varsayınız, ancak bireyin stresli yaşam durumu deneyimleri sayısı hakkında herhangi bir bilginiz yok. En iyi yordamanız lnSemptoms[8] ($\overline{Y}$) tüm katılımcılara ait ortalama sayısının ortalaması olabilir ve yordamadaki hata $Y$'nin standart sapması ($s_Y$) olacaktır çünkü sizin yordamanız ortalamadır ve $s_Y$ ortalama etrafındaki sapmayı belirtir. $s_Y$'nin

$$s_Y = \sqrt{\frac{\sum (Y - \overline{Y})^2}{N - 1}}$$

olduğu veya başka bir ifadeyle,

$$s_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N - 1}$$

olduğu bilinmektedir.

**Y'nin karelerinin toplamı ($KT_Y$)**

Burada pay $\overline{Y}$'den (bu örnekte tahmin edilen nokta) sapmaların karelerinin toplamıdır ve ***Y'nin karelerinin toplamı* ($KT_Y$)** olarak ifade edilecektir. Payda ise serbestlik derecesidir. O halde;

$$s_Y^2 = \frac{KT_Y}{sd} \quad \text{ve} \quad s_Y = \sqrt{\frac{KT_Y}{sd}}$$

yazılabilir.

## Yordamanın Standart Hatası

Belirli bir sayıda stresli yaşam durumlarına sahip bir öğrenci için semptom sayılarını yordayacağınızı varsayınız. Eğer elde sonsuz büyüklükte bir örnekleme ait veri varsa, semptomlar için yordama, örneklemdeki belirli bir stress değerindeki tüm öğrencilerden elde edilen semptom değerlerinin ($Y$) ortalaması olur. Başka bir ifade ile şartlı ortalama $X$'in değerine şartlanmıştır. Sonsuz örnekleme sahip olunmadığından regresyon doğrusu kullanacağız. (Burada kısaca tartışılan tüm varsayımlar karşılandığında, $X$'in belirli değeri ile ilgili $Y$'nin beklenen değerleri regresyon doğrusunda yer alacaktır.) Buradaki durumda $X$'in ilgili değerleri ve regresyon eşitliği bilinmektedir ve en iyi yordama $\hat{Y}$ olacaktır. Önceki ölçümün hatası (standart sapma) doğrultusunda, şu anki yordamanın hatası yine $Y$'nin sapmasının bir fonksiyonu olacaktır, fakat bu durumda tahmin edilen nokta $Y$'den ziyade $\hat{Y}$'dır. Özellikle hatanın ölçüsünü şimdi aşağıdaki şekilde tanımlayabiliriz.

$$S_{Y \cdot X} = \sqrt{\frac{\sum (Y - \widehat{Y})^2}{N - 2}} = \sqrt{\frac{KT_{artık}}{sd}}$$

**yordamanın standart hatası**

**artık varyans**

**hata varyansı**

ve hataların toplamının sapması yine yordama ($\hat{Y}$) için alınmıştır. $\hat{Y}$ için sapmaların karesinin toplamı $KT_{\text{artık}}$ olarak gösterilir çünkü $Y$'yi kestirmede $X$'i kullandıktan sonra kalan değişkenliği ifade eder.[9] Burada $s_{Y.X}$ **yordamanın standart hatası** olarak adlandırılır. $s_{Y.X}$ simgesi $X$ değeri için kestirilen $Y$ değerinin standart sapmasını temsil eder. Bu en yaygın (her zaman en iyisi olmamasına rağman) yordama hatası ölçüsüdür. $s_{Y.X}^2$ **artık varyans** veya **hata varyansı** olarak adlandırılır ve evren için $\sigma_{Y.X}^2$ parametresininin yansız bir kestiricisi olarak görülebilir. Burada $N - 2$ *sd*'dir, çünkü regresyon doğrusunu tahminde iki serbestlik derecesi kaybedilmektedir. (Örnek veriden $a$ ve $b$'nin her ikisi de kestirilmektedir.)

8 "Semptom logları"nı sürekli tekrar etmek yerine, log dönüşümlü değer anlamında semptom kullanılacaktır.

9 Yordama hatalarının karesinin toplamı olduğu için, sıklıkla $KT_{\text{hata}}$ olarak da gösterilir.

**Tablo 9.4** Yordamanın standart hatasının doğrudan hesaplanması

| Denek | Stres ($X$) | Semptomlar ($Y$) | $\hat{Y}$ | $Y - \hat{Y}$ |
|---|---|---|---|---|
| 1 | 30 | 4,60 | 4,557 | 0,038 |
| 2 | 27 | 4,54 | 4,532 | 0,012 |
| 3 | 9 | 4,38 | 4,378 | 0,004 |
| 4 | 20 | 4,25 | 4,472 | – 0,223 |
| 5 | 3 | 4,61 | 4,326 | 0,279 |
| 6 | 15 | 4,69 | 4,429 | 0,262 |
| 7 | 5 | 4,13 | 4,343 | – 0,216 |
| 8 | 10 | 4,39 | 4,386 | 0,008 |
| 9 | 23 | 4,30 | 4,498 | – 0,193 |
| 10 | 34 | 4,80 | 4,592 | 0,204 |
| … | … | … | … | … |
| | | | $\sum (Y - \hat{Y}) = 0$ | |
| | | | $\sum (Y - \hat{Y})^2 = 3,128$ | |
| $s^2_{Y \cdot X} = \frac{\sum (Y - \hat{Y})^2}{N - 2} = \frac{3,128}{105} = 0,030$ | | | $s_{Y \cdot X} = \sqrt{0,030} = 0,173$ | |

© Cengage Learning 2013

şartlı dağılım

Özellikle sonsuz sayıda gözlem varsa, $X$'in verilen değeri için yordamanın, $X$'in o değeri ile ilişkili $Y$ değerlerinin ortalaması olabileceğini tavsiye etmiştim. Bu düşünce $s_{Y.X}$ değerinin ne olacağı hakkında yardımcı olacaktır. Eğer sonsuz sayıda örneklem var ise ve her $X$ değeri ile ilgili $Y$ değerlerinin varyansı hesaplanmışsa, bu varyansların ortalaması artık varyans, karekökü de $s_{Y.X}$ olacaktır. Belirli bir $X$'e ait $Y$ değerlerinin kümesi $Y$'nin **şartlı dağılımı** olarak adlandırılır; çünkü $X$'e göre belirli bir koşulu sağlayan $Y$ değerlerinin dağılımıdır. Buradaki standart sapmalar $X$'e koşulludur, çünkü $X$'in belirli değerlerine karşılık gelen $Y$ değerlerinden hesaplanmıştır. Bununla birlikte $Y$'nin standart sapması ($s_Y$) $X$'e koşullu değildir, çünkü $X$ değerlerine karşılık gelen değerlerden bağımsız olarak tüm $Y$ değerleri kullanılarak hesaplanmıştır.

Yordamanın standart hatasını elde etmenin bir yolu Tablo 9.4'te yapıldığı gibi, her bir gözlem için $\hat{Y}$ değerini hesaplamak ve sonra $s_{Y.X}$ değerini doğrudan elde etmektir. Standart hatayı bu yolla hesaplamak zahmetli ve gereksizdir. Neyse ki çok fazla daha basit yöntemler vardır. Bu sadece yordamanın standart hatasını elde etmeye değil aynı zamanda daha önemli konulara da bir yol gösterir.

## $r^2$ ve Yordamanın Standart Hatası

Aşağıda yer alan bölümün büyük bir çoğunluğunda, kareler toplamı ($KT$) için varyans ifadesi kullanılmayacaktır. Hatırlanacağı gibi varyans, ortalamadan sapmaların kareleri toplamının (genellikle kareler toplamı olarak bilinir) serbestlik derecesine bölünmesi ile elde edilir. Varyans ile ilgili problem aynı *sd* değerine sahip olmadıkça toplanabilir olmamasıdır. Kareler toplamı serbestlik derecesinden bağımsız olarak toplanabilir ve böylece kullanılmaya daha uygun bir ölçü haline gelir.[10]

Daha önce artık veya hata varyansını aşağıdaki gibi tanımlamıştık:

$$s^2_{Y \cdot X} = \frac{\sum (Y - \hat{Y})^2}{N - 2} = \frac{SS_{\text{artık}}}{N - 2}$$

10 Kitabın ilerleyen bölümlerinde, varyans tipi ölçümden bahsedileceği, ancak bunun varyans, kareler toplamı ya da benzer bir şey olup olmadığı ima edilmek istenmediği zaman, daha zayıf bir terim olan varyasyon kullanılacaktır.

Bu ifade cebirsel işlemlere başvurularak aşağıdaki gibi gösterilebilir:

$$s_{Y \cdot X} = s_Y \sqrt{(1 - r^2)\frac{N - 1}{N - 2}}$$

Büyük örneklemler için $(N - 1)/(N - 2)$ kesri aslında 1'dir ve istatistik kitaplarında olduğu şekliyle aşağıdaki gibi yazılır:

$$s^2_{Y \cdot X} = s^2_Y(1 - r^2)$$

veya

$$s_{Y \cdot X} = s_Y\sqrt{(1 - r^2)}$$

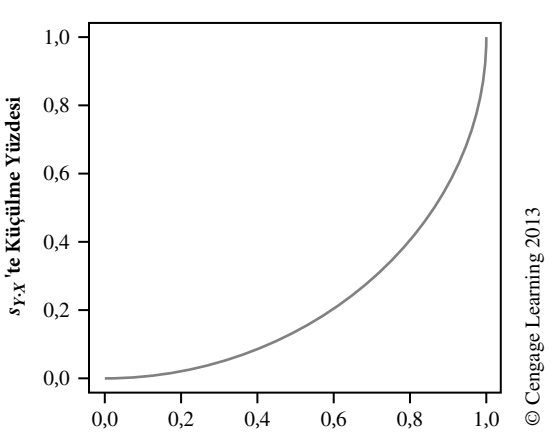
Küçük örneklemler için bu eşitliklerin sadece birer yaklaşık değer verdiğini aklınızda tutunuz ve $s^2_{Y \cdot X}$, $(N - 1)/(N - 2)$ kesri nedeniyle hata varyansını olduğunun altında kestirir. Buna rağmen herhangi bir örneklem büyüklüğünde, $KT_{artık} = KT_Y(1 - r^2)$'dir. Bu formül kitabın geri kalanı boyunca özellikle Bölüm 15 ve 16'da yer almaktadır.

## *r*'nin Bir Fonksiyonu Olarak Yordamanın Hatası

Yordamanın standart hatası için *r* cinsinden bir ifade elde ettik, *r* artarken bu hatanın nasıl azaldığını dikkate almak öğreticidir. Şekil 9.5'te, *r* değeri ,00'dan 1,00'e doğru artarken $s_{Y \cdot X}$ değerinin azaldığı görülmektedir.

Şekil 9.5'te yer alan değerler etkilerinden biraz ayrılmaktadır. Eğer *X* bilinmez ise ,20 korelasyon değeri ile yordamanın standart hatası, alacağı değerden %2 oranında azalmaktadır. Bu, *X*'in bilinmemesi durumunda, korelasyon değeri ,20 ise, yordama için $\overline{Y}$ (*X*'i dikkate alarak) yerine *Y*'nın kullanılması, standart hatanın alacağı değerin %98'i kadar hata verir anlamındadır. Daha hayal kırıklığı yaratan şey ise, *r*'nin buradaki örnekte olduğu gibi ,50 olması halinde yordamanın standart hatasının hâlâ standart sapmanın %87'si kadar olmasıdır.

X bilinmeden, hatayı, alacağı değerin yarısı kadar azaltmak için ,866 korelasyon gerekir ve ,95 korelasyon bile hatanın sadece 2/3'ünü azaltmaktadır. Tüm bunlar yordamaya dayalı bir regresyon eşitliğini kullanarak bir şey elde edilemeyeceği anlamına gelmez, sadece yordamanın dikkatle yorumlanması gerektiği anlamına gelir. Burada ilişkilerin büyüklüğünden ziyade türü ile ilgilenildi.




**Şekil 9.5** r'nin bir fonksiyonu olarak yordamanın standart hatası

## Yordanabilir Değişkenliğin Bir Ölçüsü Olarak $r^2$

Bir önceki eşitlikte ifade edilen $r^2$'nin artık hatasından, korelasyon katsayısı ile ilgili önemli bir yorum elde edilebilir. Daha önce gördüğümüz,

$$KT_{artık} = KT_Y(1 - r^2)$$

ifadesini genişleterek ve düzenleyerek,

$$KT_{\text{artık}} = KT_Y - KT_Y(r^2)$$

$$r^2 = \frac{KT_Y - KT_{\text{artık}}}{KT_Y}$$

elde edilir. Bu eşitlikte, $KT_Y$'nin $\sum(Y - \bar{Y})^2$ değerine eşit olduğunu ve $Y$'nin kareler toplamı olduğunu ve aşağıdakilerin toplamları temsil ettiğini biliyorsunuz.

1. $Y$'nin $X$ ile ilişkili kareler toplam kısmı [$KT_Y(r^2)$]
2. $Y$'nin $X$'ten bağımsız olan kareler toplam kısmı [$KT_{\text{artık}}$]

Buradaki örnek kapsamında, bireylerde gözlenen semptom sayılarının hem stresli yaşam durumları ile ilişkili kısmı hem de diğer şeylerle ilişkili kısmı hakkında konuşuldu. $KT_{\text{artık}}$ miktarı, $X$'ten bağımsız olarak $Y$'nin kareler toplamıdır ve $Y$'yi yordamak için $X$ kullanıldığında dahi kalan hata miktarının ölçüsüdür. Bu kavramlar ikinci örnekle daha açık hale gelecektir.

Sigara tüketim miktarı ($X$) ve ölüm yaşı ($Y$) arasındaki ilişki ile ilgilendiğimizi düşününüz. Bireylerin zamanla ölümleri izlenildiğinde birçok şey dikkat çekicidir. İlk olarak tüm ölümlerin tam olarak aynı yaşta olmadığını görürüz. Ölüm yaşında sigara tüketiminden bağımsız olarak değişkenlik söz konusudur ve bu değişkenlik $KT_Y = \sum(Y - \bar{Y})^2$ ile ölçülür. Ayrıca bazı kişilerin diğerlerinden daha çok içtiği de dikkat çekebilir. Bu sigara tüketimindeki değişkenlik ölüm yaşından bağımsız olarak $KT_X = \sum(X - \bar{X})^2$ ile ölçülür. Sigara içenlerin sigara içmeyenlerden ve çok sigara içenlerin az sigara içenlerden daha erken öldüğü bulunmuştur. Bu yüzden $X$'den $Y$'yi yordamak için bir regresyon eşitliği yazılabilir. Çünkü bireyler sigara içme davranışlarında ve ayrıca yordanan ortalama ömürlerinde ($\hat{Y}$) farklılaşacaklardır ve bu değişkenlik $KT_{\hat{Y}} = \sum(\hat{Y} - \bar{Y})^2$ olarak belirlenecektir. Bu son eşitlik $Y$'deki değişkenliktir ve doğrudan $X$'deki değişkenliğe atfedilebilir, çünkü $\hat{Y}$'nin farklı değerleri $X$'in farklı değerlerinden ve $\hat{Y}$'nin benzer değerleri $X$'in benzer değerlerinden ileri gelmektedir - böylece $\hat{Y}$, $X$ değişmedikçe değişmeyecektir.

Değişkenliğin son bir kaynağına daha sahibiz; bu aynı miktarda sigara içen bireylerin ortalama ömürlerindeki değişkenliktir. Bu $KT_{\text{artık}}$ ile ölçülür ve $X$'teki değişkenlik ile açıklanamayan $Y$'deki değişkenliktir (çünkü bu bireyler sigara içme miktarlarında farklılaşmamaktadırlar). Değişkenliğin birçok kaynağı (kareler toplamı) Tablo 9.5'te özetlenmiştir.

Sigara içmeyen tüm bireylerin 72 yaşında öldüğü ve aynı miktarda sigara içen tüm bireylerin tam 68 yaşında öldüğü sıra dışı bir örneği dikkate aldığımızda, ortalama ömürdeki tüm değişkenlik doğrudan sigara içme davranışındaki değişkenlik ile kestirilebilir. Eğer sigara içiliyorsa 68 yaşında, içilmiyorsa ise 72 yaşında ölünecektir. Burada $KT_{\hat{Y}} = KT_Y$, ve $KT_{\text{artık}} = 0$'dır.

Daha gerçekçi bir örnek olarak, sigara içen bireylerin içmeyenlerden daha erken ölmeye eğilimli olduğunu, fakat her bir grubun ömür beklentisi için kendi içinde belirli bir miktarda değişkenliğe sahip olduğunu varsayınız. Bu $KT_Y$'nin bir kısmının sigara içmeye $KT_{\hat{Y}}$'ye dayandırılabildiği ve bir kısmının ise dayandırılamadığı ($KT_{\text{artık}}$) bir durumdur.

**Tablo 9.5** Sigara içme ve ortalama ömür çalışması için regresyondaki varyans kaynakları

| | |
|---|---|
| $KT_X =$ | sigara miktarı değişkenliği = $\sum(X - \bar{X})^2$ |
| $KT_Y =$ | yaşam beklentisi değişkenliği = $\sum(Y - \bar{Y})^2$ |
| $KT_{\hat{Y}} =$ | sigara içme davranışı tarafından doğrudan açıklanan yaşam beklentisi değişkenliği = $\sum(\hat{Y} - \bar{Y})^2$ |
| $KT_{\text{artık}} =$ | sigara içme davranışı tarafından açıklanamayan yaşam beklentisi değişkenliği = $\sum(Y - \hat{Y})^2 = KT_Y - KT_{\hat{Y}}$ |

© Cengage Learning 2013

Yapabileceğimiz, yaşam beklentisindeki değişkenliğin tamamının ne kadarlık bir yüzdesinin sigara içme davranışına dayandırılabildiğini belirlemektir. Bir başka ifade ile

$$\frac{KT_{\hat{Y}}}{KT_Y} = \frac{KT_Y - KT_{\text{artık}}}{KT_Y}$$

ile gösterilen ölçümü istemekteyiz. Gördüğümüz gibi bu ölçüm $r^2$'dir. Diğer bir deyişle

$$r^2 = \frac{KT_{\hat{Y}}}{KT_Y}$$

olur.

$r^2$'nin bu yorumu oldukça faydalıdır. Örneğin, sigara miktarı ve ortalama ömür arasındaki korelasyon ,80 gibi gerçekçi olamayacak kadar yüksek olsa idi, ortalama ömürdeki $,80^2$ = %64'lük bir değişkenliğin doğrudan sigara içme davranışındaki değişkenlik tarafından kestirilebilir olduğunu söyleyebilirdik. (Açıkça bu, gerçeğin aşırı bir abartısıdır.) Eğer korelasyon daha çok, $r$ = ,10 gibi olsaydı ortalama ömürdeki $,10^2$ = %1'lik değişkenliğin sigara içme davranışı ile ilişkili olduğu diğer %99'luk kısmın ise diğer faktörlerle ilişkili olduğunu belirtebilirdik.

"Sebebidir", "dayandırılabilir", "yordanabilir" ve "ilişkilidir" gibi ifadeler sebep ve sonuç ifadeleri olarak yorumlanmamalıdır. Bu yüzden "Geçen yıl kırdığım ayak bileğimdeki sancıya dayanarak havadaki değişkenliğin %10'unu yordayabilirim - yağmurlu havalarda ağrımakta ve havalar düzeldiğinde ise iyi hissedilmekte." diyebilirsiniz. Bu ayak bileği ağrısının yağmura veya yağmurun ayak bileği ağrısına neden olduğuna işaret etmemektedir. Örneğin, yağmur yağdığı zaman, yağmurla ilişkili ve bir şekilde ağrıyı etkileyen düşük barometrik basınçtan dolayı ayak bileği ağrıyabilir.

Bu tartışmadan $r^2$, korelasyonun ölçümü olarak $r$'den daha kolay yorumlanabilirdir, çünkü bir ölçümdeki değişkenliğin diğer ölçümdeki değişkenliğe dayandırılabilirliğinin derecesini ifade etmektedir. Korelasyon katsayısının karesinin alınmasını önemli durumlarda bazı fikirler verebileceğinden daima tavsiye ederim.[11] Semptom ve stres örneğinde, $r^2 = ,529^2 = ,280$'dir. Bu yüzden semptomlardaki ¼'lük değişkenlik stresteki değişkenlikten yordanabilmektedir. Psikolojik semptomları etkileyen tüm faktörler düşünüldüğünde bu etkileyici bir yordama düzeyidir.

Bir değişkenin, diğer bir değişkenin yordanmasına katkısının ölçülmesinde $r^2$'nin en iyi ölçüm olduğuna evrensel bir uyuşma olmamasına rağmen en popüler ölçüm olduğu kesindir.

---

11 Robert Rosenthal gibi birçok saygıdeğer otorite, etki büyüklüğünün ölçümü olarak $r$'yi kullanırsanız, etkiyi gerçekte olduğundan daha az algılamanıza sebep olacağından karesini almamanızı savunur. Fakat, geleneksel olarak karesi alınır.

**hatadaki oransal azalma**

Judd ve McClelland (1989), $r^2$'yi şiddetle desteklemektedir, çünkü hata, hataların kareleri toplamı açısından sıralandığında, bu **hatadaki oransal azalma**dır **(HOA)**. Başka bir deyişle *Y*'yi yordarken **X** kullanılmazsa hata $KT_Y$'dir. Eğer kestirici olarak *X* kullanılırsa hata $KT_{\text{artık}}$'tır.

$$r^2 = \frac{KT_Y - KT_{artık}}{KT_Y}$$

olduğu için $r^2$'nin değeri, *X* yordayıcı olarak kullanıldığında, hatanın azalma oranı olarak görülebilir.[12]

**yordamada oransal gelişim (YOG)**

Fakat diğerleri, **yordamada oransal gelişimi (YOG)** daha iyi bir ölçüm olarak önerirler.

$$YOG = 1 - \sqrt{1 - r^2}$$

Geniş örneklemler için bu istatistik, kestirimin standart hatası büyüklüğündeki *azalmadır*. Benzer şekilde, görülebileceği gibi, bu yordama güven aralığı genişliğinin azalmasının ölçümüdür.

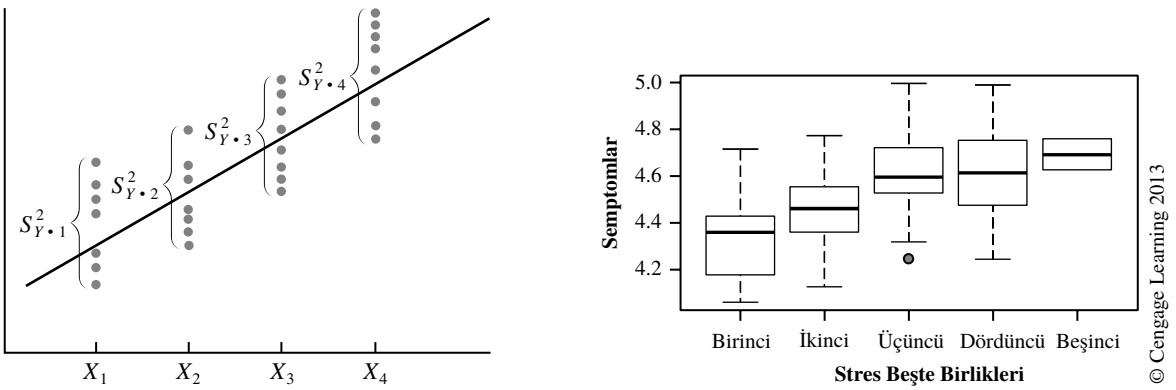
*r*, $r^2$ ve *YOG* arasındaki seçim hatayı nasıl ölçmek istediğinize bağlıdır. Eğer $r^2$'ye odaklanılırsa ölçmenin hatasına kareler toplamı açısından odaklanılıyordur. Eğer YOG'ye odaklanılırsa hata standart sapma birimlerinde ölçülüyordur.

Özel nedenler için değişkenlik yüzdesi indeksi olarak $r^2$ tartışıldı. Göreceğimiz gibi, pek çok alanda, en azından psikolojide, sadece test istatistiği ve *p* değerinden daha çok etki büyüklüğünü raporlamaya, çok güçlü bir eğilim vardır. Bölüm 7'de belirttiğim gibi, büyüklük ölçümünün 2 önemli tipi vardır. Birincisi etki büyüklüğü olarak adlandırılır ve çoğunlukla ölçmenin *d*-ailesine atfedilir. Cohen'ın *d*'si ile temsil edilir ve iki ya da daha fazla grubun ortalamalarının olması durumunda daha uygun bir istatistiktir. Ölçümünün ikinci tipi ise, sıklıkla *r*-ailesi olarak adlandırılır ve genellikle $r^2$ olarak temsil edilen değişkenliğin yüzdesidir. Anılan ölçümü, bu bölümde ilk olarak psikolojik semptomlardaki değişimin %25,6'sının stresteki değişim ile ilgili bulunduğunda gördük. Bölüm 10'da nokta çift serili korelasyonu işlerken tekrar göreceğiz. Varyans analizi bölümlerinde (özellikle Bölüm 11 ve 13'te) eta-kareler ve ilişkili ölçümlerde örtük olarak tekrar gündeme gelecektir. Sonuç olarak çoklu regresyon hakkında konuşulduğunda önemli bir yol olarak tekrar ortaya çıkacaktır. Tüm bunlarda ortak konu bağımlı değişkendeki değişkenliğin ne kadarının kategorik ya da sürekli olan bağımsız değişkendeki değişim ile ilgili olduğunu istiyor olmamızdır. Bazı kişiler gibi ölçmenin değişim yüzdesine bayılmıyorum, çünkü birçoğumuzun benzeri ölçümlerden anlam çıkarabileceğini düşünmüyorum. Buna rağmen yaygın olarak kullanılmaktadırlar ve bunlara aşina olunması gerektir.

## 9.9 Regresyon ve Korelasyon Altındaki Varsayımlar

Yordamanın standart hatasını ve diğer istatistikleri verinin çekildiği evren(ler)e ilişkin herhangi bir varsayım yapmadan türettik. Bunun gibi, $\sigma_{Y \cdot X}$'in yansız kestiricisi olarak $s_{Y \cdot X}$'i kullanmada da bazı varsayımlara ihtiyaç olmamıştır. Eğer $s_{Y \cdot X}$ herhangi bir şekilde kullanılacaksa bazı parametrik varsayımlardan bahsedilmelidir. Neden olduğunu anlamak için Şekil 9.6'da grafiği çizilmiş veri dikkate alınmalıdır. $s^2_{Y \cdot 1}$, $s^2_{Y \cdot 2}$, $s^2_{Y \cdot 3}$ ve $s^2_{Y \cdot 4}$ ile belirtilen dört

12 Bölüm 9.5'te tanımlanan $r^2_{\text{düzeltilmiş}}$, denklemdeki kareler toplamına karşılık gelen varyasyon teriminin oranına neredeyse eşittir.




**Tablo 9.6** a) Regresyon varsayımlarını gösteren grafik; b) Semptom ve Stresler için benzer grafik

istatistiğe dikkat ediniz. Her biri $X$ serisindeki regresyon doğrusu etrafında bulunan noktaların varyansını temsil eder ($Y$'nin artık varyansı belirli bir $X$'e koşulludur). Daha önce de belirtildiği gibi, bu varyansların ortalamaları, artık veya hata varyansı olarak, $s^2_{Y.X}$, her bir serideki serbestlik derecesi ile ağırlıklandırılmıştır. Eğer $s^2_{Y.X}$, ortalama olarak çeşitli terimleri temsil ederse pratik bir anlamı olur. Bu bizi, her bir $X$ değeri için $Y$'deki varyansın sabit olması (evren için) anlamına gelen **serilerdeki varyansın homojenliğ**inin varsayımına yönlendirir. Bu varsayım, $s^2_{Y.X}$'i kullanarak manidarlık testini uyguladığımızda önemli olmaya başlayacaktır.

**serilerdeki varyansın homojenliği**

Hipotez testlerine gelindiğinde bir zorunlu varsayım da **serilerde normallik**tir. Evrende belirli herhangi bir $X$ değerine ait $Y$ değerinin -$X_i$ için **Y'nin koşullu serisi**- $\hat{Y}$'nin etrafında normal dağıldığını varsayacağız. Bu varsayım $t$–testi ile yapılan normallik varsayımının -her bir işlem örnekleminin ortalamaları etrafında normal dağıldığı- doğrudan benzeridir ve benzer nedenlerle yapılır.

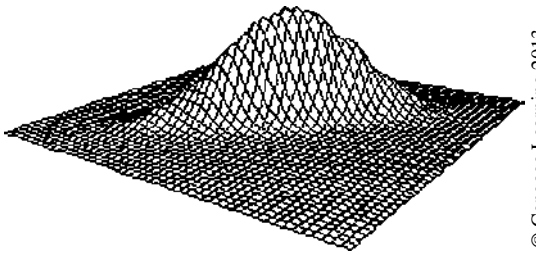
**serilerde normallik**

**koşullu seri**

Stres değişkenini yeniden beş kategoride veya beşte birlikte tanımlayarak var olan stres ve semptom verisi için, bu varsayımların uygunluğunu inceleyebiliriz. Stres değişkenin her bir kategorisi için lnSemptomlara ait kutu grafikleri çizilebilir. Bu grafik Şekil 9.6'da görülmektedir. Her bir kategoride yaklaşık 20 veri bulunmaktadır. Şekil 9.6b varsayımın uygunluğunu yansıtır.

Serilerdeki varyansın homojenliğinin ve normalliğin, farklı işlemlerle ilişkili evren varyansının homojenliğine ve normalliğine eşdeğer olduğunu, Bölüm 11'de varyans analizinde tartışacağız. Bölüm 11'de verinin çekildiği işlemlere ait evrenlerin dağılımlarının normal olduğu ve aynı varyansa sahip olduklarını varsayacağız. Şekil 9.6a ve 9.6b, farklı deney şartları ($X$'in farklı düzeylerinin farklı deneysel durumları temsil ettiği) düşünülürse, regresyon ve varyans analizi varsayımları arasındaki ilişkiyi görebilirsiniz.

Serilerdeki normallik ve varyansların homojenliği varsayımları $X$'in sabit değerlerinin söz konusu olduğu regresyon modeli ile ilişkilidir. Öte yandan $X$ ve $Y$ arasındaki korelasyon ile ilgileniyorsak, $X$ ve $Y$'nin her ikisinin tesadüfî değişken olduğu iki değişkenli model söz konusudur. Bu durumda temel olarak evren için korelasyon katsayısının ($\rho$) kestirimi olarak örneklem korelasyonu ($r$) ile ilgileniriz. Burada regresyon modeli varsayımları yerine iki değişkenli normal dağılımdan çekilen örneklemin varsayımlarını koyuyoruz.



© Cengage Learning 2013

**Tablo 9.7** İki değişkenli normal dağılım $r = ,90$

İki değişkenli normal dağılım kabaca damperli kamyonların köprü temeli inşaatı sırasında oluşturduğu toz yığınına benzer. Toz yığının etrafa yayılması normal dağılımı andırır. (Eğer $X$ ve $Y$ arasında korelasyon olmasaydı yığın, bütün toz yığının merkezine düşmüş gibi ve bütün yönlere simetrik yayılmış gibi görülecekti. Eğer $X$ ve $Y$ ilişkili ise yığın bir sokak boyunca dökülmüş ve boylu boyunca yayılmış gibi uzayacaktır.) $r = ,90$ olan iki değişkenli normal dağılımın bir örneği Şekil 9.7'de gösterilmiştir. Eğer bu dağılımı verilen herhangi bir $X$ değerine göre bir doğru ile dilimleseydik, son kesimin normal dağılım olduğunu görürdük. Eğer yığını verilen herhangi bir $Y$ değerine göre bir doğru boyunca dilimleseydik yine bir normal dağılım elde ederdik. Bunlar **koşullu normal dağılım** olarak adlandırılır, çünkü ilki, verilen belirli bir $X$ değeri için (şart) $Y$'nin dağılımını temsil ederken ikincisi, verilen belirli bir $Y$ değeri için $X$'in dağılımını temsil eder. Eğer bunun yerine $X$'ten bağımsız olarak $Y$'nin bütün değerlerine bakarsak (veya $Y$'den bağımsız olarak bütün $X$ değerlerine), $Y$'nin (veya $X$'in) **marjinal dağılım**ı elde ederiz. İki değişkenli normal dağılım için, şartlı ve marjinal dağılımların her ikisi de normal olacaktır. (Regresyon modeli için, şimdi $Y$'nin şartlı normalliği olduğu bilinen, $X$ serisinde $Y$'nin sadece normal dağıldığının varsayıldığı hatırlanmalıdır. Regresyon modeli için $X$'in şartlı ya da marjinal dağılımının normallik varsayımı yoktur.)

koşullu normal dağılım

marjinal dağılım

## 9.10 $\hat{Y}$'ya Ait Güven Sınırları

$\hat{Y}$'ya ait güven sınırlarını oluşturmadan önce neden buna ihtiyaç duyduğumuza karar vermeliyiz. Gelecekteki bazı katılımcıların stres düzeylerini belirledikten sonra semptom düzeylerinin yordanması için isteniyor olabilir. Alternatif olarak, regresyon doğrusunda güven sınırlarını belirlemek isteyebiliriz, bu aynı stres puanlarına sahip katılımcılar için ortalama semptom puanlarında güven aralığını belirlemeye denktir. Bunlar hemen hemen aynı sorularmış gibi görülebilir, fakat gerçekte değillerdir. Eğer gelecekteki bireyler için yordama yapacaksak $X$ için $Y$'nin ortalaması ile ilişkili hatasını ve ayrıca $Y$ değerlerinin varyansını elde edeceğiz. Bununla birlikte eldeki strese dayalı veriyi, belirli bir stres derecesine koşullu ortalama semptom puanlarını yordamak için kullanmak istersek başka hata kaynakları için endişe etmemize gerek yoktur. Bazıları önceki aralığı **yordama aralığı** ve sonrakini ise güven aralığı olarak adlandırmaktadır. Yordama aralığı hem $Y$'nin sabit değerine koşullu $Y$'nin ortalamasının belirsizliğini hem de bu ortalama etrafında gözlem değişkenliğini dikkate almalıdır.[13]

yordama aralığı

[13] Bu farklılığın güzel bir tartışması http://www.ma.utexas.edu/users/mks/statmistakes/CIvsPI.html. adresinde bulunabilir.

Yordamanın standart hatası tüm hataların ölçümünde kullanışlı olmasına rağmen, herhangi bir tek yordama ile ilişkili hatanın iyi bir kestiricisi değildir. $X$ değeri ortalamanın yakınında olan bir birey için $Y$ değerini yordamak istersek yordamadaki hata, $X$ değerinin $\overline{X}$'ten uzak olduğu durumdan daha düşüktür. (Bunu daha iyi anlayabilmek için $\overline{X}$, $\overline{Y}$ noktası hafifçe regresyon doğrusunu etrafında döndürdüğümüzde $X$'in farklı değerleri için yordamanın ne olabileceğini düşününüz. Ortalama yakınında ihmal edilebilir değişimler olabilirken uç değerlerde önemli değişimler olabilir.)

Evrenin *yeni* bir elemanı için (orjinal örneklemde yer almayan) $X$'e dayalı $Y$ için yordama aralığı belirlenmek istenirse, yordama için standart hatamız aşağıdaki eşitlikle verilir.

$$s'_{Y\cdot X} = s_{Y\cdot X}\sqrt{1 + \frac{1}{N} + \frac{(X_i - \overline{X})^2}{(N-1)s_X^2}}$$

Burada $X_i - \overline{X}$, bireyin $X$ puanının $X$'in ortalamasından sapma miktarıdır. Buradan yola çıkarak $\hat{Y}$ için yordama aralığı aşağıdaki ifadeyle verilir.

$$GA(Y) = \hat{Y} \pm (t_{\alpha/2})(s'_{Y\cdot X})$$

Bu ifade regresyon doğrusu etrafında, $X = \overline{X}$ için en dar olan ve $|X_i - \overline{X}|$ arttıkça artan eliptik güven sınırları sonucunu doğurur. (Şekil 9.6'da doğruların eliptik olduğunu göstermek için cetvele ihtiyaç olabilir ama gerçekten öyleler.)

Alternatif olarak, $X$'in belirli bir değerine koşullanmış ortalama $Y$'deki güven sınırları için, standart hata şu şekildedir:

$$s''_{Y\cdot X} = s_{Y\cdot X}\sqrt{\frac{1}{N} + \frac{(X_i - \overline{X})^2}{(N-1)s_X^2}}$$

ve bu aşağıda verilen güven aralığını doğurur.

$$GA(Y) = \hat{Y} \pm (t_{\alpha/2})(s''_{Y\cdot X})$$

Konu ile ilgili belirli bir örnek olarak, stres puanı 10 -stresin iyice düşük düzeyi- olan yeni bir katılımcı için semptom sayısı üzerinde ($Y$) güven sınırlarını belirlemek istediğinizi varsayınız. Burada

$$s_{Y\cdot X} = 0,173$$
$$s_X^2 = 156,05$$
$$\overline{X} = 21,290$$
$$\hat{Y} = 0,0086(10) + 4,31 = 4,386$$
$$t_{,025} = 1,984$$
$$N = 107$$

olduğu bilinmektedir. Buradan,

$$s'_{Y\cdot X} = s_{Y\cdot X}\sqrt{1 + \frac{1}{N} + \frac{(X_i - \overline{X})^2}{(N-1)s_X^2}}$$

$$s'_{Y\cdot X} = 0,173\sqrt{1 + \frac{1}{107} + \frac{(10 - 21,290)^2}{(106)156,05}}$$

$$= 0,173\sqrt{1,017} = 0,174$$

bulunur.

Buradan da,

$$GA(Y) = \hat{Y} \pm (t_{\alpha/2})(s'_{Y \cdot X})$$
$$= 4,386 \pm 1,984(0,174)$$
$$= 4,386 \pm ,345$$
$$4,041 \leq Y \leq 4,731$$

elde edilir. Yordama aralığı 4,041 - 4,731'dir ve ,95 olasılıkla bu şekilde hesaplanan aralık o birey tarafından bildirilen semptomların düzeyini içerecektir. Bu aralık geniştir, ancak $X$ kullanılmadığında, güven aralığı $X$'e koşullu olmak yerine sadece $Y$'nin (ve $s_Y$) elde edilen değerlerine bağlı olan %95 olasılıkla elde edilecek $3,985 \leq Y \leq 4,787$ aralığı kadar geniş değildir.

Bunun yerine eğer $X$ puanı 10 olanlar için $Y$'nin ortalama değerini yordamak istersek, standart hata kestirimi aşağıdaki gibi bulunur:

$$s''_{Y \cdot X} = s_{Y \cdot X}\sqrt{\frac{1}{N} + \frac{(X_i - \overline{X})^2}{(N-1)s_X^2}}$$

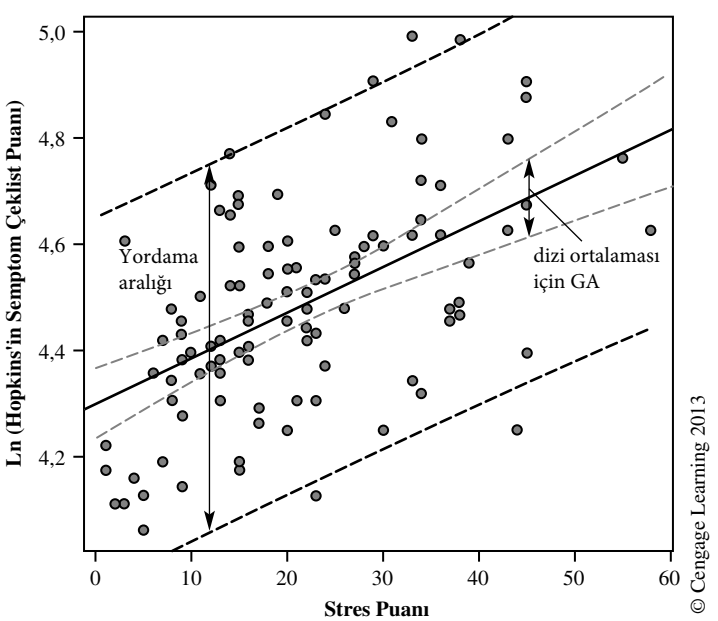
$$s''_{Y \cdot X} = 0,173\sqrt{\frac{1}{107} + \frac{(10 - 21,290)^2}{(106)156,05}}$$

$$= 0,173\sqrt{,017} = 0,022$$

ve güven aralığı da

$$GA(Y) = \hat{Y} \pm (t_{\alpha/2})(s'_{Y \cdot X})$$
$$= 4,386 \pm 1,984(0,022)$$
$$= 4,386 \pm 0,045$$
$$4,341 \leq Y \leq 4,431$$

Şekil 9.8'de dizi ortalamaları için güven aralığı olarak etiketlenmiş olan doğru etrafındaki güven sınırları gösterilmiştir ayrıca gelecek yordamalar için yordama aralığı belirtilmiştir. İkincinin çok geniş olduğuna dikkat ediniz. SPSS'te yordama veya güven aralığını




**Şekil 9.8** Yeni Stres değerleri için log(Semptomlar)ın yordanmasında güven sınırları

**Save/Prediction Intervals—Mean and Individual** düğmelerine tıklayarak oluşturabilirsiniz. Bu menü her bir durum ve aralık türü için üst ve alt sınırları ekleyecektir.

## 9.11 Test Alma Becerilerinin Rolünü Gösteren Bir Bilgisayar Örneği

Birçoğumuz bir materyali çalıştıktan sonra, bu materyale dayalı bir sınav aldığımızda makul ölçüde iyi yapabiliriz. Fakat materyale hiç bakmadan sınav aldığımızda nasıl yapabiliriz? (Bazılarınız bu deneyime sahip olabilir.) Katz, Lautenschlager, Blackburn ve Harris (1990) bu soruyu araştırmışlar ve bazı öğrencilere metni okumaları ve sonra çoktan seçmeli sorular serisini cevaplamalarını istemişlerdir. Ayrıca diğer öğrencilerden de metni görmeden soruları cevaplamalarını istemişlerdir. Burada ikinci gruba odaklanacağız. Test maddeleri Kuzey Amerikalı öğrencilerin üniversiteye girişte aldıkları SAT sınavlarında yer alan maddelere oldukça benzerdir. Bu durum araştırmacıların SAT'ta iyi yapan öğrencilerin bu başarıları üzerine şüphelenmelerine yol açtı çünkü her iki sınavda seçenekler arasından farklı olanı eleme gibi temel test alma becerisini içeriyordu.

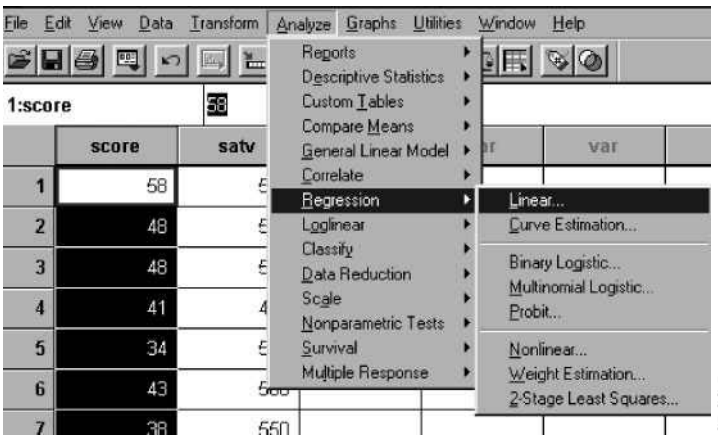
Katz ve arkadaşları tarafından elde edilen veri ile benzer özelliğe sahip veri Tablo 9.6'da verilmiştir. Puan değişkeni, öğrencinin metni görmeden yanıtladığı doğru cevap yüzdesini ve SATV değişkeni ise öğrencinin üniversite başvurusundaki sözel SAT puanını temsil etmektedir.

Gösterim 9.1 SPSS'de regresyon kullanılarak yapılan analizi göstermektedir. Burada birçok şey dikkat çekmektedir. İlk olarak, hangi değişkenin bağımlı, hangisinin bağımsız olduğuna karar verilmelidir. Eğer sadece değişkenler arasında korelasyon hesaplanacaksa, bu, sonucu etkilemeyecektir ancak regresyonda önemlidir. Bu örnekte akıllı tahmin yapan bireylerin aynı zamanda SAT'tan iyi puan alabileceklerini düşünerek nispeten keyfî bir karar verdim. Bu yüzden araştırma öncesinde elde edilmiş olsa bile SATV'yi bağımlı değişken olarak kullanıyorum. Gösterim 9.1'deki ilk 2 panel SPSS için gerekli menü seçimlerini gösterir. Ortalama ve standart sapma değerleri çıktının orta kısmında yer almakta ve buradan yüksek başarı puanlarına sahip (hemen hemen ortalama 600 puan ve standart sapma 60) gruplarla ilgilenildiğini görebilirsiniz. Bu, SAT için ortalamanın yaklaşık 100 puan üzerini ifade etmektedir. Bu öğrenciler ayrıca Katz'in testinde de maddelerin yaklaşık %50'sini doğru

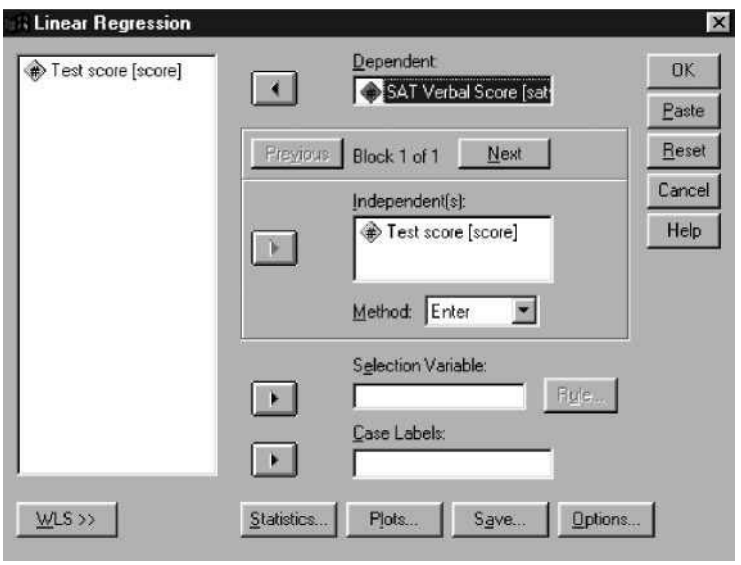
**Tablo 9.6** Katz ve arkadaşlarına (1990) göre metni okumayan grubun verisi

| Puan | SATV | Puan | SATV |
|---|---|---|---|
| 58 | 590 | 48 | 590 |
| 48 | 580 | 41 | 490 |
| 34 | 550 | 43 | 580 |
| 38 | 550 | 53 | 700 |
| 41 | 560 | 60 | 690 |
| 55 | 800 | 44 | 600 |
| 43 | 650 | 49 | 580 |
| 47 | 660 | 33 | 590 |
| 47 | 600 | 40 | 540 |
| 46 | 610 | 53 | 580 |
| 40 | 620 | 45 | 600 |
| 39 | 560 | 47 | 560 |
| 50 | 570 | 53 | 630 |
| 46 | 510 | 53 | 620 |

© Cengage Learning 2013




SPSS Inc.




SPSS Inc.

**Betimsel İstatistikler**

| | Ortalama | Standart Sapma | *N* |
|---|---|---|---|
| SAT Sözel Puanı | 598,57 | 61,57 | 28 |
| Test Puanı | 46,21 | 6,73 | 28 |

Adapted from output by SPSS, Inc.

**Gösterim 9.1** Katz ve arkadaşlarının (1990) test-alma davranışı çalışmasına ait SPSS çıktısı

**Korelasyonlar**

| | | SAT Sözül Puanı | Test Puanı |
|---|---|---|---|
| Pearson Korelasyon | SAT Sözel Puanı | 1,000 | ,532 |
| | Test Puanı | ,532 | 1,000 |
| Manidarlık (tek kuyruklu) | SAT Sözel Puanı | | ,002 |
| | Test Puanı | ,002 | |
| N | SAT Sözel Puanı | 28 | 28 |
| | Test Puanı | 28 | 28 |

**Model Özeti**

| Model | R | R Kare | Düzeltilmiş R Kare | Kestirmenin Std. Hatası |
|---|---|---|---|---|
| 1 | ,532[a] | ,283 | ,255 | 53,13 |

[a]Kestiriciler: (Sabit), Test puanı

**ANOVA[b]**

| Model | | Kareler Toplamı | sd | Kareler Ortalaması | F | Manidarlık |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 28940,123 | 1 | 28940,123 | 10,251 | ,004[b] |
| | Artık | 73402,734 | 26 | 2823,182 | | |
| | Toplam | 102342.9 | 27 | | | |

[a]Kestiriciler: (Sabit), T

[b]Bağımlı Değişken: SAT Sözel Puanı

**Korelasyonlar[a]**

| | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | | |
|---|---|---|---|---|---|---|
| Model | | B | Std. Hata | Beta | t | Manidarlık |
| 1 | (Sabit) | 373,736 | 70,938 | | 5,269 | ,000 |
| | Test puanı | 4,865 | 1,520 | ,532 | 3,202 | ,004 |

[a]Bağımlı Değişken: SAT Sözel Puanı

**Gösterim 9.1** Devam

cevaplayarak başarılı olmuşlardır. Bu istatistiklerin altında Puan ve SATV değişkenleri arasındaki korelasyon istatistiğinin ,532 olduğu görülmektedir. Bu değere manidarlık testi yapılacaktır. Bu korelasyon değerinin, SAT puanlarının eski test alma becerilerine çok bağımlı olabileceği ve öğrencinin kendisinden ne istenildiğinin farkında olmasını daha az yansıtabileceğinden dolayı çok yüksek olmasının istenemeyeceğini belirtebilirim.

Model Özeti bölümünde $R$ ve $R^2$ değerlerinin her ikisi de yer almaktadır. Burada "$R$" büyük harfle yazılmıştır, çünkü eğer çoklu yordayıcılar varsa çoklu korelasyon olacaktır ve bu sembol her zaman büyük harfle gösterilir. Burada $R$, $R^2$'nin kare kökü alınarak he-

saplanır ve ilişki negatif olsa bile $R^2$ her zaman pozitif olacaktır. Bu çoklu yordayıcılar için geçerli olan sürecin bir sonucudur.

ANOVA tablosu evrende korelasyonun ,00 olduğuna dair bir sıfır hipotezi testidir. Hipotez testlerini sonra tartışacağız, fakat burada önemli olan test istatistiği *F*'dir ve *F*'ye ait manidarlık düzeyi $p$ = ,004'tür. $p$ değerinin ,05'ten küçük olması nedeniyle sıfır hipotezini reddedeceğiz ve değişkenlerin doğrusal bağımsız olmadığı sonucunu çıkaracağız. Başka bir deyişle, öğrencilerin test alma becerisini yansıtan puanları ile SAT performansları arasında doğrusal ilişki vardır. Bu ilişkinin gerçek doğası çıktının en son kısmında yer almaktadır. Burada "katsayılar" başlıklı tabloda sabit değer ve eğim verilmektedir. Sabit değer burada "sabit" olarak etiketlenmiştir, çünkü her bir yordamaya sabit olarak eklenmektedir. Bu örnekte 373,736'dır. Teknik olarak eğer bir öğrenci Katz'in testinde 0 doğru cevap vermişse ortalama SAT puanının 370 olmasını bekleyeceğiz. Çünkü 0 puanı, öğrencilerin gerçekten aldığı puanlardan oldukça uzaktır (herhangi birinin kötü tahmin ile dahi 0 almasını zihinde canlandırmak oldukça zordur), bu değere çok fazla dikkat edilmeyecektir.

Bu tabloda eğim yordayıcı değişkenin ismi ile etiketlenmiştir. (Tüm yazılımlar böyle yapmaktadır, çünkü eğer çoklu yordayıcılar varsa, hangi yordayıcının hangi eğim değeri ile birlikte kullanılacağı bilinmelidir. Böyle yapmanın en kolay yolu değişkenin ismini etiket olarak kullanmaktır.) Bu örnekte eğim 4,865'tir, bunun anlamı Katz'in testinde 1 puanlık farklılığa sahip iki öğrencinin SAT'ta 4,865 puanlık bir farkla yordanmasıdır. Regresyon denklemimiz şimdi $\hat{Y}$ = 4,865 x Puan + 373,736 olarak yazılabilir.

Standart regresyon katsayısı ,532 olarak gösterilmiştir. Bu, test puanlarındaki 1 standart sapmalık değişimin, SAT puanlarında yaklaşık ½ standart sapmalık değişimle ilişkili olduğu anlamına gelir. Buradaki örnekte sadece bir yordayıcı olduğundan, standartlaştırılmış katsayının korelasyon katsayısına eşit olduğuna dikkat ediniz.

Standartlaştırılmış regresyon katsayısının sağında eğim ve sabit değerlerin manidarlık testinin *t* ve *p* değerleri görülmektedir. Eğimin manidarlık testini kısaca tartışacağız. Sabit değer testi ile çok nadir ilgilenilse de yorumlanması eğimin testi hakkında söylenenlere bir kanıt olabilecektir.

## 9.12 Hipotez Testi

İki değişken arasındaki ilişkinin yordayıcısı olarak *r*'nin ve *X*'in bir fonksiyonu olan *Y*'deki değişim oranının ölçümü olarak eğim (*b*) değerinin nasıl hesaplandığını gördük. *r* ve *b*'nin kestirimlerine ek olarak, evren parametre değerinin sıfıra eşit olduğuna dair sıfır hipotezinin manidarlık testinin de sıklıkla yapılması istenir. Örneklemden hesaplanan *r* ve *b* değerlerinin sıfır değerine eşit olmaması, evrene ait parametrelerde de sıfıra eşit olmayacağına kanıt değildir.

### *r*'nin Manidarlık Testi

Örneklem korelasyonu için test edilen en yaygın hipotez, evrende *X* ve *Y* arasındaki korelasyonun - $\rho$ (ro) ile gösterilmekte- sıfıra eşit olmasıdır. Bu mantıklı bir testtir çünkü test edilen sıfır hipotezi *X* ve *Y*'nin doğrusal bağımsız olduğuna ilişkin bir hipotezdir. Bu hipotezin reddedilmesi, değişkenlerin bağımsız olmadığı ve aralarında bir kısım doğrusal ilişki olduğu sonucuna götürür.

*N* için $\rho = 0$ olarak gösterildiği zaman, *r* yaklaşık olarak sıfırın etrafında normal dağılacaktır. ($\rho \neq 0$ olduğunda *r* ortalamanın etrafında normal dağılım gösteremeyecektir.)

*t* testi için test istatistiği aşağıdaki eşitlikle ifade edilir:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Bu oran *N*-2 sd'de *t* dağılımıdır.[14] Gösterim 9.1'de verilen örneğe geri dönüldüğünde, *r* = ,532 ve *N* = 28 dir. Böylece,

$$t = \frac{,532\sqrt{26}}{\sqrt{1 - ,532^2}} = \frac{,532\sqrt{26}}{\sqrt{,717}} = 3,202$$

elde edilir.

*t*'nin bu değeri $\alpha$ = ,05 (çift kuyruklu) düzeyinde manidardır ve buradan SAT puanları ve Katz testinden alınan puanlar arasında manidar bir ilişki olduğu sonucu çıkarılabilir. Diğer bir deyişle, mutlaka nedensel bir ilişki anlamına gelmemesine rağmen SAT puanlarındaki değişimin test puanlarındaki değişim ile ilişkili olduğu sonucunu çıkarabiliriz.

Bölüm 7'de kısaca *F* istatistiğinden söz ettik. Bölüm 11-16'da ayrıntılı olarak göreceğiz. Herhangi bir serbestlik derecesinde *F* istatistiğinin elde edilmesi için *t* istatistiğinin karesinin alınması gerektiğini bilmeniz gerekir. Birçok istatistik paketi hipotez testinde *t* yerine *F* istatistiğini kullanmaktadır. Bu örnekte sadece *F*'nin karekökünü alarak burada tartışılan *t* istatistiği elde edilebilir. (Gösterim 9.1'de *F* = 10,251 elde edilmiştir. Bu değerin karekökü 3,202'dir ve burada hesaplanan *t* değeri ile uyuşmaktadır).

İkinci örnek olarak, Tablo 9.2'deki stres ve psikolojik semptomlar verisine dönersek, *r* = ,506 ve *N* = 107 bulunur. Böylece

$$t = \frac{,529\sqrt{105}}{\sqrt{1 - ,529^2}} = \frac{,529\sqrt{105}}{\sqrt{,720}} = 6,39$$

elde edilir.

Burada yine $H_0$: $\rho = 0$ hipotezini reddedeceğiz. Stres ve semptomlar arasında manidar bir ilişki olduğu sonucunu çıkaracağız. Stresdeki değişimler rapor edilen psikolojik semptomlardaki değişimlerle ilişkilendirilir.

Korelasyon katsayısı için hipotez testini yapmamız her zaman mantıklı olduğu anlamına gelmez. İstatistiksel manidarlığın mantıklı olmadığı birçok durum söz konusudur. Akademik başarıyı yordama amacıyla bir ölçek geliştirdiğimde, korelasyon sadece ,19 ise hipotez testi bizim için çok kullanışlı olmayacaktır. Burada önemli olan *r* = ,19'un istatistiksel olarak ,00'dan çok farklı olup olmadığı değil, çok az değişimi açıkladığından kullanımının pek mümkün olmamasıdır. Ve test-tekrar test güvenirlik katsayısının istatistiksel olarak manidar olduğunu düşünenler, gerçekten ne yaptıkları hakkında fikir sahibi değildirler.

## *b*'nin Manidarlık Testi

Problem üzerinde bir an düşünürseniz, bu bölümde tartıştığımız tek-yordayıcılı durumda *b*'nin testinin *r*'nin testine denk olduğunun farkına varacaksınız. Eğer *X* ve *Y*'nin ilişkili olduğu bir gerçekse o halde *Y*'nin *X* ile değişebileceği de bir gerçektir ve eğim sıfır değildir. *b*'nin testi *r*'nin testi ile aynı sonucu doğurur ve *b*'nin testi için birlikte uygulanmalıdır. Bununla birlikte regresyon katsayısı çoklu regresyonda önemli bir rol oynadığından ve manidar korelasyon her bir yordayıcı değişken için manidar bir eğim değerini vermek zorunda olmadığından, testin tam formu burada verilecektir.

*b*'nin parametrik denkliğini (tüm evrenden *X* ve *Y* ölçümleri elde edilebilirse hesaplanan eğim) *b** olarak göstereceğiz.[15]

---

14 Bu, Bölüm 7'de gördüğümüz *Studen's t* ile aynıdır.

15 Pek çok ders kitabı b* yerine $\beta$ kullanır, fakat bu, standartlaştırılmış korelasyon katsayısı ile karışıklık yaratır.

*b'nin b** etrafında aşağıdaki standart hata ile normal dağıldığı gösterilebilir.[16]

$$s_b = \frac{s_{Y \cdot X}}{s_X\sqrt{N-1}}$$

Böylece, evrendeki regresyon doğrusunun gerçek eğiminin sıfır olduğu hipotezini test etmek istersek ($H_o$: $b^* = 0$), oranını sadeleştirerek

$$t = \frac{b - b^*}{s_b} = \frac{b}{\frac{s_{Y \cdot X}}{s_X\sqrt{N-1}}} = \frac{(b)(s_X)(\sqrt{N-1})}{s_{Y \cdot X}}$$

elde ederiz, bu *N*-2 *sd*'de *t*'nin dağılımıdır.

SAT performansı ve test alma becerisine ilişkin örnek verimizde, $b = 4{,}865$; $s_X = 6{,}73$ ve $s_{Y \cdot X} = 53{,}127$'dir.

Buradan,

$$t = \frac{(4{,}865)(6{,}73)(\sqrt{27})}{53{,}127} = 3{,}202$$

bulunur. Bu değer *r*'nin testinden elde edilen değerle aynıdır. Çünkü $t_{\text{elde edilen}} = 3{,}202$ ve $t_{,025}(26) = 2{,}056$ olduğu için, $H_0$'ı ret edeceğiz ve regresyon doğrusunun eğiminin de sıfırdan farklı olduğu sonucuna ulaşacağız. Başka bir ifade ile, yüksek seviyede test alma becerileri, yüksek yordanan SAT puanları ile ilişkilidir.

*b*'nin örneklem dağılımı hakkında bilinenlerden, *b**'nin güven sınırlarını belirlemek mümkündür.

$$GA(b^*) = b \pm (t_{\alpha/2})\left[\frac{(s_{Y \cdot X})}{s_X\sqrt{N-1}}\right]$$

Burada $t_{\alpha/2}$ , *t*'nin *N*-2 *sd*'de çift kuyruklu kritik değeridir.

Buradaki veri için, ilişkili istatistik Gösterim 9.1'den elde edilebilir. %95 güven sınırları aşağıdaki gibidir:

$$GA(b^*) = 4{,}865 \pm 2{,}056\left[\frac{53{,}127}{6{,}73\sqrt{27}}\right]$$
$$= 4{,}865 \pm 3{,}123 = 1{,}742 \leq b^* \leq 7{,}988$$

%95 olasılıkla bu sınırlar gerçek *b** değerlerini sınırlayacaktır. Güven sınırlarının sıfır değerini içermediğine dikkat ediniz. $H_0$: $b^* = 0$ hipotezinin ret edildiği, *t* testinin sonucu ile aynıdır.

## İki Bağımsız *b* Değeri Arasındaki Farkın Test Edilmesi

Bu test tek eğim testinden daha az yaygın olmasına rağmen neden kullanıldığı sorusu anlamlıdır. Bir bireyin tükettiği sigara miktarı ve yaşam beklentisi arasındaki ilişki üzerine iki veri kümesi olduğunu varsayınız. Bir küme veri kadınlar, diğeri erkeler için toplanmış olsun. Tek geniş bir veri kümesinden çok iki ayrı veri kümesi söz konudur, çünkü sonuç-

16 b'nin standart hatası için en iyi yaklaşım konusunda anlaşmazlık vardır. Paydası farklı çeşitlerde $s_X\sqrt{N}, s_X\sqrt{N-1}, s_X\sqrt{N-2}$ olarak verilir.

ların kadın ve erkeklerin ortalama ömürleri arasındaki normal farklılıktan etkilenmesini istemiyoruz. Aşağıdaki verinin elde edildiğini varsayınız:

| | Erkek | Kadın |
|---|---|---|
| $b$ | −0,40 | −0,20 |
| $s_{Y \cdot X}$ | 2,10 | 2,30 |
| $s_X^2$ | 2,50 | 2,80 |
| $N$ | 101 | 101 |

Tablodaki veriden gördüğümüze göre, erkeklere ait regresyon doğrusu kadınlara ait olana göre daha diktir. Eğer bu farklılık manidar ise, her bir sigara tüketim miktarının artışı için, erkeklerin ortalama ömürlerinin kadınlara göre daha fazla azalmakta olduğu anlamı çıkarılabilir. Eğer bu doğru olsaydı önemli bir bulgu olurdu ve bu yüzden $b_1$ ve $b_2$ arasındaki farkın testi ile ilgilenilirdik.

İki bağımsız regresyon katsayısı arasındaki fark için $t$ testi, iki bağımsız ortalama arasındaki farkın testinin benzeridir. Eğer $H_0$ doğru ise ( $H_0 : b^*_1 = b^*_2$), $b_1$ - $b_2$'nin örneklem dağılımı sıfır ortalama ve aşağıdaki standart hata ile normaldir.

$$s_{b_1 - b_2} = \sqrt{s_{b_1}^2 + s_{b_2}^2}$$

Bu oran

$$t = \frac{b_1 - b_2}{\sqrt{s_{b_1}^2 + s_{b_2}^2}}$$

$t$ dağılımı verir ve serbestlik derecesi $N_1 + N_2$ - 4'tür. $b$'nin standart hatasının kestiriminin

$$s_b = \frac{s_{Y \cdot X}}{s_X \sqrt{N - 1}}$$

olduğunu biliyoruz, böylece örneklemlerin hata varyansları $s^2_{Y \cdot X_1}$ ve $s^2_{Y \cdot X_2}$ olmak üzere,

$$s_{b_1 - b_2} = \sqrt{\frac{s^2_{Y \cdot X_1}}{s^2_{X_1}(N_1 - 1)} + \frac{s^2_{Y \cdot X_2}}{s^2_{X_2}(N_2 - 1)}}$$

ifadesi yazılabilir. Ortalamalar ile ilgili durumda olduğu gibi, eğer hata varyanslarının homojen olduğunu varsayarsak, varyansları serbestlik dereceleri ile ağırlıklandırarak toplamak suretiyle,

$$s^2_{Y \cdot X} = \frac{(N_1 - 2)s^2_{Y \cdot X_1} + (N_2 - 2)s^2_{Y \cdot X_2}}{N_1 + N_2 - 4}$$

ifadesini yazabiliriz. Örneğimiz için bu değer aşağıdaki gibidir:

$$s^2_{Y \cdot X} = \frac{99(2,10^2) + 99(2,30^2)}{101 + 101 - 4} = 4,85$$

Bu değeri birleştirilmiş varyansı denklemin içine yerleştirerek

$$s_{b_1 - b_2} = \sqrt{\frac{s^2_{Y \cdot X_1}}{s^2_{X_1}(N_1 - 1)} + \frac{s^2_{Y \cdot X_2}}{s^2_{X_2}(N_2 - 1)}}$$

$$= \sqrt{\frac{4,85}{(2,5)(100)} + \frac{4,85}{(2,8)(100)}} = 0,192$$

elde edilir. Verilen $s_{b_1 - b_2}$ değerini $t$'ye göre düzenleyerek;

$$t = \frac{b_1 - b_2}{s_{b_1 - b_2}} = \frac{(-0,40) - (-0,20)}{0,192} = -1,04$$

buluruz. 198 serbestlik derecesi için $t_{0,025}(198) = 1,97$ olduğundan $H_0$ reddedilmez ve böylece kadın ve erkeklerde aynı oranda sigara tüketiminin bir fonksiyonu olarak ortalama ömrün azaldığından şüphe duyacak bir sebep olmadığı sonucuna ulaşırız.

Şu unutulmamalıdır ki $H_0 : b^* = 0$ hipotezi $H_0 : \rho = 0$'a denk olmasına rağmen, $H_0 : b_1^* - b_2^* = 0$'ın $H_0 : \rho_1 - \rho_2$'ye denk olmasını gerektirmez. Eğer bir süreliğine düşünürseniz, iki dağılım diyagramının benzer regresyon doğrusuna $(b_1^* - b_2^*)$ sahip olduğunu görebilirsiniz, fakat doğrunun etrafındaki dağılımın derecesi farklıdır (bu yüzden $\rho_1 \neq \rho_2$'dir). Tersi de ayrıca düşünülebilir - iki farklı regresyon doğrusu, ait oldukları veri kümeleriyle eşit uyum gösterebilir.

## İki Bağımsız *r* Değeri Arasındaki Farkın Test Edilmesi

İki bağımsız $r$ arasındaki farkın test edilmesinde küçük bir zorluk belirir. Eğer $\rho \neq 0$ ise, $r$'nin örneklem dağılımı yaklaşık olarak normal olmayacaktır ($\rho \Rightarrow \pm 1,00$'e göre oldukça daha fazla çarpık olacaktır) ve standart hata kolay kestirilemeyecektir. Benzer durum $r_1 - r_2$ için de geçerlidir. Bu, açıkça bir problem oluşturacaktır, çünkü zihinde canlandırılabileceği gibi, eğer farklılığa ilişkin $t$ testi yapılacaksa korelasyonlar arasındaki farkın standart hatasının bilinmesi zorunluluğu vardır. Bu problemin çözümü R.A. Fisher tarafından bulunmuştur.

Fisher (1921), $r$,

$$r' = (0,5)\log_e \left| \frac{1 + r}{1 - r} \right|$$

eşitliği yardımıyla $r'$ye dönüştürülürse $r'$, $\rho'$ ($\rho$'nun dönüştürülmüş değeri) civarında yaklaşık olarak

$$s_{r'} = \frac{1}{\sqrt{N - 3}}$$

standart hata ile normal dağılıma sahip olacağını göstermiştir. (Fisher bu istatistiği "$z$" ile göstermiştir, fakat standart normal sapma ile karıştırılmasını engellemek için "$r'$" kullanılmaktadır.) Standart hata bilindiğinden, $\rho_1 - \rho_2 = 0$ olan sıfır hipotezi şimdi her bir "$r$", "$r'$" değerine dönüştürülerek aşağıdaki test istatistiği edilebilir:

$$z = \frac{r_1' - r_2'}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

Buradaki test istatistiğinin $t$'den çok z olduğuna dikkat ediniz, çünkü buradaki standart hata örneklemden ($N$'den farklı) hesaplanan istatistiğe bağlı değildir ve bu yüzden bir parametredir.

Ek $r'$, $r$'nin farklı değerleri için $r'$ değerlerini tablolaştırır, böylece $r'$ için yukarıdaki eşitliği çözme ihtiyacı ortadan kalkar.

Meijs, Chillessen, Scholte, Segers ve Spijkermann (2010) akademik başarı, sosyal zekâ ve ergenlerdeki popülerlik arasındaki ilişki çalışmasını yapmışlardır. Popülerlik, daha da ileri gidilerek algılanan popülerliğe bölünmüştür (AP), bu sosyal baskınlık ve prestiji yansıtır ve bir öğrencinin kaç kez "en popüler" ve kaç kez "en az popüler" olarak aday gösterilmesi arasındaki fark ile ölçülür. Sosyometrik popülerlik (SP) bireyin beğenilme

**Tablo 9.7** Cinsiyete göre temel değişkenlerin korelasyonları

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Akademik Başarı | | –,05 | –,05 | ,08 |
| Sosyal Zekâ | ,10 | | ,31* | ,19* |
| Algılanan Pop. | –,01 | ,20* | | ,57* |
| Sosyometrik Pop. | ,04 | ,13 | ,33* | |

(* $p < ,05$) Köşegenin altındaki korelasyonlar erkekler, üstündekiler kızlar içindir.
$N_{\text{erkekler}} = 225$; $N_{\text{kızlar}} = 287$.

© Cengage Learning 2013

derecesi ile ilgilidir, "en çok sevilen" ve "en az sevilen" için aday gösterilme farklılığı olarak hesaplanmıştır (popülerlik ve sevilmenin aynı olduğu düşünülebilir ancak değildir). Veri cinsiyete göre ayrılmıştır, çünkü erkekler ve kızlar arasında bu değişkenin farklılık gösterebileceği düşünülmüştür. Tablo 9.7'de, korelasyonlar erkekler ve kızlar için ayrı ayrı verilmiştir.

Hem kızlar hem de erkekler için akademik başarı ve diğer değişkenler arasında açıkça bir ilişkinin olmadığı ilginç bir bulgudur. Korelasyon önemsenmeyecek kadar düşüktür. Erkekler ve kızlarda Sosyal Zekâ ve Algılanan Popülerlik arasındaki korelasyonu karşılaştırmayı istediğinizi varsayınız. Kızlar için korelasyon ,31 ve erkekler için ,20'dir. Buradan,

| | Erkekler | Kızlar |
|---|---|---|
| $r$ | ,20 | ,31 |
| $r'$ | ,203 | ,320 |
| $N$ | 225 | 287 |

$$z = \frac{,203 - ,320}{\sqrt{\frac{1}{225-3} + \frac{1}{287-3}}} = \frac{-,117}{\sqrt{,008}} = \frac{-,117}{,089} = -1,32$$

bulunur. $z_{\text{elde edilen}} = -1,32$ değeri $z_{,025} = -1,96$'dan büyük olduğundan $H_0$ reddedilemez ve ile $\alpha = ,05$ manidarlık düzeyinde çift kuyruklu test için Sosyal Zekâ ve Algılanan Popülerite arasındaki korelasyonun kızlar ve erkekler için benzer olduğu sonucu çıkarılır. Buna rağmen Algılanan ve Sosyometrik Popülerlik arasındaki korelasyon kızlar için manidar olarak daha yüksektir.

Genellikle örneklem çok büyük olmadıkça, herhangi bir karşılaştırmada iki bağımsız $r$ değeri arasında manidar bir fark bulmanın zor olduğunun altını çizmeliyim. Şüphesiz manidar olarak farklı iki korelasyon bulunabilir, teorik ve pratikte ilişkiyi test etmede araştırmacı kendisini kısıtlarsa istatistiksel olarak manidar bir farklılık elde etmek genellikle zordur.

## $\rho$'nun Belirli Bir Değere Eşit Olduğu Hipotezinin Test Edilmesi

$r'$ kavramını tartıştığımıza göre, şimdi $\rho$'nun sıfır dışında herhangi bir değere eşit olduğu sıfır hipotezini de test etmek durumundayız. Bu hipotezi kullanacağınız birçok durumu düşünemeyebilirsiniz, ben de. Fakat bunu kullanabilme yeteneği daha kullanışlı bir yöntem olarak $\rho$'da güven sınırlarının belirlenmesine imkân sağlar.

Yukarıda gördüğümüz gibi, $\rho$'nun herhangi bir değeri için, $r'$'nün örneklem dağılımı $\rho'$ ($\rho$'nun dönüştürülmüş değeri) etrafında $\frac{1}{\sqrt{N-3}}$ standart hata ile normal dağılmaktadır. Buradan

$$z = \frac{r' - \rho'}{\sqrt{\frac{1}{N-3}}}$$

Buradaki $z$ standart normal sapmadır. Eğer $\rho$ = ,50 olan bir evrenden çekilen $r$ = ,30 olan bir örneklem için sıfır hipotezi test edilmek isteniyorsa

$$r = ,30 \qquad r' = ,310$$
$$\rho = ,50 \qquad \rho' = ,549$$
$$N = 103 \qquad s_{r' = 1/\sqrt{N-3}} = 0,10$$
$$z = \frac{,310 - ,549}{0,10} = -0,239/0,10 = -2,39$$

elde edilir.

$z_{\text{elde edlen}}$ = -2,39 değeri $z_{,025}$= –1,96'dan daha uç değer olduğundan, $\alpha$ = 0,5'te (çift kuyruklu) $H_0$ reddedilir ve örneklemimiz $\rho$ = ,50 olan bir evrenden gelmemektedir sonucunu çıkarırız.

## $\rho'$nun Güven Sınırları

Önceki tartışmadan yola çıkılarak, eşitlik $z$ yerine $\rho$ için çözülerek $\rho$ için güven sınırlarını belirleyebiliriz. Bunu yapmak için önce $\rho'$ için güven sınırları belirlenir, sonra $\rho'$ tekrar $\rho$'ya dönüştürülür.

$$z = \frac{r' - \rho'}{\sqrt{\frac{1}{N-3}}}$$

böylece

$$(\pm z) * \sqrt{\frac{1}{N-3}} = r' - \rho'$$

ve buradan

$$GA(\rho') = r' \pm z_{\alpha/2}\sqrt{\frac{1}{N-3}}$$

bulunur. Stres örneği için $r$ = ,529 ($r'$ = ,590) ve $N$ = 107, böylece %95 güven sınırları

$$GA(\rho') = ,590 \pm 1,96\sqrt{\frac{1}{104}}$$
$$= ,590 \pm 1,96(0,098) = ,590 \pm 0,192$$
$$= ,398 \le \rho' \le ,782$$

elde edilir. $\rho'$ tekrar $\rho'$ya dönüştürüldüğünde ve yuvarlandığında

$$,380 \le \rho \le ,654$$

bulunur. Böylece güven sınırlar $\rho$ = ,380 ve $\rho$ = ,654'tür. Elde edilen sınırlar ,95 olasılıkla $\rho$'nun gerçek değerini kapsayacaktır. Sınırlar arasında $\rho$ = 0 olmadığına dikkat ediniz, bu $H_0$: $\rho$ = 0 eşzamanlı testi önerebilir, bu bilgi ile ilgilenilmelidir. Ayrıca güven sınırlarının $r$ etrafında asimetrik olarak dağıldığına dikkat ediniz çünkü $r$'nin örneklem dağılımı çarpıktır.

## Manidarlık Testine Karşı Güven Sınırları

En azından davranış bilimlerinde, birçok kitap, kurs ve basılmış araştırmada manidarlık testine odaklanılmış olup güven sınırlarına daha az önem verilmektedir. Bazı durumlarda bu uygundur, fakat diğer durumlarda okuyucuyu sınırlar.

Bu bölümde stres ve psikolojik semptom örneğine sıklıkla değindik. Bu konuyu araştıran ilk birkaç kişi için, bu iki değişken arasında manidar ilişki olup olmadığı gerçekten önemli bir problemdi. Ancak şimdi buna inanan herkes için ilişkinin ne kadar büyük olduğu problemi ortaya çıkmıştır. Ve bu problem için, uygun cevap iki değişken arasındaki korelasyon ,529'dur, %95 olasılıklı güven aralığı $,380 \leq \rho \leq ,654$'tür. (Karşılaştırılmalı ifade $r = ,529 \pm ,15$ hata payı (yaklaşık) biçiminde olabilirdi.)[17]

## Bağımsız Olmayan İki *r* Arasındaki Farklılığın Testi

Bazen birbirinden bağımsız olmayan iki korelasyon arasındaki farkın test edilmek istendiği durumlarla karşılaşırız. (Muhtemelen her yıl birkaç defa bu soruyu sorarım.) Bir durum, iki korelasyon ortak bir değişkeni paylaştığında ortaya çıkar. Böyle bir örneği aşağıda göreceğiz. Diğer bir durum ise, Zaman 1'de iki değişken arasındaki ilişkiye bakıldığında ve sonraki bir noktada tekrarlandığında (Zaman 2) ve zamana bağlı olarak korelasyonda manidar bir değişkenlik olup olmadığı sorulduğunda karşımıza çıkar. Bu durumla ilgili bilgi vermeyeceğim, ancak bu duruma özgü iyi bir tartışmayı core.ecu.edu/psyc/wuenschk/StatHelp/ZPF.doc adresinde ve Raghunathan, Rosenthal, ve Rubin (1996) tarafından yapılan çalışmada bulabilirsiniz.

Korelasyonların ortak bir değişkeni paylaşmasına örnek olarak Reilly, Drudge, Rosen, Loew ve Fischer (1985) birinci sınıf öğrencilerine, önce iki zekâ testini (WISC-R McCarthy), bundan 2 yıl sonra aynı öğrencilere (Geniş Kapsamlı Başarı Testi) Wide Range Achievement (WRAT) uygulamışlar ve aşağıda verilen korelasyon katsayılarını elde etmişlerdir.

| | WRAT | WISC-R | McCarthy |
|---|---|---|---|
| WRAT | 1,00 | ,80 | ,72 |
| WISC-R | | 1,00 | ,89 |
| McCarthy | | | 1,00 |

Burada WISC-R ve McCarthy'nin yüksek korelasyona sahip olduğuna dikkat ediniz fakat WISC-R, WRAT (okuma) ile McCarthy'den daha yüksek korelasyon göstermiştir. Burada ilgilenilen soru, WISC-R ve WRAT korelasyonu (,80) ve McCarthy ve WRAT korelasyonu (,72) arasındaki farklılığın manidar olup olmadığıdır, bu sorunun cevabı bağımsız olmayan korelasyon testini gerektirir, çünkü her ikisinde WRAT ortaktır ve aynı örnekleme dayanmaktadır.

Eğer bağımsız olmayan iki korelasyonumuz varsa -buradakiler değil, çünkü test puanları aynı 26 öğrenciye dayanmaktadır- bağımsızlığın olmadığı dikkate alınmalıdır. Özellikle, ilişkili iki testin derecesini temsil eden bir terimi eklemeliyiz. Hotelling (1931) geleneksel bir çözüm önerdi, fakat daha iyisi Williams (1959) tarafından düzenlendi ve Steiger (1980) tarafından doğrulandı. Bu son test aşağıdaki eşitlikle ifade edilmiştir:

$$t = (r_{12} - r_{13}) \sqrt{\frac{(N-1)(1+r_{23})}{2\left(\frac{N-1}{N-3}\right)|R| + \frac{(r_{12}+r_{13})^2}{4}(1-r_{23})^3}}$$

Burada

[17] Yukarıda görülebileceği gibi, limitler tam olarak *r* etrafında simetrik olmadığı için yaklaşık kelimesi kullanılmıştır.

$$|R| = (1 - r_{12}^2 - r_{13}^2 - r_{23}^2) + (2r_{12}r_{13}r_{23})$$

Bu oran *N*-3 *sd*'de *t* dağılımı gösterir. Bu eşitlikte, $r_{12}$ ve $r_{13}$ farklılığı test edilen korelasyon katsayılarını ve $r_{23}$ bu iki yordayıcı arasındaki korelasyonu temsil eder. |R|, elemanları korelasyonlar 3x3 biçiminde korelasyonlar arası matrisin determinantıdır, fakat siz determinant hakkında bir şey bilmeden hesaplayabilirsiniz.

Buradaki örnekte;

$r_{12}$ = WISC-R ve WRAT arasındaki korelasyon = ,80
$r_{13}$ = McCarthy ve WRAT arasındaki korelasyon = ,72
$r_{23}$ = WISC-R ve McCarthy arasındaki korelasyon = ,89

Buradan,

$$|R| = (1 - ,80^2 - ,72^2 - ,89^2) + (2)(,80)(,72)(,89) = ,075$$

$$t = (,80 - ,72)\sqrt{\frac{(25)(1 + ,89)}{2\left(\frac{25}{23}\right)(,75) + \frac{(,80 + ,72)^2}{4}(1 - ,89)^3}}$$

$$= 1,36$$

elde edilir. Bulunan $t_{\text{elde edilen}}$ = 1,36 değeri 23 *sd*'de manidar değildir. Buna rağmen, bu sonuç testlerin üçüncü sınıf öğrencilerinin performanslarını WRAT testinin okuma ölçeğinde eşit etkililikte yordadığı iddiasının bir kanıtı değildir, çünkü sıfır hipotezini kanıtlayamacağınız için bu iddia ile tutarlıdır ve destekler.

## 9.13 Son Bir Örnek

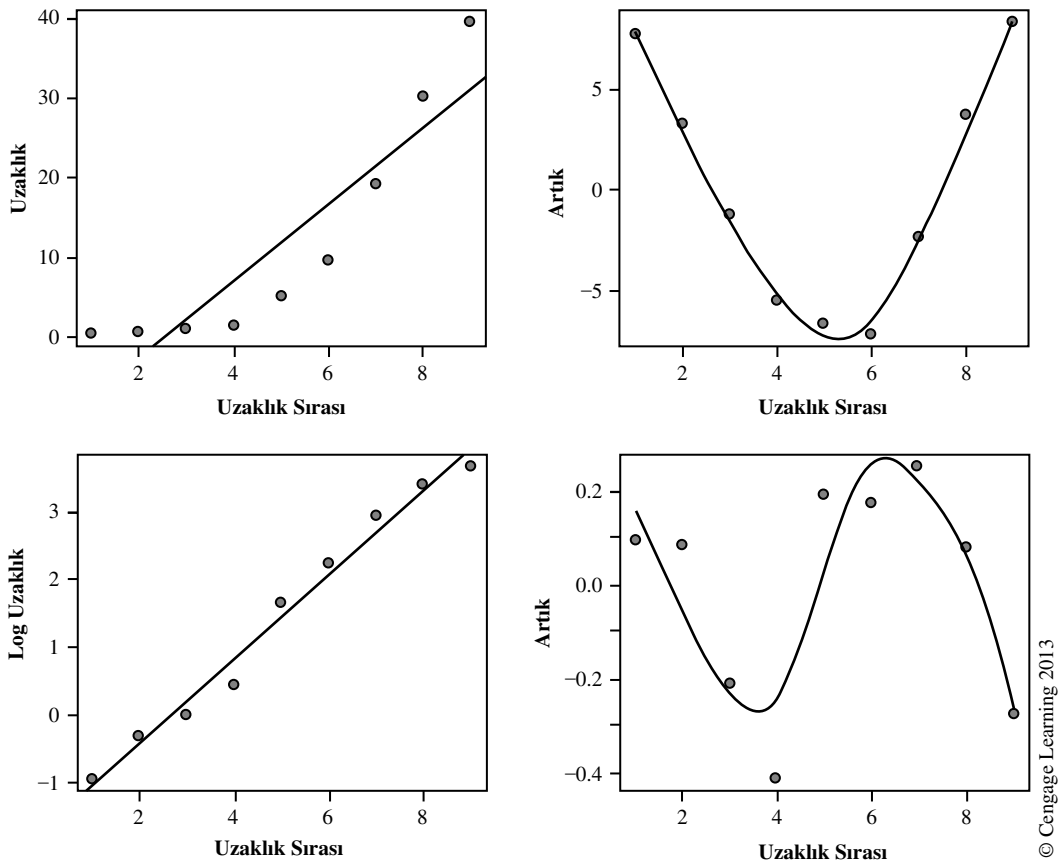
Korelasyon ve regresyon hakkında birçok önemli noktayı gösterecek son bir örnek vermek istiyorum. Bu, psikologlar hakkında olduğu kadar, fizikçi ve astronomlara da aittir, Todman ve Dugard'dan (2007) alınmış anlamlı bir örnek olup çok önemli bir noktaya değinir. Bin yıldan uzun zamandır Güneş ile Güneş sistemimizde bulunan gezegenler arasındaki uzaklığın düzenli bir örüntü izlediği bilinmektedir. Aşağıdaki tabloda uzaklıklar gösterilmiştir. Her ne kadar kısa bir zaman önce Plüton'un gezegen olmadığına karar verilse de tabloda yer almaktadır. (Daha sonra göreceğimiz gibi diğer gezegenlerin örüntüsüne düzenli şekilde uyduğu gerçeğine rağmen, onun daha düşük bir seviye olan "cüce gezegen" durumuna düşürülmesinin adil olmadığı izlenimi uyandırmaktadır.)

Eğer bu uzaklıklar kendi orijinal birimlerinde grafiğe dönüştürülürse, doğrusal olmayan düzenli bir grafik elde edilir. Grafik Şekil 9.9a'da gösterilmiştir. Aradaki ilişki açıkça doğrusal olmamasına rağmen, doğrusal regresyon doğrusu grafik üzerine eklenmiştir. Şekil 9.9b'de dizilimin fonksiyonu olarak çizilen önceki regresyonun artıkları, eklenen yiv şeklinde bir eğri ile görülebilir. Artıklar düzenli bir örüntüyü izlediğinden kesinlikle bir şeyler olduğunu gösterir. Bu örüntü verinin, uzaklığın logaritmik dönüşümü ile daha iyi

**Tablo 9.8** Astronomik birimde güneşe olan uzaklıklar

| Gezegen | Merkür | Venüs | Dünya | Mars | Jüpiter | Satürn | Uranüs | Neptün | Plüton |
|---|---|---|---|---|---|---|---|---|---|
| Sıra | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Uzaklık | 0,39 | 0,72 | 1 | 1,52 | 5,20 | 9,54 | 19,18 | 30,06 | 39,44 |

© Cengage Learning 2013




**Şekil 9.9** Gezegenlerin Güneş'ten uzaklıkları ile ilgili grafikler

uyum göstereceği izlenimi uyandırmaktadır. Şekil 9,9'un sol alt köşesinde, uzaklık dizilimine karşı çizilen uzaklığın logaritması görülmektedir, buradaki değişken seçimine dikkat ediniz. Noktaların regresyon doğrusuna ne kadar yakın olduğundan da görülebileceği gibi ilişki neredeyse doğrusaldır. Fakat sağ altta dizilim mesafesine karşı çizilen regresyon artıklarının da doğruladığı gibi, burada görülen örüntüden ilişkinin logaritmik olduğunu ifade etmek mümkündür. Artıklarda halen düzgün bir örüntü olduğuna dikkat edilmelidir. Bu, veriyi uydurma konusunda her ne kadar başarılı olunsa da, artıklarda halen sistematik bir çeşitlilik olduğunu gösterir. İkinci grup artıklar için astronomların halen bir açıklaması olmadığı bilinmektedir, fakat bir açıklama ihtiyacı olduğu kesindir.

Bu örneği birçok nedenle seçtim. İlk olarak, psikoloji ve fizik arasındaki farklılıkları göstermektedir. Psikologların, fizik bilimindeki değişkenlerin duyarlığına sahip bir değişkenle çalıştığını düşünemem. Birincisi psikolojide verinin bu kadar iyi uyum sağladığı asla görülmez. İkinci olarak, bu örnek artıklara bakmanın önemini gösterir - temel olarak modelin nerede yanlış olduğunu gösterirler. Sol üstteki ilk grafik, çok sistematik ve doğrusal olmayan bir şeylerin olduğunun kanıtı olsa da, devamında, dizilim uzaklığına karşı grafikleştirilen log uzaklıkta da bu söz konusu olmaya devam etmiştir. Burada artıklar, daha çok şeyin açıklanması gerektiğini ortaya koymaktadır. Son olarak, bu örnek regresyon analizi ile teori arasındaki etkileşimi çok iyi göstermektedir. Hiç kimse her gezegenin Güneş'e

olan uzaklığını yordamada regresyon kullanmayı düşünmez. Bu uzaklıklar zaten bilinmektedir. Önemli olan bu ilişkinin ne olduğu tanımlanarak teori üretilebilir ya da doğrulanabilir olmasıdır. Büyük ihtimalle bir fizikçi için ilişkinin logaritmik olduğunu söylemenin ne anlama geldiği açıktır. (Bu yerçekimi kuvvetinin, uzaklığın karesiyle ters orantılı olarak değişmesiyle ilişkilidir.) Fakat logaritmik ilişki açıklandıktan sonra bile, açıklanması gereken çok şey olduğu görülmektedir. Psikologlar, değişkenlerin kesin yargılara varmayı zorlaştıran yeterince tesadüfî hata içermesine rağmen, regresyonu aynı amaçla kullanırlar. Bölüm 14'te çoklu regresyona gelindiğinde regresyon analizinin amacının teori üretme olduğunu tekrar göreceksiniz.

## 9.14 Korelasyon ve Regresyonda Varsayımlarının Rolü

Korelasyon ve regresyon tekniklerinin kullanılması konusunda, altında yatan varsayımlarla ilgili, literatürde çok sayıda karışıklık söz konusudur. Karışıklığın büyük bir kısmı, korelasyon ve regresyon modellerinin birçok benzer sonuç üretmesine rağmen, farklı varsayımlara dayandığı gerçeğinden kaynaklanmaktadır. Karışıklık, istatistikçilerin tüm varsayımları başlangıçta yapma ve bu varsayımların bazılarının belirli amaçlar için gerekli olmadığını belirleyememelerinden dolayı da ortaya çıkmaktadır.

**regresyonun doğrusallığı**

**eğrisel (doğrusal olmayan)**

Doğrusal regresyon ve iki değişkenli normal model ile ilgili en önemli varsayım ve tüm yorumlar **regresyonun doğrusallığıdır**. $X$ ve $Y$ arasındaki ilişki ne olursa olsun, doğrusaldır - bunun anlamı, verinin doğruya iyi uyduğudur. Burada sadece **eğrisel (doğrusal olmayan)** ilişki örneği görüldü ancak aksi belirtilmedikçe korelasyon ve regresyonda doğrusallık varsayılır. (Yordamanın hatasının iyi kestirilememesine rağmen ve nadiren, eğrisel veriye doğru uyumu sağlatılır; bu, doğrunun yeterince uyum sağlaması varsayımı için böyle yapılır. Doğrusal regresyondan başka regresyon türleri de vardır, ancak burada tartışılmayacaktır.)

Önceden belirttiğimiz gibi, ne yapmak istediğimize göre olsun ya da olmasın çeşitli varsayımlar yaparız. Eğer amacımız veriyi basitçe betimlemek ise varsayım zorunlu değildir. Verinin çekildiği evren hakkında herhangi bir varsayım zorunluluğu olmadan, veriyi en iyi şekilde regresyon doğrusu ve $r$ tanımlar.

Eğer amacımız $X$'teki varyansla $X$ ile doğrusal ilişkili olan $Y$'deki varyansın derecesini saptamak ise, yine varsayıma gerek yoktur. $s^2_Y$ ve $s^2_{Y \cdot X}$ altta yatan herhangi bir varsayımdan bağımsız olarak ait oldukları parametrelerin yansız kestiricileri olduğundan bu doğrudur ve cebirsel olarak

$$\frac{KT_Y - KT_{\text{artık}}}{KT_Y}$$

$r^2$'ye eşdeğerdir.

Eğer $b$ veya $Y$'de güven sınırlarını belirlemek veya $b^*$ için hipotez testi yapmak istersek, $Y$'nin seri değerlerinde varyansların homojenliği ve normallik varsayımlarına ihtiyaç duyarız. Varyansın homojenliği varsayımı, her bir seri için varyansın $s^2_{Y \cdot X}$ ile temsil edildiği durumlarda ve normallik varsayımı da standart normal dağılım kullanıldığında zorunludur.

Eğer $\rho = 0$ hipotezini test etmek için $r$ kullanılmak veya $\rho$'da güven sınırları tayin etmek istersek ($X$, $Y$) çiftinin iki değişkenli normal dağılımdan tesadüfi olarak çekildiğini varsaymak zorundayız, ancak birçok çalışmada $r$'nin manidarlığı önemli değildir ve $r$'nin güven sınırlarını genellikle belirlenmediğini unutmamalıyız.

# 9.15 Korelasyonu Etkileyen Faktörler

Korelasyon katsayısı örneklem özelliklerinden önemli derecede etkilenebilir. Bu özellikler *X* veya *Y*'de ranj daralması (veya varyans) ve heterojen alt örneklemlerin kullanılmasıdır.

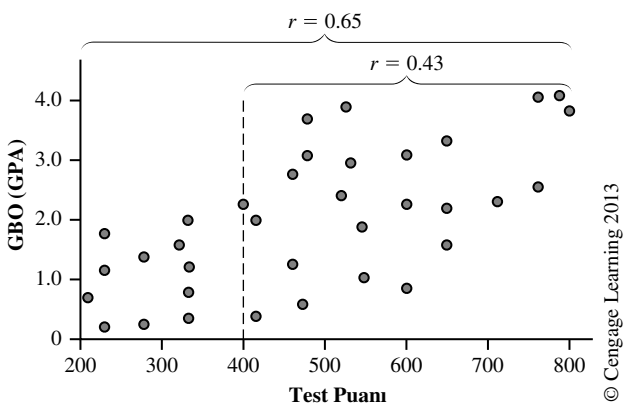
## Ranj Daralmasının Etkisi

**ranj daralmaları**

*X* ve *Y* değişkenlerinin ranjları farklı olduğunda ranj daralması yaygın bir problemdir. Böyle **ranj daralmaları**nın etkisi, ranj daralması olmadığındaki duruma göre *X* ve *Y* arasındaki korelasyonun alacağı değeri değiştirmesidir. Ranjın daralmasında, *r*'nin genel olarak azalmasına rağmen verinin doğasına bağlı olarak korelasyon artabilir veya azalabilir.

Alışılmadık bir durum hariç, eğer sınırlama ile bazı eğrisel ilişkiler yok edilebiliyorsa *X*'in ranjının daralması *r*'yi artırır. Örneğin, okuma yeteneği yaş ile ilişkilendiriliyorsa (burada yaş aralığı 0-70'tir) veri kesinlikle eğrisel olacaktır (4 yaşa kadar yatay, 17 yaşa kadar artan ve sonra düzelen) ve doğrusal ilişkiyi ölçen korelasyon bağıl olarak düşük olacaktır. Eğer yaş aralığı 5'ten 17'ye kadar sınırlanırsa, *Y*'nin, *X*'in doğrusal fonksiyonu olarak dağılmayan değerleri dahil edilmediğinden korelasyon yüksek olabilir.

*X* veya *Y*'nin ranjının daralmasının daha olağan etkisi korelasyonu düşürmesidir. Bu problem özellikle test geliştirmede geçerlidir, çünkü burada ölçüt değişken (*Y*) sadece *X*'in yüksek değerleri için elde edilebilir. Şekil 9.10'da verilen hipotetik veriye dikkat ediniz. Bu şekil öğrencilerin hipotetik örneklemi için üniversite Genel Başarı Ortalamaları (GBO) ile bazı standart testlerden elde edilen puanların (SAT gibi) arasındaki ilişkiyi temsil etmektedir. Test geliştiricilerin ideal dünyasında, sınavı alan tüm bireylerin GBO'ları yeterli bulunarak üniversiteye girmişlerdir ve başarı test puanları ile GBO'ları arasındaki korelasyon hesaplanmıştır. Şekil 9.10'da görülebileceği gibi, bu korelasyon yüksek olacaktır. Buna rağmen gerçek dünyada herkes üniversiteye kabul edilmemektedir. Üniversiteler, başarı test puanlarına, lise performansına veya başka sınıflandırmalara dayanarak sadece daha yetenekli öğrencileri almaktadır. Bu GBO puanlarının daha çok standartlaştırılmış testlerden yüksek puan alan öğrenciler için var olduğu anlamındadır. Bunun *X*'in sadece 400'den büyük değerleri için *X* ve *Y* arasındaki ilişkinin değerlendirilmesini sağlayan bir etki olduğunu varsayınız. Şekil 9.10'da yer alan veri için korelasyon bağıl olarak düşük olacaktır; bu, testin iyi geliştirilmiş olmamasından değil, ranj daralmasından kaynaklanır.




**Şekil 9.10** Daraltılmış aralığın (ranjın) etkisini gösteren hipotetik veri

Bir başka ifade ile Şekil 9.10'da yer alan noktaların tüm örneklemi kullanıldığında korelasyon ,65'tir. Buna rağmen örneklem en az 400 test puanı alan bireyler için daraltıldığında korelasyon ,43'e düşecektir. ($X < 400$ için tüm veriyi kapatırsanız bunu görebilmeniz kolaydır.) Daraltılmış örnekleme dayalı korelasyon çalışmalarında ranj daralmasının etkisini dikkate almalıyız. Katsayı, çalışılan problem için uygun olmayabilir. Aslında burada yapılan iş standartlaştırılmış bir testin bireyin üniversite için uygunluğunu ne kadar iyi yordayabileceğini sormaktır, fakat bu sorunun cevabı sadece üniversiteye kabul edilen bireyler için elde edilmiştir.

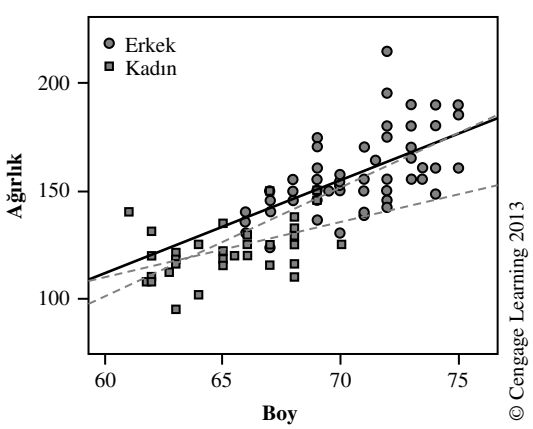
Dunning ve Friedmann (2008) buna benzer bir örnek kullanarak, ranj daralmasının, *r*'nin üzerinde ciddi etkileri var iken, regresyon doğrusunu etkilemeden bırakabileceğine değinmişlerdir. (Bunu Şekil 9.10'da gösterilen tüm veri ve sınırlandırılmış veriye regresyon doğrusunu uydurarak gösterebilirsiniz.) Buna rağmen etki, toplanmayan verinin, benzer yolla toplanılan veri ile ilişkili olduğu varsayımına dayanır.

## Heterojen Alt Örneklemlerin Etkisi

**heterojen alt örneklemler**

Korelasyon analizinin sonuçlarının değerlendirilmesinde diğer bir önemli husus **heterojen alt örneklemler**dir. Bu, erkek ve kadınlarda boy ve kilo arasındaki ilişkiyi içeren basit bir örnekle açıklanabilir. Bu değişkenlerin psikoloji ile ilgisi az gibi görünebilir ancak her iki değişken bireyin kendilerini nasıl gördüğünün değerlendirilmesinde önemli bir rol oynamaktadır. Veri kümelerinin grafikleri Minitab kullanılarak Şekil 9.11'de gösterilmiştir. Bu veri kümesi Minitab kılavuzundaki örneklerden alınmıştır (Ryan ve diğerleri, 1985). Veri kümesi, 92 üniversite öğrencisinden alınan boy, kilo, cinsiyet ve diğer değişkenlerle ilişkilidir. (Verinin, bireyin kendini anlatmaya dayalı yöntemle elde edilmiş olduğunu ve sistematik ifade yanlılığı içerebileceğini unutmayınız.)

Eğer erkek ve kadınların her ikisi için veri kümeleri birleştirilirse, ilişki çarpıcı bir şekilde iyi görünüyor, korelasyon ,782'dir. Eğer inceleme her iki cins için ayrı ayrı yapılırsa, korelasyon erkekler için ,60, kadınlar için ,49'a düşecektir. (Erkekler ve kadınlar için veri kümelerinin grafikleri farklı sembollerle gösterilmiştir; kadınlara ait veri kümesi şeklin genellikle sol aşağı tarafındadır. Regresyon denklemi erkekler için $\hat{Y}_{\text{erkek}} = 4{,}36 \text{ x boy}_{\text{erkek}} - 149{,}93$ ve kadınlar için $\hat{Y}_{\text{kadın}} = 2{,}58 \text{ x boy}_{\text{kadın}} - 44{,}86$ dır.)




**Şekil 9.11** Erkekler ve kadınlar için birleştirilmiş boy ve kilo arasındaki ilişki (kesik çizgi = kadın, düz çizgi = erkek, noktalı çizgi = birleştirilmiş)

Önemli bir nokta, veri kümeleri cinsiyete göre birleştirildiğinde yüksek korelasyon elde edilmesi sadece boy ve kilo arasındaki ilişkiden kaynaklanmayıp büyük ölçüde erkeklerin ortalama olarak kadınlara göre daha uzun ve ağır olduğu gerçeğinden de kaynaklanmaktadır. Aslında bir kâğıda yapılabilecek karalamalar, her bir cinsiyetin ağırlığı ile boyu arasındaki negatif ilişkinin, cinsiyetler birleştirildiğinde pozitif olabileceğini gösterir. Burada belirtilmek istenen nokta, araştırmacıların farklı kaynaklardan elde edilen veri kümelerini birleştirirken dikkatli olması gerektiğidir. İki değişken arasındaki ilişki başka bir üçüncü değişken tarafından gizlenmiş veya artırılmış olabilir. Böyle bir bulgu tek başına önemlidir.

Heterojen alt örneklemlerle ilgili benzer bir etki veren ikinci bir örnek, erkek ve kadınlarda kalp damar hastalığı ve kolesterol düzeyi arasındaki ilişkidir. Eğer cinsiyetleri birleştirirseniz ilişki etkileyici değildir. Fakat veriyi erkek ve kadınlara göre ayrı ayrı incelerseniz, kolesterol düzeyi için artışın kalp damar hastalığında artış eğilimi olduğunu görürsünüz. Bu ilişki birleştirilmiş veride bulanıklaşmaktadır, çünkü erkekler kolesterol düzeyinden bağımsız olarak kadınlara kıyasla üst seviyede kalp damar hastalıklarına sahiptir.

## 9.16 Pearson'ın *r*'si İçin Güç Hesaplaması

Televizyondaki şiddet ile agresif davranışlar arasındaki ilişkinin gösterilmek istendiği bir problemi dikkate alınız. Bu çalışmanın desenlenmesiyle ilgili olarak araştırmacının tüm gerçek durumları oluşturmaya çalıştığını ve iki değişken arasında korelasyon elde etmek için bir yol bulduğunu varsayınız. Araştırmacı evrende korelasyon katsayısının ($\rho$) yaklaşık ,30 olduğunu düşünmektedir. (Bu korelasyon düşük gibi görülebilir ancak, agresif davranışların içerdiği tüm değişkenler dikkate alındığında etkileyicidir. Bu değer Huesmann, Moise-Titus, Podolski ve Eron (2003) tarafından yapılan çalışmadan elde edilen değer ile uyumludur, ancak Block ve Crain (2007) tarafından yapılan çalışmada, bu ilişkinin büyüklüğü tartışılmıştır.) Araştırmacımız böyle bir korelasyon elde etmek için araştırmayı yapmak istemektedir, ancak başlamadan önce çalışmasının gücü hakkında bilgi sahibi olmak da istemektedir. Bu durumda güç hesaplaması yapmak kolaydır.

Tekrar hatırlayacağınız gibi, eğer güç hesaplanacaksa önce etki büyüklüğü (*d*) tanımlanır. Daha sonra örneklem büyüklüğü açıklanır ve $\delta$ hesaplanır. Son olarak $\delta$ değeri, Ek'te yer alan Güç tablosundan araştırmanın güç değerini hesaplamak için kullanılır. Güç hesaplamasına

$$d = \rho_1 - \rho_0 = \rho_1 - 0 = \rho_1$$

tanımlanarak başlanır. Burada $\rho_1$, $H_1$ tarafından tanımlanan evrendeki korelasyondur, bu örnekte ,30'dur. Ardından,

$$\delta = d\sqrt{N-1} = \rho_1\sqrt{N-1}$$

tanımlanır.

50 büyüklüğündeki bir örneklem için,

$$\delta = ,30\sqrt{50-1} = 2,1$$

Güç tablosundan $\delta$ = 2,1 ve $\alpha$ = ,05 (iki kuyruklu) için güç = ,56 bulunur.

,56'lık bir güç katsayısı araştırmacıyı memnun etmez ve araştırmacı gücü artırmanın bir yolunu arar. Gücün ,80 olmasını ister. Güç tablosundan bunun için $\delta$ = 2,8'e ihtiyaç olduğunu görür. Bu yüzden,

$$\delta = \rho_1\sqrt{N-1}$$
$$2,8 = ,30\sqrt{N-1}$$

Her iki tarafın karesi alınarak,

$$2,8^2 =, 30^2(N-1)$$
$$\left(\frac{2,8}{,30}\right)^2 + 1 = N = 88$$

bulunur.

Böylece .80 güç elde etmek için, araştırmacının yaklaşık 90 katılımcıdan veri toplaması gerekecektir. (Televizyondaki şiddetin etkisini ele alan birçok çalışma, bundan daha fazla katılımcıya dayanır.) R'da yazılmış, güç için korelasyon katsayısını hesaplayan kısa bir programa kitabın internet sitesinde CorrelationPower.R adıyla ulaşılabilir.

## Ek Örnekler

Korelasyon ve regresyonla ilgili birkaç ek ve faydalı materyal çektim. Bunlar http://www.uvm.edu/~dhowell/methods8/Supplements/CorrReg.html. adresindedir.

Bunlar, şu ana kadar gördüklerimizden daha karmaşık, her bir veri kümesi için birkaç farklı istatistiksel süreçle ilgilidirler. Bu örneklere bakarak korelasyonun pratikte kullanılışına dair iyi bir fikir edinebilirsiniz ve tanımadığınız materyalleri geçebilirsiniz. Sadece basit korelasyonla ilgili materyalle ilgileniniz, burada şimdiye kadar görmediğimiz konularla ilgili ileri düzeyde süreçlerle ilgili olan materyaller bulunabilir.

## Anahtar Terimler

İlişkiler (Giriş)
Faklılıklar (Giriş)
Korelasyon (Giriş)
Regresyon (Giriş)
Tesadüfî değişken (Giriş)
Sabit değişken (Giriş)
Doğrusal regresyon modeli (Giriş)
İki değişkenli normal model (Giriş)
Yordama (Giriş)
Grafik (9.1)
Diyagram (9.1)
Yordayıcı ((9.1)
Ölçüt (9.1)
Regresyon doğrusu (9.1)
Doğrusal ilişki (9.2)
Eğrisel ilişki (9.2)
Korelasyon (*r*) (9.1)
Kovaryans $kov_{XY}$ veya $s_{XY}$ (9.3)
Evrendeki korelasyon katsayısı $\rho$ (rho) (9.4)
Düzeltilmiş korelasyon katsayısı ($r_{düzel.}$) (9.4)
Eğim (9.5 )
Sabit (9.5)
Yordamanın hatası (9.5)
Artık (9.5)
Normal denklemler (9.5)
Standartlaştırılmış regresyon katsayısı $\beta$ (beta) (9.5)
Grafik düzgünleştiricileri (9.7)
Spliness (9.7)
Loess (9.7)
Kareler toplamı (*SSy*) (9.8)
Yordamanın standart hatası (9.8)
Artık varyans (9.8)
Hata varyansı (9.8)
Koşullu dağılım (9.8)
Hatada oransal azalma (9.8)
Yordamada oransal gelişim (9.8)
Seri/Dizi (9.9)
Serilerdeki varyansın homojenliği (9.9)
Serilerde normallik (9.9)
Koşullu seri (9.9)
Koşullu dağılım (9.9)
Marjinal dağılım (9.9)
Yordama aralığı (9.10)
Regresyonun doğrusallığı (9.13)
Eğrisel (9.13)
Ranj daralması (9.15)
Heterojen alt örneklemler (9.15)

# Alıştırmalar

9.1 Sahra Çölü'nün güney kısmında annelerin yarısı en az bir çocuğunu, bir yaşını doldurmadan kaybetmektedir. Aşağıda, bölgedeki 36 ülkeye ait, ülke, bebek ölümleri, kişi başı gelir (Amerikan doları), 20 yaş altı annelerde doğum oranı, 40 yaş üstü annelerde doğum oranı, 2 yıldan kısa aralıkla gerçekleşen doğum oranı, evli kadınlarda doğum kontrolü kullanma oranları ve aile planlaması ile tanışmamış kadınların oranına ait veri verilmiştir. (http://www.guttmacher.org/pubs/ib_2-02.html)

| Ülke | Bebek Ölümü | Gelir | % Anne < 20 | % Anne > 40 | <2 yıl aralıklı | Doğum kontrolü kullanma | Aile planlaması ihtiyacı |
|---|---|---|---|---|---|---|---|
| Benin Rep | 104 | 933 | 16 | 5 | 17 | 3 | 26 |
| Burkina Faso | 109 | 965 | 17 | 5 | 17 | 5 | 26 |
| Cameroon | 80 | 1,573 | 21 | 4 | 25 | 7 | 20 |
| Central African Rep | 102 | 1,166 | 22 | 5 | 26 | 3 | 16 |
| Chad Rep | 110 | 850 | 21 | 3 | 24 | 1 | missing |
| Côte d'Ivoire | 91 | 1,654 | 21 | 6 | 16 | 4 | 28 |
| Eritrea | 76 | 880 | 15 | 7 | 26 | 4 | 28 |
| Ethiopia | 113 | 628 | 14 | 6 | 20 | 6 | 23 |
| Gabon | 61 | 6,024 | 22 | 4 | 22 | 12 | 28 |
| Ghana | 61 | 1,881 | 15 | 5 | 13 | 13 | 23 |
| Guinea | 107 | 1,934 | 22 | 5 | 17 | 4 | 24 |
| Kenya | 71 | 1,022 | 18 | 3 | 23 | 32 | 24 |
| Madagascar | 99 | 799 | 21 | 5 | 31 | 10 | 26 |
| Malawi | 113 | 586 | 21 | 6 | 17 | 26 | 30 |
| Mali | 134 | 753 | 21 | 4 | 26 | 5 | 26 |
| Mozambique | 147 | 861 | 24 | 6 | 19 | 5 | 7 |
| Namibia | 62 | 5,468 | 15 | 7 | 22 | 26 | 22 |
| Niger | 136 | 753 | 23 | 5 | 25 | 5 | 17 |
| Nigeria | 71 | 853 | 17 | 5 | 27 | 9 | 18 |
| Rwanda | 90 | 885 | 9 | 7 | 21 | 13 | 36 |
| Senegal | 69 | 1,419 | 14 | 7 | 18 | 8 | 35 |
| Tanzania | 108 | 501 | 19 | 5 | 17 | 17 | 22 |
| Togo | 80 | 1,410 | 13 | 6 | 14 | 7 | 32 |
| Uganda | 86 | 650 | 23 | 4 | 28 | 8 | 35 |
| Zambia | 108 | 756 | 30 | 4 | 19 | 14 | 27 |
| Zimbabwe | 60 | 2,876 | 32 | 4 | 12 | 50 | 13 |

*Kaynak*: Guttmacher Institute, Community health centers and family planning, Issues in Brief, New York: Guttmacher, 2002, http://www.guttmacher.org/pubs/ib_2-02.html, accessed May 16, 2011.

a. Bebek ölümü ve gelir dağılımı grafiğini çiziniz.

b. Veriye en iyi uyum sağlayan doğruyu göz ile (yaklaşık olarak) çiziniz.

c. Gelirdeki iki aykırılığın nasıl bir etkiye sahip olduğunu düşünüyorsunuz?

9.2 Alıştırma 9.1.'deki tüm sayısal değişkenler için korelasyonu SPSS kullanarak hesaplayınız.

9.3 Ek *t*'deki tabloyu kullanarak, Alıştırma 9.2'de gösterilen ilişkinin manidar olması için ne büyüklükte bir korelasyona ihtiyaç vardır?

9.4 Alıştırma 9.2'de bebek ölümleri için en güçlü yordayıcılar nelerdir?

9.5 Bebek ölümleri ile ilgili veriden ne sonuç çıkarabiliriz?

9.6 Alıştırma 9.1'de 40 yaş üzeri kadınların oranı manidar görünmemektedir, ancak başka toplumlarda bu bir risk faktörüdür. Sizce neden böyledir?

9.7 Bebek ölümlerinde iki yordayıcı manidar görünmektedir. Eğer aynı zamanda her ikisi de yordayıcı olarak kullanılabilseydi nasıl bir sonuç elde edilirdi?

9.8 Daha önceki örneklerden, düşük gelirin bebek ölümlerine sebep olduğu sonucu çıkarılabilir mi?

9.9 Bebek ölümleri toplum için çok ciddi bir problemdir. Neden psikologlar bu problemle diğer mesleklerdeki insanlardan daha fazla ilgilenmelidir?

9.10 Down sendromu psikologların çalıştığı başka bir problemdir. İleri yaşlarda anne olanlarda Down sendromlu çocuk sahibi olma ihtimallerinin yüksek olduğu ileri sürülmüştür. Yaş ve hastalığın görülme sıklığı arasındaki ilişkiyi gösteren aşağıdaki verinin grafiğini çiziniz. Veri Geyer'dan (1991) alınmıştır.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Yaş | 17,5 | 18,5 | 19,5 | 20,5 | 21,5 | 22,5 | 23,5 | 24,5 | 25,5 |
| Doğumlar | 13,555 | 13,675 | 18,752 | 22,005 | 23,796 | 24,667 | 24,807 | 23,986 | 22,860 |
| Downs | 16 | 15 | 16 | 22 | 16 | 12 | 17 | 22 | 15 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Yaş | 26,5 | 27,5 | 28,5 | 29,5 | 30,5 | 31,5 | 32,5 | 33,5 | 34,5 |
| Doğumlar | 21,450 | 19,202 | 17,450 | 15,685 | 13,954 | 11,987 | 10,983 | 9,825 | 8,483 |
| Downs | 15 | 27 | 14 | 9 | 12 | 12 | 18 | 13 | 11 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Yaş | 35,5 | 36,5 | 37,5 | 38,5 | 39,5 | 40,5 | 41,5 | 42,5 | 43,5 |
| Doğumlar | 7,448 | 6,628 | 5,780 | 4,834 | 3,961 | 2,952 | 2,276 | 1,589 | 1,018 |
| Downs | 23 | 13 | 17 | 15 | 30 | 31 | 33 | 20 | 16 |

| | | | |
|---|---|---|---|
| Yaş | 44,5 | 45,5 | 46,5 |
| Doğumlar | 596 | 327 | 249 |
| Downs | 22 | 11 | 7 |

© Cengage Learning 2013

Yaşın bir fonksiyonu olarak Down sendromunun gerçekleşme yüzdesinin dağılım grafiğini çiziniz.

9.11 Alıştırma 9.10'daki veride Pearson korelasyonunu hesaplamak neden uygun değildir?

9.12 Alıştırma 9.11'deki problemin üstesinden gelmenin bir yolu, Down sendromu sıklığını sıralı veriye çevirmektir. Sıralanmış sıklığı kullanarak grafiği tekrar çiziniz ve korelasyonu hesaplayınız. Bu, bir sonraki bölümde göreceğimiz Spearman korelasyonudur.

9.13 Önemli bir gelişimsel sorun, düşük kilo ile doğan bebeklerde beyin kanaması şiddeti ve aynı çocukların 5 yaşındaki kavrama eksikliği arasındaki ilişki ile ilgilidir. 25 çocuk kullanmanın planlandığı ve ,20 korelasyonun beklendiği varsayılırsa, bu çalışmanın gücünün büyüklüğü nedir?

9.14 Alıştırma 9.13'te ,80'lik bir güç elde etmek için kaç çocuk kullanılması gerekir?

9.15 Tablo 9.2'deki bilgiyi ve hesaplanan katsayıyı kullanarak, 8 stres puanı için log (semptomlar) puanını yordayınız.

9.16 Tablo 9.3'teki veri için ortalama stres puanı 21,467 idi. Bu stres puanına sahip bir bireye ait log(semptomlar) için yordamanız ne olur? $\bar{Y}$ ile nasıl karşılaştırılabilir?

9.17 Yeni durumlar için psikolojik semptomların yordanmasında $\hat{Y}$'nun %95 olasılıklı güven aralığını yazınız. Şekil 9.2'de yer alan güven sınırlarını kullanabilirsiniz.

9.18 1886'da İngiliz bilim adamı Sir Francis Galton, vasatlık boyunca regresyondan bahseder, bugün buna iyimser bir yaklaşımla ortalama boyunca regresyon denir. Temel prensip, herhangi bir dizinin sonlarındaki bireylerin (boy, IQ ya da müzik yeteneği) çocuklarının

ortalama değere kendilerinden daha yakın olma eğilimleri vardır. Galton'ın fikrini açıklamak için standartlaştırılmış veri ile regresyon katsayısı (eğimi) olarak *r* kavramını kullanınız.

9.19 Bu kitabın Web sitesinde bulunan Galton.dat isimli dosya, "ortalama regresyon" başlığı altında Galton'ın ebeveynler ve çocuklarının boyları hakkındaki gerçek veri kümesini içerir. Bu veride Galton erkeklerin boyları ile aynı ortalamayı verebilmek için annelerin ve kızların boylarını 1,08 ile çarpmıştır ve "ortalama ebeveyn" boyunu üretebilmek için anne ve babanın boylarının ortalamasını almıştır. Veri Stigler'dan (1999) alınmıştır.

a. Çocukların boylarının ebeveynlerin boylarına karşı regresyonunu elde ediniz.

b. Ebeveynlerin boylarına bağlı olarak çocukların yordanan boylarını hesaplayınız.

c. Veri dosyası, 1 en düşük olmak üzere 1 ile 4 arasında Çeyrek (Quartile) adı verilen bir değişken içermektedir. Her bir çeyreğe bağlı olarak çocuk ortalamasını vermek için SPSS (Analyze/Compare Means/One-way ANOVA) kullanınız. (Çocuğu bağımlı değişken, çeyreği bağımsız değişken olarak alınız.) Aynısını ebeveyn ortalaması için de yapınız.

d. En yüksek çeyrekteki ebeveynlerin çocuklarını, ebeveynlerinden daha düşük ortalamaya sahip yapınız ve tam tersini en düşük çeyrekteki ebeveynlerin çocukları için yapınız.

e. Ebeveyn çeyrek ortalamalarının *X* ekseninde, çocuk quartile ortalamalarının *Y* ekseninde olduğu bir dağılım grafiği çiziniz. Ayrıca aynı ortalama boydaki çocuklara sahip ebeveynleri temsil eden 45 derecelik bir doğru çiziniz.

9.20 Okulun öğrenci aldığı bölgenin eğitime yaptığı harcama miktarı ile öğrencilerin SAT gibi standart testlerdeki performansı arasındaki ilişkiyi göstermek istiyorsunuz. Sadece gerçek korelasyonun en az ,40 olduğu bir korelasyon elde etmek istiyorsunuz. Eğer 30 okul bölgeniz var ise, manidar örnek bir korelasyon bulma şansınız nedir?

9.21 Alıştırma 9.20'de ,80 güç elde etmek için kaç bölgeye ihtiyacınız vardır?

9.22 Guber (1999) alıştırma 9.20 ve 9.21'deki temel soruya işaret eden veriyi bir araya getirmiştir. Bütün 50 eyalet için eğitim harcamaları, SAT performansı, SAT sınavına giren öğrencilerin yaşa göre oranı ve diğer değişkenleri de içeren okul performansı ile ilişkili birçok değişkenin verisini elde etmiştir. Bu veri kümesine ileride daha derinlemesine bakılacaktır, ancak aşağıdaki tablo Guber'ın verisi için SPSS bilgisayar çıktısını içermektedir.

| Model | R | R kare | Düzeltilmiş R Kare | Tahminin Std. Hatası |
|---|---|---|---|---|
| 1 | ,453[a] | ,205 | ,188 | 65,49 |

[a]Kestiriciler: (Sabit), her bir öğrenci için harcama

[b]Bağımlı Değişken: Birleştirilmiş SAT Ortalaması 1994-1995

**ANOVA[b]**

| Model | | Kareler Toplamı | sd | Kareler Ortalaması | F | Manidarlık |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 50920,767 | 1 | 50920,767 | 11,872 | ,001[a] |
| | Artık | 197303,0 | 46 | 4289,197 | | |
| | Toplam | 248223,8 | 47 | | | |

[a]Kestiriciler: (Sabit), her bir öğrenci için harcama

[b]Bağımlı Değişken: Birleştirilmiş SAT Ortalaması 1994-1995

Adapted from output by SPSS, Inc.

**Katsayılar[b]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | | |
|---|---|---|---|---|---|---|
| | | B | Std. Hata | Beta | t | Manidarlık |
| 1 | (Constant) | 1112,769 | 42,341 | | 26,281 | ,000 |
| | Test puanı | –23.918 | 6,942 | –,453 | –3,446 | ,001 |

[a]Bağımlı Değişken: Birleştirilmiş SAT Ortalaması 1994-1995

Adapted from output by SPSS, Inc.

Bu veri beklediğimiz örüntüyü ortaya çıkarmaz. Bu veri neyi gösterir? (Bölüm 15'te, diğer değişkenler kontrol edilebildiğinde, beklenen örüntünün elde edilebileceğini göstereceğiz. Test yapımcısı kuruluşların, çok fazla şaşırtıcı değişken olduğundan, test puanlarının böyle amaçlarla kullanılmasını istemediklerini söylemek zorundayım.)

9.23 Bu bölümde ve Alıştırma 7.13 ve 7.29'da kullanılan Katz, Lautenschlager, Blackburn ve Harris (1990) tarafından yapılan çalışmada, SAT'ta yer alan okuduğunu anlama sorularını ilk önce pasajı okumadan cevaplayan öğrencilerin, şans başarısından daha iyi performans gösterdikleri görülmüştür. Bu SAT'ın kullanışlı bir test olmadığı anlamına gelmez. Katz ve arkadaşları kendi katılımcılarının, gerçek SAT sözel puanları ve 100 maddelik testteki performansları arasındaki korelasyonu hesaplamışlardır. Bu pasajı okuyan katılımcılar için korelasyon ,68 ($N = 118$), pasajı okumadan cevaplayanlar için korelasyon ,53'tür ($N = 28$).

a. Bu korelasyonlar birbirinden manidar olarak farklı mıdır?

b. Bu veri kümesi ne sonuç çıkarırsınız?

9.24 Katz ve arkadaşları, deneylerini, SAT Sözel puanları ilk çalışmalarındaki bireylerden daha çok grup içi varyans gösteren bireyleri kullanarak tekrarlamışlardır. Bu durumda pasajı okuyan grup için korelasyon ,88 ($N = 52$) iken, okumayan grup için ise ,72'dir ($N = 74$). Bu korelasyon katsayıları birbirinden manidar olarak farklı mıdır?

9.25 Alıştırma 9.23 ve 9.24'te verilen korelasyonlar arasındaki farklılıktan ne gibi sonuçlar çıkarabilirsiniz?

9.26 "Sigara tüketimine karşı ortalama ömür" örneğinden yola çıkarak kendi örneğinizi oluşturunuz ve $r^2$ ile açıklanabilen değişkenlik arasındaki ilişkiyi gösteriniz.

9.27 Moore ve McCabe (1989) alkol ve tütün tüketimine ilişkin önemli bir istatistiksel kavramı gösteren bazı ilginç veri kümeleri elde etmişlerdir. Veriyi İngiliz İstihdam Departmanı'nın Aile Harcamaları Araştırmalarından toplamışlardır. Bağımlı değişkenler Büyük Britanya'nın 11 bölgesinden alkol ve tütün tüketimi için ortalama haftalık hane harcamalarıdır.

| Bölge | Alkol | Tütün |
|---|---|---|
| North | 6.47 | 4.03 |
| Yorkshire | 6.13 | 3.76 |
| Northeast | 6.19 | 3.77 |
| East Midlands | 4.89 | 3.34 |
| West Midlands | 5.63 | 3.47 |
| East Anglia | 4.52 | 2.92 |
| Southeast | 5.89 | 3.20 |
| Southwest | 4.79 | 2.71 |
| Wales | 5.27 | 3.53 |
| Scotland | 6.08 | 4.51 |
| Northern Ireland | 4.02 | 4.56 |

© Cengage Learning 2013

a. Bu iki değişken arasındaki ilişki nedir?

b. İrlandalıların çok içtiğine dair popüler ön yargı vardır. Bu veri, bu inancı desteklemekte midir?

c. Kuzey İrlanda'nın eklenmesinin bu sonuçlara nasıl bir etkisi vardır? (Bir dağılım grafiği çözümde yardımcı olabilir.)

9.28 Mireault ve Bond (1992) tarafından kullanılan http://www.uvm.edu/~dhowell/methods8/DataFiles/DataSets.html'de yer alan Mireault.dat dosyasındaki veride, üniversitedeki öğrencilerin performansları (GPA ile değerlendirilen) ile öğrencilerin psikolojik semptomları (GSIT ile değerlendirilen) arasında bir ilişki var mıdır?

9.29 Alıştırma 9.28'de sözü edilen veri kullanılarak

a. Kısaltılmış Semptom Envanteri alt ölçeğinden elde edilen tüm korelasyonları hesaplayınız. (İpucu: Hemen hemen tüm istatistik programları basit bir şekilde bu korelasyonları hesaplayabilir. Her birini ayrı ayrı hesaplamaya gerek yoktur.)

b. (a)'ya verilen cevap ayrı ölçekler arasındaki ilişkiler hakkında bize ne söyler?

9.30 Regresyon kullanımı için gerekli olan varsayımlardan birisi de serideki varyansın homojenliği varsayımıdır. Bu varsayımın ihlâli için verinin incelenmesinin bir yolu $Y$'nin yordanan değeri ve ilgili artıkların $(Y - \hat{Y})$ hesaplanmasıdır. Eğer artıkların yordanan değerlere karşı grafiğini çizerseniz, daha fazla veya az noktalar kümesi görebilirsiniz. Dikey dağılım sağdan sola hareket ederken ne düzenli olarak artmalı/azalmalı ne de belirgin bir örüntüye sahip olmalıdır. Bu kitabın internet sitesinde yer alan Cancer.Dat dosyasındaki veri için dağılım grafiğini çiziniz. Birçok bilgisayar programı bu grafiği çizdirmenize olanak sağlar. Sağlamazsa, ilk olarak regresyon eşitliğini belirleyerek daha sonra bilgisayar komutu ile bu eşitliği doldurarak, uygun değişkenleri basitçe üretebilirsiniz (örneğin;"set Pred = 0.256*GSIT+4.65," and "set Resid = TotBPT – Pred").

9.31 Aşağıdaki veri kümesi erkek üniversite öğrencileri için daha önce belirtilen gerçek boy ve ağırlıkları temsil etmektedir.

| Boy | Kilo | Boy | Kilo |
|---|---|---|---|
| 70 | 150 | 73 | 170 |
| 67 | 140 | 74 | 180 |
| 72 | 180 | 66 | 135 |
| 75 | 190 | 71 | 170 |
| 68 | 145 | 70 | 157 |
| 69 | 150 | 70 | 130 |
| 71.5 | 164 | 75 | 185 |
| 71 | 140 | 74 | 190 |
| 72 | 142 | 71 | 155 |
| 69 | 136 | 69 | 170 |
| 67 | 123 | 70 | 155 |
| 68 | 155 | 72 | 215 |
| 66 | 140 | 67 | 150 |
| 72 | 145 | 69 | 145 |
| 73.5 | 160 | 73 | 155 |
| 73 | 190 | 73 | 155 |
| 69 | 155 | 71 | 150 |
| 73 | 165 | 68 | 155 |
| 72 | 150 | 69.5 | 150 |

(Devam edecek)

© Cengage Learning 2013

| | | | |
|---|---|---|---|
| 74 | 190 | 73 | 180 |
| 72 | 195 | 75 | 160 |
| 71 | 138 | 66 | 135 |
| 74 | 160 | 69 | 160 |
| 72 | 155 | 66 | 130 |
| 70 | 153 | 73 | 155 |
| 67 | 145 | 68 | 150 |
| 71 | 170 | 74 | 148 |
| 72 | 175 | 73.5 | 155 |
| 69 | 175 | | |

a. Verinin grafiğini çiziniz.

b. Bu veri için boydan yordanan ağırlığın regresyon eşitliğini hesaplayınız. Eğim ve sabit değeri yorumlayınız.

c. Bu veri için korelasyon katsayısı nedir?

d. Korelasyon katsayısı ve eğim sıfırdan manidar derecede farklı mıdır?

9.32 Aşağıdaki veri bu bölümde söz edilen üniversite kız öğrencilerinin gerçek ağırlık ve boylarına ait kümeleridir.

a. Verinin grafiğini çiziniz.

b. Bu veri için regresyon katsayısını hesaplayınız. Eğim ve sabit değeri yorumlayınız.

c. Bu veri için korelasyon katsayısı nedir? Eğim sıfırdan manidar olarak farklı mıdır?

| Boy | Kilo | Boy | Kilo |
|---|---|---|---|
| 61 | 140 | 65 | 135 |
| 66 | 120 | 66 | 125 |
| 68 | 130 | 65 | 118 |
| 68 | 138 | 65 | 122 |
| 63 | 121 | 65 | 115 |
| 70 | 125 | 64 | 102 |
| 68 | 116 | 67 | 115 |
| 69 | 145 | 69 | 150 |
| 69 | 150 | 68 | 110 |
| 67 | 150 | 63 | 116 |
| 68 | 125 | 62 | 108 |
| 66 | 130 | 63 | 95 |
| 65.5 | 120 | 64 | 125 |
| 66 | 130 | 68 | 133 |
| 62 | 131 | 62 | 110 |
| 62 | 120 | 61.75 | 108 |
| 63 | 118 | 62.75 | 112 |
| 67 | 125 | | |

© Cengage Learning 2013

9.33 Kendi boyunuzu kullanarak Alıştırma 9.31 veya 9.32'den uygun regresyon eşitliğini kullanarak ağırlığınızı yordayınız.

a. Gerçek ağırlığınız sizin yordanan ağırlığınızdan ne kadar çok veya azdır? (Artık değeri zaten hesapladınız.)

b. Verinin alındığı öğrenci grubunun yanlı raporlaması, kendi ağırlığınızı yordamanızda nasıl bir etkiye neden olur?

9.34 Kendi cinsiyetinizdeki öğrenciler için verinin dağılım grafiğini kullanınız ve artıkların büyüklüğünü elde ediniz. (İpucu: Artıkları regresyon doğrusundan dikey uzaklıktaki noktalarda görebilirsiniz.) Dağılım grafiğiniz için en geniş artık nedir?

9.35 Her ikisinin de boyu 5'6˝ olarak verilen bir erkek ve bir kız öğrenci için ağırlığın birbirinden ne kadar farklı olması beklenir? (İpucu: Her biri için kendi cinsiyetine özgü regresyon eşitliğini kullanarak yordanan ağırlığı hesaplayınız.)

9.36 Erkeklerin boylarından ağırlıklarının yordanmasında kullanılan eğim ($b$), kadınlar için elde edilen eğimden büyüktür. Bu manidar mıdır, eğer öyleyse ne anlama gelir?

9.37 Bölüm 2'de verdiğim basamağın, bir karşılaştırma kümesinin parçası olup olmadığına karar verme hızının verisi sunulmuş ve karşılaştırma kümesinin bir, üç veya beş basamaktan oluştuğu denemelerden veri aktarmıştım. Üç koşul kıyaslamak istemekteyim (yalnızca uyarıcı basamağı kümenin bir parçası olan denemelerin verisini kullanarak) fakat denemelerin bağımsız olmadığından şüphelenmekteyim. Eğer birey görev devam ederken kendini geliştiriyorsa, sonraki denemelerde daha iyi yapacaktır ve nasıl yaptığı deneme sayısına bağlı olacaktır. Böylece cevapların bağımsız olduğu söylenemez. Sadece karşılaştırma kümesinde beş basamak olan durumda, $Y$ ile etiketlenmiş denemelerin verisi kullanılarak, deneme sayısındaki cevabın regresyonunu elde ediniz. Denemeler boyunca performans önemli ölçüde gelişmekte midir? Zamana bağlı olarak sistematik doğrusal bir eğilim olmadığını varsayabilir miyiz?

## Tartışma Soruları

9.38 Yakın zamandaki bir e-mailde birisi, yanyana duran ve bütün gün veri toplayan iki hava kirliliği gözlemcisini nasıl karşılaştırabileceğini sordu. 50 günün her birisi için gözlemci başına ortalama okumayı elde etmişler ve iki gözlemciyi karşılaştırmak istemişlerdir. İlk düşünceleri iki gözlemcinin okuma ortalamaları arasında bir $t$ testi yapmaktı. Eğer hava kirliliğinin iki ölçümü için, dışa dönüklüğün iki ölçümünü ve iki ölçümün aynı 50 birey için gerçekleştirilip veri toplanmış olduğunu düşünürsek bu soru hem psikologlar hem de diğer davranış bilimciler için eşit derecede önemlidir. Gözlemcileri (veya ölçümleri) nasıl karşılaştırırsınız? Hangi tür sonuçlar sizi benzer veya farklı ölçüm yaptıkları sonucuna götürür? Bu ilk görüldüğünden daha karmaşıktır, böylelikle sadece $t$ testi yapılması veya korelasyon katsayısı elde edilmesini düşünmezsiniz. Eğer veri ile ilgilenmek isterseniz örnek veri Web sayfasında AirQual.dat isimli dosyada yer almaktadır.

9.39 2005 yılında Plüton'un ötesinde önce Xena şimdi de Eris olarak adlandırılan bir cisim keşfedildi. Bu cisim Plüton'dan daha büyüktür ancak gezegen olarak düşünülmemektedir. Yeni adı "plütoid" dir. Güneş'ten 96,7 astronomik birim uzaktadır. Bir cisim Tablo 9.7'deki veriye nasıl uyar?

9.40 1801 yılında Güneş'ten 2,767 astronomik birim uzaklıkta Ceres adı verilen bir gök cismi Giuseppi Piazzi tarafından keşfedildi. Bu cisim cüce gezegen olarak adlandırılmıştı, ancak şimdi plutiod olarak adlandırılmaktadırlar. Eğer bu cisim bir gezegen olarak sınıflandırılsaydı, Tablo 9.7'de gösterilen bildiğimiz diğer gezegenlerle nasıl uyum sağlardı?

9.41 Popülerlik ve akademik başarı ile ilgili örneğe dönersek, erkekler ve kadınlarda algılanan popülerlik ve sosyometrik popülerlik arasındaki korelasyonları karşılaştırmak için uygun testi çalıştırınız.

9.42 Regresyonda güven aralığı ve yordama aralığı arasındaki farklılığı anlamak genelde zordur. Farklılığın sade bir anlamını bulmak için Google araması yapınız.



Bölüm 10

# Diğer Korelasyon Teknikleri

## Hedefler

İki değerli değişkenlerde, sıralama ölçeğindeki veri kümelerinde korelasyon ve regresyonu tartışma ve kategorik veri kümeleri arasındaki ilişki ölçülerini tanıtma.

## Konular


**10.1** Nokta-Çiftserili Korelasyon ve Fi (Phi) Pearson Korelasyonlarının Diğer Adı

**10.2** Çift Serili ve Tetrakorik Korelasyon: Pearson-Dışı Korelasyon Katsayıları

**10.3** Sıralı Veriden Oluşan Veri Kümeleri İçin Korelasyon Katsayıları

**10.4** Sıralı Veriden Oluşan Çapraz Tabloların Analizi

**10.5** Kendall'ın Konkordans Katsayısı (*W*)

PEARSON MOMENTLER ÇARPIMI KORELASYON KATSAYISI ($r$) mevcut korelasyon katsayılarından sadece birisidir. Bu katsayı genellikle, her iki değişkenin hemen hemen sürekli ölçek üzerinde ölçüldüğü, bazı normallik ve varyansların homojenliği varsayımlarının kabul edildiği, iki değişken arasındaki ilişkinin temel olarak doğrusal olduğu durumlarda kullanılır. Bu bölümde gösterileceği gibi, $r$ farklı adlar altında ve çoğu zaman gerçekte ne olduğunun anlaşılmadığı durumlarda da anlamlı bir şekilde yorumlanabilmektedir.

Bu bölümde farklı veri türlerine uygulanan çeşitli katsayıları tartışacağız. Örneğin; veri, sıraları temsil edebilir, değişkenlerden biri ya da ikisi iki kategorili olabilir veya kategorik olabilir. Verinin temel doğası hakkında yapmak istediğimiz varsayımlara bağlı olarak, farklı katsayıların değişik durumlarda kullanılması uygun olur. Bu katsayıların bazıları Pearson $r_s$ gibi hesaplanırken, bazıları öyle hesaplanmayacaktır. Önemli olan nokta, bunların tümünün iki değişken arasındaki ilişki ölçülerinin elde edilmesi için kullanılması ve *regresyon*dan çok *korelasyon* genel başlığı altında yer almasıdır.

**ilişkisel ölçümler**

**ilişkilerin (benzerliklerin) ölçümü**

Bu değişkenlerin doğasında herhangi bir sınırlama olmaksızın iki değişken arasındaki ilişkiden söz ederken, **ilişkisel ölçümler** ve **ilişkilerin (benzerliklerin) ölçümü** arasındaki farkı ayırt etmek zorundayız. Yüksek puanların daha fazla (az) özelliği temsil etmesi gibi, her bir değişken düzeyine en azından belli bir sıra verilebildiğinde korelasyondan söz edilebilir. Bir değişkendeki artışlar, başka bir değişkendeki artışlarla ilişkili olduğunda anlamlı artıştan söz edebiliriz. Bununla birlikte, pek çok durumda bir değişkenin farklı düzeyleri, bazı miktarlarda düzenli bir artış veya azalış göstermez. Örneğin, insanları farklı kampüs organizasyonlarındaki üyeliklerine ve daha sonra bazı konulardaki görüşlerine göre sıralayabiliriz. Bundan sonra, insanların organizasyonlardaki üyelikleri ve görüşleri arasında bir ilişki bulabiliriz, ancak bu değişkenlerin ikisi de düzenli bir süreklilik göstermeyebilir. Bu gibi durumlarda, hesaplayacağımız katsayı korelasyon katsayısı değildir. Onun yerine, bundan ilişkinin ölçümü olarak söz edeceğiz.

**geçerlilik**

Her türlü korelasyon katsayısının hesaplanmasının üç temel nedeni vardır. En önemlisi olmasa da en belirgin nedeni, evrene ilişkin korelasyonun $r$ kestirimini elde etmektir. Böylece, bir testin **geçerliği**yle ilgilenen bir kişi, aslında kendi testi ve bazı ölçütler arasındaki gerçek ilişkiyi önemser ve bu amaçla bir katsayıyı hesaplama yoluna gider. Bu kullanım amaca hizmet etse de, alternatif tekniklerin en az tatmin edici olanıdır.

Korelasyon katsayılarının ikinci kullanımı, çoklu regresyon ve faktör analizi gibi tekniklerde ortaya çıkar. Bu durumda korelasyonun kendisi son üründen ziyade daha sonraki istatistiklerin hesaplanabilmesi için bir girdi olarak kullanılır. Bu amaçla ele alınacak katsayıların pek çoğu tatmin edicidir.

Korelasyon katsayısının hesaplanmasındaki son neden, bir değişkendeki değişkenliği başka bir değişkendeki değişkenlik tarafından açıklayabilmek için, katsayının karesinin kullanılmasıdır. Bu, etki büyüklüğünün bir ölçümüdür (ölçümlerin $r$-ailesinden) ve bulduğumuz etkinin büyüklüğünü ifade etmede sıklıkla yararlanırız. Burada, bu amaç için tartışılacak katsayılar birçok durumda tatmin edicidir. Aşağıda, özellikle $r$-ailesi, etki büyüklüğünün ölçümü tartışılacaktır.

## 10.1 Nokta-Çift Serili Korelasyon ve Fi (Phi): Pearson Korelasyonlarının Diğer Adı

Önceki bölümde standart Pearson momentler çarpımı korelasyon katsayısını ($r$) her iki ölçmeden elde edilen ölçümlerin nispeten sürekli olduğu durumlarda tartıştım. Ancak, bu aynı formül ayrıca bir veya iki ölçümün iki kategorili (iki düzeyli) olan değişken çifti için

de geçerlidir. Yorumlamada biraz daha dikkatli olmamız gerekebilir, bu korelasyonlar ve diğer ele alınan istatistikler arasında bazı ilginç ilişkiler bulunur, ama bu temel prosedür daha genel durumlar için kullanıldığı gibi özel durumlar için de kullanılabilir.

## Nokta-Çift Serili Korelasyon ($r_{nç}$)

**iki kategorili**

Değişkenler sıklıkla **iki kategorili** formda ölçülebilir, örneğin, erkek-kadın, geçti-kaldı, deney grubu-kontrol grubu vb. Bu değişkenler nadiren sayısal olarak ölçülür (küçük bir sorun), onların sürekli olarak da ölçülemediği oldukça açıktır. Bunu, böyle normal dağılım gibi sürekli bir dağılım olarak varsayabilecek hiçbir yol yoktur, örneğin, erkek-kadın gibi iki kategorili değişkenler üzerinden elde edilen puanları ele alalım. Değişkenler arasındaki ilişkinin bir ölçüsü olarak $r$'yi kullanmak isterseniz, açıkça bir probleminiz vardır, çünkü $\rho$'nun kestirimi gibi $r$ için arzu edilen bazı özelliklere sahip olmak, $X$ ve $Y$'nin birleşik evrenindeki yaklaşık normalliğini varsaymak gerekir.

$X$'in sayısal olarak ölçülmesindeki zorluk, iki kategorili değişkenler için önemsiz hale gelir. Örneğin, $X$'in evlenmişe karşı evlenmemişi belirttiği ve evlenmişlerin 0 evlenmemişlerin 1 ya da tam tersi olarak puanlandığını düşünelim. (Aslında herhangi iki değer olabilir, basit bir nedenle, sayısal olarak kolay olduğundan, 0 ve 1 tercih edilir, bazen 1 ya da 2 de olabilir.) Korelasyonun işaretinin sadece 0 ve 1'i atamak için seçtiğimiz keyfî yola bağlı olması açıktır ve böyle bir sistemi göz önüne aldığımızda nicelikleme, çoğu amaçla anlamsızdır.

**nokta-çiftserili katsayı**

Bölümün sonuna kadar $r$ problemini $\rho$'nun bir kestirimi olarak bir kenara bırakırsak, işler daha da belirginleşir. Bu **nokta-çiftserili katsayısı** $(r_{nç})$ şeklinde isimlendirilmesine rağmen, başka bir amaç için, alışıldığı gibi standart Pearson korelasyon katsayısını ($r$) hesaplamaya devam edebilirsiniz. Bu nedenle, değişkenlerden birisi iki kategorili ve diğeri yaklaşık olarak sürekli ve diziler daha az veya daha fazla normal dağıldığında, cebirsel olarak $r_{nç} = r$ olur.[1] Nokta-çift serili korelasyon katsayısını hesaplamak için özel formüller vardır, ancak böyle yapmakla hiçbir şey kazanılmamaktadır ve bu sadece öğrenmek ve hatırlamak için ek bir şeydir.

## $r_{nç}$'nin Hesaplanması

İnternet üzerindeki istatistiksel tartışma grupları arasındaki yaygın sorulardan biri, "Nokta-çiftserili korelasyonu hesaplayan bir programı bilen var mı?" şeklindedir. Cevabı çok basittir - bilinen herhangi bir istatistiksel paket nokta-çiftserili korelasyonu hesaplar çünkü bu sadece Pearson $r$ verisine uygulanan özel bir türdür.

Nokta-çiftserili korelasyonun hesaplanmasında, Tablo 10.1'de yer alan veri kümesini örnek olarak kullanacağız. Burada, Alıştırma 9.31 ve 9.32'deki erkek kilolarını belirten ilk 12 veri (Cinsiyet = 0) ve kadın kilolarını belirten ilk 15 veri (Cinsiyet = 1) yer almaktadır. Bunu yapmanın mümkün olduğunu göstermek için kadın ve erkek kiloları eşit sayıda seçilmemiştir. Bunun gerçek konulardan alınan gerçek kişisel rapor verisi olduğunu aklınızda tutunuz.

Bu veri kümesine dağılım grafiği ve regresyon doğrusu Şekil 10.1'de aynı grafikte üst üste verilmiştir. Burada 27 noktadan daha az nokta bulunmaktadır çünkü bazı noktalar üst üste binmiştir. Regresyon doğrusunun her bir ışının ortalamasından geçtiğine dikkat ediniz. Sabit, $X = 0$ olduğunda, $\hat{Y}$'deki değer olmakta ve erkeklerin ortalama ağırlıklarına

[1] Orada açık bir ölçüt değişken olduğunda ve bu değişken iki kategorili (örneğin, başarılı-başarısız) olduğunda, lojistik regresyonu göz önünde bulundurmak isteyebilirsiniz (Bölüm 15'e bakınız).

**Tablo 10.1** Erkek ve kadın ağırlıkları için nokta-çift serili korelasyonunun hesaplanması

| Cinsiyet | Kilo | Cinsiyet | Kilo |
|---|---|---|---|
| 0 | 150 | 1 | 130 |
| 0 | 140 | 1 | 138 |
| 0 | 180 | 1 | 121 |
| 0 | 190 | 1 | 125 |
| 0 | 145 | 1 | 116 |
| 0 | 150 | 1 | 145 |
| 0 | 164 | 1 | 150 |
| 0 | 140 | 1 | 150 |
| 0 | 142 | 1 | 125 |
| 0 | 136 | 1 | 130 |
| 0 | 123 | 1 | 120 |
| 0 | 155 | 1 | 130 |
| 1 | 140 | 1 | 131 |
| 1 | 120 | | |

$\text{Ortalama}_{\text{erkek}} = 151{,}25$

$s_{\text{erkek}} = 18{,}869$

$\text{Ortalama}_{\text{kilo}} = 140{,}222$

$s_{\text{kilo}} = 17{,}792$

$\text{kov}_{XY} = -5{,}090$

$\text{Ortalama}_{\text{kadın}} = 131{,}4$

$s_{\text{kadın}} = 10{,}979$

$\text{Ortalama}_{\text{cinsiyet}} = 0{,}556$

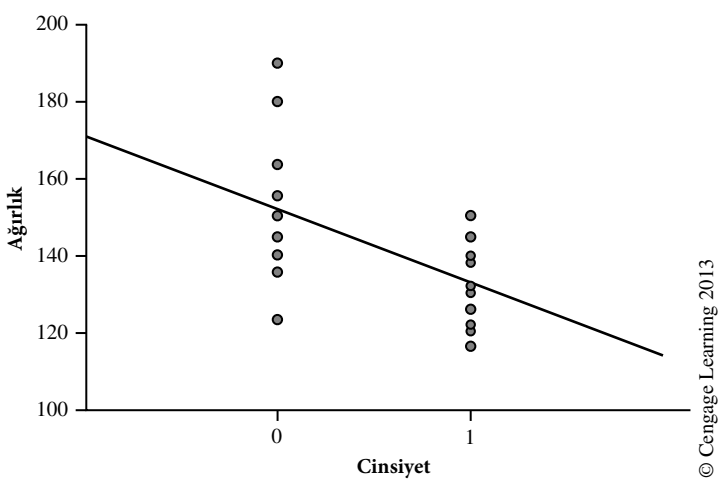
$s_{\text{cinsiyet}} = 0{,}506$

$$r = \frac{\text{kov}_{XY}}{s_X s_Y} = \frac{-5.090}{(0{,}506)(17{,}792)} = -{,}565$$

$$b = \frac{\text{kov}_{XY}}{s_X^2} = \frac{-5{,}090}{(0{,}506)^2} = -19{,}85$$

$$a = \overline{Y} - b\overline{X} = 151{,}25$$

© Cengage Learning 2013




© Cengage Learning 2013

**Şekil 10.1** Cinsiyetin bir fonksiyonu olarak ağırlık

eşit olmaktadır ve $X = 1$ olduğunda, $\hat{Y}$, kadınların ortalama kilolarına eşit olmaktadır. Bu değerler korelasyon katsayısıyla birlikte Tablo 10.1'de gösterilmiştir. Doğrunun eğimi negatiftir, çünkü "kadın" = 1 kodlanmıştır ve bu nedenle kadınlar erkeklerin sağına çizilmiştir. Eğer puanlama tersine çevrilirse, eğim pozitif olacaktı.

Tablo 10.1'de ağırlık ve cinsiyet arasındaki korelasyonun -,565 olduğunu görebilirsiniz. Belirtildiği gibi cinsiyeti kodlama konusundaki kararın keyfi olmasından dolayı korelasyonun işaretini göz ardı edebilirsiniz. Negatif katsayı, 1 kodlu grubun ortalamasının, 0 kodlu grubun ortalamasından daha düşük olduğunu gösterir, pozitif korelasyon ise bu durumun tam tersini belirtir. Alışıldığı gibi $r^2$ yi, ağırlıktaki değişkenliğin $(-,565)^2$ = %32'sinin cinsiyet tarafından açıklanabileceği şeklinde yorumlayabiliriz. Burada neden ve sonuçtan söz edilmemektedir. Ağırlığın daha öncelikli nedenlerinden biri ise, kesinlikle cinsiyetle ilgili olan erkeklerin boylarındaki fazlalıktır, ancak buradaki resimde cinsiyete bağlı diğer özelliklerin de çok fazla payı bulunmaktadır.

Şekil 10.1 ile ilgili diğer bir ilginç durum, regresyon doğrusunun denklemiyle ilgilidir. Sabitin, $X = 0$ olduğunda $\hat{Y}$'deki değer olduğunu hatırlayınız. Bu durumda, erkekler için $X = 0$ olduğunda erkekler için yordanan $\hat{Y} = 151,25$ olmaktadır. Başka bir ifadeyle, sabit, 0 kodlu grubun ortalama ağırlığı olmaktadır. Ayrıca regresyon doğrusunun eğimi, $X$'teki bir birimlik değişikliğin $\hat{Y}$'deki değişimi olarak tanımlanır. Çünkü $X$'teki bir birimlik değişim erkekten kadına bir değişime karşılık gelir ve yordanan değer ($\hat{Y}$) erkeklerin ortalama ağırlığından kadınların ortalama ağırlığına değişir, eğim (-19,85), iki ortalama arasındaki farkı belirtir. Bölüm 16'da bu kısma tekrar döneceğiz, ancak basit olarak bunun farkına varmak önemlidir.

## $r_{pb}$ ve $t$ Arasındaki İlişki

$r_{pb}$ ve $t$ arasındaki ilişki çok önemlidir. İspatı burada verilmeyecek olsa da aşağıdaki gibi gösterilebilir:

$$r^2_{nç} = \frac{t^2}{t^2 + sd}$$

$t$, ortalamalar arası fark (örneğin, erkek ve kadınların ağırlıklarının ortalamaları arasındaki) $t$ testinden elde edilir, $sd = t$ için serbestlik derecesini yani $N_1 + N_2 - 2$'yi belirtir. Örneğin, kadınlar ve erkeklerin ağırlıklarının arasındaki ortalama fark için $t$ testi yapmak istediğimizde, eşit olmayan örneklem büyüklüklerinde bağımsız iki grup için $t$ testi kullanırız.

$$s^2_{birleşik} = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

$$= \frac{11(18,869^2) + 14(10,979^2)}{12 + 15 - 2} = 224,159$$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s^2_{birleşik}}{N_1} + \frac{s^2_{birleşik}}{N_2}}}$$

$$= \frac{151,25 - 131,4}{\sqrt{\frac{224,159}{12} + \frac{224,159}{15}}}$$

$$= \frac{19,85}{5,799} = 3,42$$

25 *sd* ile iki grup arasındaki fark manidardır. Şimdi hesapladığımızda,

$$r_{sd}^2 = \frac{t^2}{t^2 + sd} = \frac{3,42^2}{3,42^2 + 25} = ,319$$

$$r_{sd} = \sqrt{,319} = ,565$$

katsayının keyfi işareti dışında doğrudan hesaplamayla uyuşur.

## $r_{nç}$'nin Manidarlık Testi

Sıfır hipotezi $H_0 : \rho = 0$'a karşı, $r_{nç}$'nin testini oluşturmak kolaydır. Çünkü $r_{nç}$ Pearson momentler-çarpımı katsayısıdır, herhangi bir *r* gibi test edilebilir. Yani, *N-2* serbestlik derecesinde aşağıdaki test istatistiği yazılabilir.

$$t = \frac{r_{nç}\sqrt{N-2}}{\sqrt{1 - r_{nç}^2}}$$

Ayrıca, bu eşitliğin $r_{nç}^2$'nin tanımından doğrudan elde edilebilir olmasından dolayı, buradan elde edilen $t = 3,42$, iki kategorili değişkenlerin iki düzeyi arasındaki *t* testiyle aynıdır (muhtemelen işareti hariç). Erkek ve kadınların ağırlıklarındaki farklılığın açıklanması, ağırlığın cinsiyetle değişmesi ifadesiyle aynı olduğunu fark ettiğinizde anlamlı olur.

## $r_{nç}^2$ ve Etki Büyüklüğü

Burada atılacak bir önemli adım daha vardır. Başka bir yerde, Cohen (1988) tarafından formülü aşağıdaki gibi verilen ve bir işlemin etkisinin diğerine göre bir ölçümü olarak tanımlanan etki büyüklüğünü düşünmeliyiz.

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Burada biraz dikkatli olmak zorundayız, çünkü Cohen etki büyüklüğünü parametrelerle açıklamıştır (bunlar, evren ortalaması ve standart sapmasıdır). Diğerleri (Glass (1976) ve Hedges (1981)), örneklem istatistikleri açısından kendi istatistiklerini (sırasıyla *g´* ve *g*) belirtmişlerdir. Hedges payda olarak birleştirilmiş evren varyansının kestirimini kullanırken (birleştirilmiş kestirim için bkz: Bölüm 7) Glass bir grubun standart sapmasını kullanmıştır. Bu etki büyüklüğü ölçümlerinin herhangi biri hakkında güzel olan şey, ortalamalar arası farkı standart sapmanın büyüklüğü açısından ifade etmeleridir. Doğru olması güzel olmakla birlikte, bazen daha açık ve tutarlı olmalıdır. Başka bir yerde yaptığım gibi, Hedges ve Glass'den özür dileyerek, burada etki büyüklüğünün ölçümü için Cohen'ın *d*'sine başvurarak devam edeceğim. Farklı ölçümler ve bunların tanımlarına ilişkin mükemmel tartışma Kline'da (2004) bulunabilir.

Nokta-çift serili korelasyon katsayısının karesi ile *d* arasında doğrudan bir ilişki vardır.

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s_{birleşik}} = \sqrt{\frac{sd(n_1 + n_2)r_{nç}^2}{n_1 n_2 (1 - r_{nç}^2)}}$$

Erkek ve kadın ağırlıkları verisine uyguladığımızda, aşağıdaki ifadeyi elde ederiz:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s_{birleşik}} = \sqrt{\frac{sd(n_1 + n_2)r_{nç}^2}{n_1 n_2 (1 - r_{nç}^2)}}$$

$$= \frac{151,25 - 131,4}{14,972} = 1,33 = \sqrt{\frac{25(12+15)(-,565)^2}{12x15(1-,565^2)}} = \sqrt{1,758} = 1,33$$

Buradan kadın ve erkeklerin ortalama ağırlıkları arasındaki farkın yaklaşık 1 1/3 standart sapma olduğu sonucuna ulaşırız. Bana göre, ağırlıktaki değişkenliğin %32'sinin cinsiyetten geldiğini söylemekten daha anlamlıdır.

Burada önemli olan nokta, bu istatistiklerin anlamlı bir şekilde ilişkili olduğunu görmektir. Hangi istatistiğin daha anlamlı göründüğüne bağlı olarak, $r^2_{nç}$'ten $d$'ye sonra $t$'ye veya tam tersine gidilebilir. Etki büyüklüğü ve benzer ölçümlerin raporlanması üzerine vurgunun artmasıyla, bu ilişkilerin tanınması önemlidir.

## *d* İçin Güven Aralığı

Evren parametresine karşılık gelen $\delta$'nın güven aralığının hesaplanması için bir adım daha ileri gidebiliriz. Bu aralığın belirlenmesinde kullanılacak uygun yol merkezî olmayan $t$ dağılımını kullanmaktır, bu yaklaşımı Kesim 7.5'te açıkladım. Bu yaklaşımı kabul edersek standart $z$ dağılımını kullanabiliriz. Ardından birleştirilmiş standart sapmayı kullanarak ($s_{birleşik}$), standartlaştırılmış ortalama farkın standart hatasını aşağıdaki gibi hesaplayabiliriz.

$$s_{fark} = \sqrt{\frac{d^2}{2(n_1 - 1)(n_2 - 1)} + \frac{N}{n_1 n_2}}$$

$z_{,025}$ = 1,96 ile %95 olasılıklı güven aralığı

$d \pm s_{fark}(z_{0,25}) = d \pm s_{fark}$ x 1,96

eşitliğiyle verilir. Elimizdeki erkek ve kadınların ağırlıklarıyla ilgili veri için

$d = 1,33$

$$s_{fark} = \sqrt{\frac{d^2}{2(n_1 - 1)(n_2 - 1)} + \frac{N}{n_1 n_2}}$$

$$= \sqrt{\frac{1,33^2}{2(11)(14)} + \frac{27}{12x15}} = \sqrt{0,0057 + 0,150} = 0,395$$

$$1,33 \pm 0,395 \times 1,96 = 1,33 \pm 0,774$$

$$0,556 \leq \delta \leq 2,10$$

(Cumming'in ESCI programı tarafından verilen daha kesin aralık $0,4725 \leq \delta \leq 0,774$'tür. Kitabın Web sitesinde bulunan, bu aralıkları hesaplamak için R'da yazılmış program, temelde aynı sonucu verir.)

## Fi (Phi) Katsayısı ($\phi$)

Nokta-çift serili korelasyon katsayısı, değişkenlerden herhangi birinin iki kategorili olması durumuyla ilgilenir. İki değişken de iki kategorili olduğunda, başka bir istatistiğe ihtiyaç duyulur. Örneğin; cinsiyet ve çalışma durumuyla ilgilenildiğinde, bireylerin cinsiyeti kadın-erkek ve çalışma durumu da çalışan-işsiz şeklinde puanlanır. Benzer şekilde, bireylerin çalışma durumu (çalışan-işsiz) ile içkili araba sürüp tutuklanıp tutuklanmama durumları arasındaki ilişkiyle ilgilenilebilir. Son bir örnek olarak, sigara içme (sigara içenler - sigara içmeyenler) ile kanserden ölme (başka bir nedenden ölme) arasındaki korelasyonu bilmek isteyebilirsiniz. Değişkenlerin sürekliliğiyle ilgili özel varsayımlar yapmadığımız sürece,

**Tablo 10.2** Gibson verisi için f'in hesaplanması

X: 0 = Eğitim Alan
1 = Eğitim Almayan
Y: 0 = Cinsel İstismara Uğramış
1 = Cinsel İstismara Uğramamış

**Kısmi Veri:**

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **X:** | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| **Y:** | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |

**Hesaplamalar (tüm veri kümesi üzerinden):**

$\overline{X} = 0{,}3888$ $s_X = 0{,}4878$ $\text{kov}_{XY} = -0{,}0169$

$\overline{Y} = 0{,}8863$ $s_Y = 0{,}3176$ $N = 818$

$$\phi = \text{r} = \frac{\text{kov}_{XY}}{s_X s_Y} = \frac{-0{,}0169}{(,4878)(,3176)} = -,1094$$

$$\phi^2 = ,012$$

© Cengage Learning 2013

**$\phi$ katsayısı**

en uygun korelasyon katsayısı **fi ($\phi$) katsayısı**dır. Bu, Bölüm 6'da kısaca sözü edilen $\phi$ ile aynıdır.

## $\phi$'nin Hesaplanması

Tablo 10.2, Gibson ve Leitenberg'ın (2000) okuldaki cinsel istismar eğitimi (bazılarınız "yabancı kaynaklı tehlike" ya da "iyi dokunuş-kötü dokunuş" şeklinde hatırlayabilir) ve sonraki cinsel istismar arasındaki ilişkiye ait veri kümelerinin (Alıştırma 6.33'te kullanılmıştı) küçük bir kısmını içerir. Her iki değişken 0, 1 –birey eğitim almış ya da almamış, istismara uğramış ya da uğramamış- şeklinde puanlanır.

Uygun olan korelasyon katsayısı $\phi^2$ katsayısıdır, bu veri üzerinden hesaplanan Pearson'ın *r*'sine eşdeğerdir. Yine özel formüller kullanılabilir ama gereksizdir.

Tablo 10.2'de, okulda cinsel istismar nasıl önlenir üzerine eğitim alan almayan öğrenci ile sonradan cinsel istismara uğrayan uğramayan öğrenci arasındaki korelasyonun –,1094 ve $\phi^2 = ,012$ olduğu görülmektedir. Korelasyon doğru yönde ancak, çok etkileyici görünmemektedir. Ama bu yanıltıcı olabilir. (Bu veriyi kullanmayı seçtim çünkü, bu bir açıdan çok küçük bir etki, başka bir açıdan da çok daha büyük bir etki gibi görünebilir.) Kısaca bu konuya geri dönülecektir.

## $\phi$'nin Manidarlığı

$\phi$ hesaplandıktan sonra, istatistiksel manidarlık için, bunun test edilmesi gerekir. $N\phi^2$'nin dağılımı 1 *sd*'nde $\chi^2$ olduğundan $H_0 = \rho = 0$ hipotezini test etmek için uygun olan test bir ki-kare testidir, Verimiz için,

$$X^2 = N\phi^2 = 818(-,1094^2) = 9,79$$

olur. Bu istatistik bir *sd*'de manidardır. Bu nedenle, cinsel istismar eğitimi ve sonraki istismar arasındaki ilişkinin inandırıcı bir kanıtının olduğu sonucuna varırız.

Alternatif olarak aşağıdaki bağıntıdan, $\chi^2$'den $\phi$'ye gidebiliriz

$$\phi = \sqrt{\frac{X^2}{N}}$$

**Tablo 10.3** Gibson'ın cinsel istismar üzerindeki verisi için $\chi^2$ hesaplaması ($\chi^2$ tablodaki yuvarlama hataları nedeniyle "yaklaşık" olarak basitçe gösterilmektedir).

| | Eğitim Almış | Eğitim Almamış | |
|---|---|---|---|
| İstismarlı | 43 (56,85) | 50 (36,15) | 93 |
| İstismarsız | 457 (443,15) | 268 (281,85) | 725 |
| | 500 | 318 | 818 |

$$x^2 = \frac{(43 - 56,85)^2}{56,85} + \frac{(50 - 36,15)^2}{36,15} + \frac{(457 - 443,15)^2}{443,15} + \frac{(268 - 281,85)^2}{281,85}$$
$$= 9,79 \text{ (yaklaşık)}$$

© Cengage Learning 2013

Bizim örneğimiz için,

$$\phi = \sqrt{\frac{9,79}{818}} = \sqrt{0,0120} = ,1095$$

(çok az yuvarlama ile), önceki hesaplamayla uyum göstermektedir.

## $\phi$ ve $X^2$ Arasındaki İlişki

Tablo 10.3'te gösterildiği gibi, Tablo 10.2'nin temelini oluşturan veri kümesi başka bir formda yeniden biçimlendirilebilir. İki tablo (10.2 ve 10.3) aynı bilgiyi içerir, sadece farklı gösterir. Tablo 10.3'ün bir çapraz tablo olduğunu hemen tanıyacaksınız. Bunun dışında, değişkenlerin bağımsız olduğu yokluk hipotezini test etmek için $\chi^2$'nin değeri hesaplanabilir. Bunu yaparken hesaplanan $\chi^2 = 9,79$ 1 *sd*'de manidar bulunmuştur. Bu aynı zamanda bir sonraki kesimde hesaplanacak olan $\chi^2$ ile aynı değerdir.

$\phi$ ve $\chi^2$ hesaplanırken, aynı soruyu iki farklı şekilde soruyoruz: Sürpriz olmayarak aynı sonuca ulaştık. Aynı soruyu $\phi$'yi hesapladığımızda ve manidarlığını test ettiğimizde, $\chi$ ve $Y$ arasında herhangi bir korelasyonun (ilişkinin) olup olmadığını sorarız. Tablo 10.3'teki ki-kare testini yaptığımızda da değişkenler arasında ilişkinin (korelasyonun) olup olmadığını sorarız. Bu sorular aynı olduğundan aynı cevabın geleceğini umarız. Bir yandan $\chi^2$, ilişkinin manidarlığıyla ilgilidir; öte yandan $\phi$, bu ilişkinin derecesini veya büyüklüğünü ölçer.

## $\chi^2$'nin Manidarlığının Pratik Ölçüsü Olarak $\phi^2$

Aslında $\chi^2$'den $\phi$'ye gitmemizin anlamı, iki kategorili iki değişken arasındaki ilişkinin pratik manidarlığının değerlendirilmesinin bir yoludur. Gibson'ın verisi için $\chi^2$'den $\phi^2$'ye dönüşümde, 9,79 olan $\chi^2$'nin, değişimin yaklaşık %1,2'sinden sorumlu olduğunu daha önce görmüştük. Dediğim gibi, bu manidar olsa bile, çok etkileyici görünmemektedir.

Rosenthal ve Rubin (1982), "sosyal/yumaşak bilimlerdeki" psikologların ve diğerlerinin; $\phi^2$ ve $r^2$'nin küçük değerlerine bakıp bunu önemsiz bir etki gibi etiketlemek için çoktan hazır olduklarını iddia etmektedirler. Bunlar, $r^2$'nin çok küçük değerlerinin aslında önemli etkilerle ilişkili olabileceğini iddia ederler. Bunların çalışmasının burada tartışılmasın nedeni, bu çalışmayı $\phi$'ye göre ifade etmenin çok kolay olmasıdır.

Rosenthal ve Rubin, psikoterapinin etkililiğinde, 400'den fazla çalışmanın geniş ölçekli değerlendirilmesine (meta-analiz olarak adlandırılır) dikkat çekmiştir. Yazarlar, Smith

ve Glass (1977), ne olursa olsun ölçüm ve psikoterapinin varlığı veya yokluğu ile gelişmenin varlığı veya yokluğu arasında ,32 korelasyona eşdeğer bir etki bildirmişlerdir. Bir eleştirmen daha sonra bu ilişkinin karesini ($r^2$ = ,1024) almış ve sonuçtaki değişkenliğin sadece %10'unun psikoterapiden oluştuğuna üzülmüştür. Rosenthal ve Rubin, eleştirmenin anlayışından etkilenmediler. Onlar, 100 bireyi kontrol grubuna, 100 bireyi deney grubuna alıp, onları gelişmiş ya da gelişmemiş şeklinde ikiye ayırdıklarında, $\phi$ = ,32 korelasyon katsayısının $\chi^2$ = 20,48'e karşılık geleceğine dikkat çekmişlerdir. Bu, aşağıdaki hesaplama yapılarak görülebilir:

$$\phi = \sqrt{\chi^2)N}$$
$$\phi = \chi^2/N$$
$$,1024 = \chi^2/200$$
$$\chi^2 = 20,48$$

İlginç bir gerçek de, böyle bir $\chi^2$'nin, deney grubundaki 100 denekten 64'ünün, kontrol grubundaki 100 denekten 36'sının geliştiği çapraz tablodan kaynaklanmaktadır. (Böyle bir tabloda kendiniz $\chi^2$ hesaplayarak bunu kolaylıkla gösterebilirsiniz.) Bu gelişme oranlarında çarpıcı bir fark bulunmaktadır.

Ancak, bende daha çok örnek vardır. Rosenthal (1990), iyi bilinen bir çalışmada, kalp krizi sıklığını azaltmak için günlük dozda aspirin ya da etkisiz ilâç alan (erkek) hekimlere dikkat çekmiştir. (Daha önceki bölümlerde bu çalışmayı farklı bir amaç için değerlendirmiştik.) Bu çalışma, kontrol grubuna etkisiz ilâç vermeye devam edilmesinin etik olmaması ve bu durumun çok açık şekilde aspirin grubu lehine olmasından dolayı, gözden geçirilme paneli değerlendirme sonuçları sonucunda erken sonlandırılmıştır. Fakat Rosenthal, çalışmayı kısa kesmek için etkili olan aspirin ve kalp krizi arasındaki korelasyonun ne olduğunu sordu. $\phi$ = ,034 ($\phi^2$ = ,001) olduğuna inanır mısınız?

Rosenthal ayrıca AIDS tedavisi için AZT çalışmasına dikkat çekti. AZT ölümleri %61,5'ten %38,5'e kadar azalttığından çalışma durduruldu. Ancak $r^2$ sadece ,05 oldu. Rosenthal ve diğerleri $r^2$ yerine $r$'nin referans olarak alınabileceği şeklindeki öneriler gibi, söylenecek çok şeyin olduğunu belirttiler.

Rosenthal'ın çalışması, önemli bir etkiye sahip olmak için $r^2$ veya $\phi^2$'nin büyük değerlerinin gerekmediğini içerir. Bazı durumlarda küçük değerler oldukça etkili olabilir. Diğer örnekler için Rosenthal'a (1990) bakınız.

Gibson'ın cinsel istismar veri kümesinde küçük etki gibi görünene geri dönmek için Bölüm 6'da olasılık oranlarıyla benimsenen yaklaşımı ele alacağız. Gibson'ın verisinde, herhangi bir eğitim verilmeyen 318 çocuktan 50'si dışındakiler sonradan istismara uğramıştır, bu gruptaki istismar riski 50/318 = 0,157'dir. Diğer taraftan, eğitim alan 500 çocuktan 43'ü dışındakiler sonradan istismara uğramıştır, olasılık oranı ise 43/500 = 0,086'dır. Bu bize risk oranını (hesaplanan iki riskin oranı) 0,157/0,086 = 1,83 verir. Okulda cinsel istismar eğitimi almamış bir çocuğun sonradan istismar edilmesi yaklaşık iki kat olasıdır. Bu sadece 0,12 korelasyonun karesinden farklı görünüyor, yani neden istatistik seçerken dikkatli olmamız gerektiğini gösteriyor.

Bu noktada belki de kafanız iyice karıştı. Burada iki kategorik veri arasındaki korelasyonun hesaplanması gösterilmeye başlandı. Daha sonra bu ilişkinin ya Pearson korelasyon katsayısı gibi hesaplanması ya da karşılık gelen çapraz tablo üzerindeki ki-kareden doğrudan çıkarılabildiğini gösterdim, çünkü burada $\phi$ ve $\chi^2$ arasında güzel bir ilişki bulunmaktadır. $\phi$ ve $\phi^2$, bağımsız değişkenlerin etkililiğin $r$-ailesi etki büyüklüğü ölçümünün (değişimin bir ölçümünden sorumlu) sağlanabilmesi için kullanılabileceğini savundum. Sonra bir adım daha ileri giderek $\phi^2$ hesapladığınızda, onun nasıl küçük olduğuna şaşırabileceğinizi söyledim. Bu bağlamda, Rosenthal ve Rubin'in çalışmalarına dikkat çektim ve Gibson'ın verisinde iki

farklı yolla, sadece küçük miktarlar için hesaplanan değişimin hâlâ etkileyici ve önemli olabileceğini gösterdim. "Önem"in (istatistiksel manidarlık, açıklanabilir değişkenliğin yüzdesi, etki büyüklüğü [*d*] ve risk oranları) farklı ölçülerini karıştırıyorum ve bunlar zihin karıştırıcı olabilir, bu durum problemin doğasında vardır. İstatistiksel manidarlık iyi bir şeydir ancak her şey demek değildir. Değişkenliğin yüzdesi önemli bir ölçü türüdür ancak çok sezgisel değildir ve önemli durumlarda küçük olabilir. Etki büyüklüğünün ölçülmesinde, *d*-ailesi daha somut bir açıdan (standart sapma açısından ortalamalar arası fark) farklılık sunma avantajına sahiptir. Bağıl risk ve olasılık oranları 2 x 2'lik tablolarda çok kullanışlıdır, fakat daha kompleks veya daha basit durumlarda daha az kullanşlıdır.

## 10.2 Çift Serili ve Tetrakorik Korelasyon: Pearson-Dışı Korelasyon Katsayıları

Nokta çift serili ve fi (phi) katsayıları göz önüne alındığında, bir veya her iki değişkenin iki kategorili olarak ölçüldüğü veri kümelerine bakılmaktadır. Hatta buna "gerçek iki kategorili" diyebilirsiniz, çünkü genellikle böyle değişkenler "ikisinden biri-veya" şeklindeki değişkenlerdir. Bir kişi ya kadın ya da erkek şeklinde sınıflandırılır, bu ikisinin ortası yoktur. Bunlar her zaman iki kategorili veri ile hesaplanan katsayılardır ve neredeyse tüm bilgisayar programları bu katsayıları varsayılan olarak hesaplar.

**çift serili korelasyon**

**tetrakorik korelasyon**

Örneklerde karşılaştığımız ancak kullanımı az olan diğer iki katsayı ise **çift serili** ve **tetrakorik korelasyon**lardır. Kitabın önceki baskılarında bu katsayıların nasıl hesaplandığını göstermiştim, bu yüzden bundan sonra daha fazla bu noktaya değinmeye gerek görünmüyor. Burada tartıştığım katsayıların birbirlerinden nasıl farklılaştığığını kısaca açıklayacağım:

Daha önce söylediğim gibi, insanlara kadın ya da erkek olmalarına, onların bir testte başarılı ya da başarısız olmalarına, istismara uğrayıp uğramadıklarına göre davranırız. Ancak bildiğimiz gibi bu ikiye ayrılmalar, özellikle son ikisi, keyfîdir. İnsanlar, çok kötü başarısız, neredeyse başarısız, neredeyse başarılı vb. şekillerde sınıflanabilir. Başka bir ifadeyle, tüm istismarlar kötü olmasına rağmen bunlar kendi içinde derecelendirilebilir, bazıları diğerlerine göre daha kötüdür. Bunun temelinde yatan sürekliliğin hesaba katılmasında istekli olunduğunda, değişken (değişkenler) ikili dağılım yerine normal dağılmış olsaydı, korelasyonun ne olacağını tahmin edebilirdik.

Çift serili korelasyon, normallik sayıltısı dışında, nokta-çift serili korelasyona benzer. Tetrakorik korelasyon da, temelindeki her iki değişkenin normallik sayıltısıyla, $\phi$ katsayısına doğrudan benzer. Bu iki katsayı hakkında bilmeniz gereken budur.

## 10.3 Sıralı Veri Kümeleri İçin Korelasyon Katsayıları

Bazı deneylerde veri doğal olarak sıralanmış şekilde oluşabilir. Örneğin, yargıcıların iki farklı durumda nesneleri tercihlerine göre sıralamaları istenir ve bu iki sıra kümesi arasındaki korelasyonun ne olduğu merak edilebilir. Kentler sık sık yaşanabilirlik yönünden sıralanabilir ve bu sıralar ile 10 yıl sonraki sıralar arasındaki korelasyona bakılabilir. Bazı sıralama işlemlerinin güvenirliğinin değerlendirilmesi istenildiğinde, genellikle en çok bu korelasyonlarla ilgileniriz, bu kentleri sıralama örneği olsa da sıralamadaki kararlılıkla ilgileniriz.

Geçmişte sıklıkla önerilmiş olan bununla ilgili yöntem, ölçmenin temelindeki ölçeklerin doğası hakkında önemli çekincelerimiz olduğunda, ölçme sonuçlarını sıralamaktır. Bu

durumda, ham puanların yerini sıra sayıları alacaktır. Sıralı ölçme verisinin gerekliliğini önemle sorgulamamıza rağmen (Kesim 1.3'te ölçeklerin tartışılmasında belirtilen nedenlerden dolayı), bu yine de yaygın bir yöntemdir. Uç değerleri daha az ağırlıklandırmanın yararlı etkisi vardır, bu, zaman zaman tercih edilir.

## Veriyi Sıralamak

Öğrenciler zaman zaman veriyi **sıralamada** zorluk çekerler. Kısa bir örnek olarak artan sırayla düzenlenmiş aşağıdaki veri kümesinin verildiğini farzedelim.

[5, 8, 9, 12, 12, 15, 16, 16, 16, 17]

En düşük değer (5) 1. sırada verilmiştir. Sonraki iki değer (8 ve 9) 2 ve 3. sıralara atanmıştır. Daha sonra sıralanması gereken iki eşit (12) değer bulunmaktadır. Eğer bunlar eşit olmasaydı bu değerlere 4 ve 5. sıraları vermemiz gerekirdi, bu nedenle farkı ikiye bölerek her ikisine 4,5 sırası verilir. Altıncı sayıya (15), 6. sıra verilir. Üç eşit değer (16) 7, 8, ve 9. sıralar içindir ve bu sıraların ortalaması 8'dir. Bu durumda hepsine 8. sıra verilir. Son değer olan 17'ye ise 10. sıra verilir. Veri ve bunlara karşılık gelen sıralar aşağıda verilmiştir.

| **X:** | 5 | 8 | 9 | 12 | 12 | 15 | 16 | 16 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sıralar:** | 1 | 2 | 3 | 4,5 | 4,5 | 6 | 8 | 8 | 8 | 10 |

## Sıralı Veri İçin Spearman'ın Korelasyon Katsayısı ($r_S$)

**sıralı veri için Spearman sıra korelasyonu ($r_S$)**
**Spearman'ın ro (rho)'su**

Veri doğal olarak sıralı olduğunda (örneğin, iki farklı durumda 20 kentin sıralamalarına bakarken) veya ham puanlar yerine sıra sayıları kullanıldığında, uygun bir korelasyon, **sıralı veri için Spearman sıra korelasyonu**dur **($r_S$).** (Bu istatistik bazen **Spearman'ın ro**'su **(rho)** olarak adlandırılır.)

## $r_S$'nin Hesaplanması

$r_s$'nin hesaplanmasının en kolay yolu, Pearson'ın orijinal formülünü sıralı veriye uygulamaktır. Alternatif formüller vardır, ancak onlar da veride eşit sayılar olmadığı sürece Pearson'ın formülü ile aynı cevabı vermek için tasarlanmıştır. Eşit sayılar olduğunda, bir düzeltme faktörü uygulanmadığı sürece alternatif formüller yanlış sonuç verir. Bu düzeltme faktörü sizi, Pearson formülünü kullanmaya başladığınız yere geri götürür, neden alternatif formüllerle uğraşasınız?

## $r_S$'nin Manidarlığı

Bölüm 9'da, $r$'nin manidarlığını test etmek (veya güven aralığını saptamak) için normallik ve homojenlik varsayımlarının konulduğunu hatırlayınız. Sıra sayılarında veri, açık olarak normal dağılmayabilir. Küçük örneklemlerde $r_S$'nin standart hatasını hesaplamak için genel kabul görmüş bir yöntem yoktur. Sonuç olarak $r_S$'nin güven aralığının bulunması pratik değildir. $r_S$'nin kritik değerlerinin tablolarını içeren sayısız kitap bulunmaktadır, ancak $N \geq 28$ için, bu tablolar kendi yaklaşık değerlerine dayanır. Bu bağlamda, tipik bir değerlendiricinin çok sayıda öğeyi sıralaması zordur, bu nedenle pratikte $r_S$ kullanılırken $N$ genellikle küçüktür. $r_S$'nin istatistiksel manidarlığı için iyi bir test yoktur; ancak, insanların çoğu işlemlerde buna normal bir Pearson korelasyonu gibi başvurmaktadır ve sınır durumları hakkında dikkatli olunmalıdır.

**Tablo 10.4** Büyük Britanya'da Alkol ve Tütün Harcamaları

| Bölge | Alkol | Tütün | Sıra A | Sıra T | Tersleri |
|---|---|---|---|---|---|
| Northern Ireland | 4,02 | 4,56 | 1 | 11 | 10 |
| East Anglia | 4,52 | 2,92 | 2 | 2 | 1 |
| Southwest | 4,79 | 2,71 | 3 | 1 | 0 |
| East Midlands | 4,89 | 3,34 | 4 | 4 | 1 |
| Wales | 5,27 | 3,53 | 5 | 6 | 2 |
| West Midlands | 5,63 | 3,47 | 6 | 5 | 1 |
| Southeast | 5,89 | 3,20 | 7 | 3 | 0 |
| Scotland | 6,08 | 4,51 | 8 | 10 | 3 |
| Yorkshire | 6,13 | 3,76 | 9 | 7 | 0 |
| Northeast | 6,19 | 3,77 | 10 | 8 | 0 |
| North | 6,47 | 4,03 | 11 | 9 | 0 |

© Cengage Learning 2013

## Kendall'ın To (Tau) Katsayısı ($\tau$)

**Kendall'ın $\tau$'su**

Spearman'ın $r_s$'sine ciddi bir rakip **Kendall'ın** $\tau$'sudur. Spearman, sıraları puanlar gibi ele alır ve iki sıra kümesi arasındaki korelasyonu hesaplarken, Kendall'ın istatistiği, sıralamadaki *çevrilmiş* çift sayısına bakar.

Bizim örneğimizdeki veri kümesi Veri ve Hikâye Kütüphanesi (Data and Story Library-DASL) Web sitesinden alınmıştır; http://lib.stat.cmu.edu/DASL/Stories/AlcoholandTobacco.html. adresinden ulaşılabilir. Bu veri kümesi, Büyük Biritanya'nın 11 bölgesinde haftalık ortalama alkol ve tütüne yapılan harcamaları gösterir (Alıştırma 9.27'de bu veriyi görmüştük). Veri kümesi Tablo 10.4'te verilmiştir, satırlar alkol tüketimindeki artışa göre düzenlenmiştir. Tek başına Alkol ya da Tütün harcamalarına bakıldığında, belirgin olmamasına rağmen, iki değişkenli grafikte Kuzey İrlanda'nın büyük uçdeğer olduğu açıktır. Benzer şekilde, alkol harcaması dağılımının normal olmadığı kesindir, sıralanmış alkol veri kümesi ise, diğer sıralar gibi, dikdörtgensel bir dağılım göstermektedir.

Alkole ilişkin sıralamalar sırayla listelendiğinde, tütüne ilişkin sıralarda ters sıralanmış çiftler bulunmaktadır (tütündeki 11. sıra tüm alt sıralardan önce gelirken, tütünde 10. sıra 3 alt sıradan önce gelmektedir). Tütün sütununda aşağı doğru giderek ters sıralanmış çift sayısı hesaplanabilir ve tablodaki aşağı doğru sıralama sayısı sayılarak kaç kez olduğu bulunabilir, böylece tablodaki sıralamada, yukarıdakinden daha düşük sıra sayısı bulunduğu saptanabilir. Örneğin, tütün tüketimine ilişkin 1. satır için 10 ters sıralama vardır; çünkü diğer 10 değer kendisinin altındadır. 2. satıra ait sadece 1 ters sıralama vardır; çünkü 2. sıradan küçük 1 değer vardır. Diğerleri de bu şekilde hesaplanır.[2]

Eğer iki sıra kümesi arasında mükemmel sıra ilişkisi varsa, herhangi bir ters sıralı çift bulunması beklenmez. Alkole en çok para harcayan bölge tütüne de en çok para harcayacaktır, ikinci sırada alkole para harcayan bölge, tütüne de ikinci sırada para harcayan bölge olacaktır ve bu şekilde devam edecektir. Bu formun tersleri Kendall istatistiğinin temelini oluşturur.

## $\tau$'nun Hesaplanması

Buradaki veri kümemizde $N(N\text{-}1)/2 = 11(10)/2 = 55$ sıra çifti vardır. Bunların 18'i terstir. Bütün ters sıralamaların sayısı Tablo 10.14'ün en sağdaki sütünların toplamı alınarak bulunmuştur ve bu çiftlerin 37'si ters değildir, bu basitçe toplam çift sayısından (55) ters çiftlerin (18) sayısı çıkarılarak bulunabilir.

2 Veriyi sıra sayılarına dönüştürmek kesinlikle gerekli değildir, sadece bir sıralı serinin her bir değerinden daha az olan ham puan gözlem sayısını hesaplayabiliriz. Ancak ilk başta sıralamalara dönüştürürsek hesaplama daha kolay olacaktır.

C uyumlu çiftlerin sayısını, D ters sıralı çiftlerin sayısını göstermek üzere,
D = Ters sıralı çiftler = 18 (uyumsuz)
C = 37

Kendall aşağıdaki tanımı vermiştir,

$$\tau = 1 - \frac{2(\text{ters sıralamaların sayısı})}{\text{tüm çiftlerin sayısı}} \quad \text{veya} \quad \frac{2D}{N(N-1)/2}$$

Bizim verimiz için,

$$\tau = 1 - \frac{2(\text{ters sıralamaların sayısı})}{\text{tüm çiftlerin sayısı}} = 1 - \frac{2(18)}{55} = ,345$$

Böylece, alkol ve tütün tüketimi sıralamaları arasındaki uyumun bir ölçüsü olarak Kendall'ın $\tau$'su = ,345 bulunur.

$\tau$'nun yorumlanması aynı veri üzerinden hesaplanan $r_S$'nin (0,37) yorumlanmasından daha basittir. Eğer $\tau$ = ,345 ise, nesne çiftleri tesadüfî örneklendiğinde, aynı sırada sıralanmış iki bölgenin olasılığı ,345, ters sıradaki sıralamaların olasılığından daha *yüksektir*.

Bunlar eşit sıralar bulunması halinde, $\tau$'nun hesaplanması değiştirilmelidir. Eşit sıraların uygun şekilde düzeltilmesi için için Hays (1981, s. 602 ) veya Google'a bakınız.

## $\tau$'nun Manidarlığı

Spearman'ın $r_S$'sinin aksine Kendall'ın $\tau$'sunun standart hatasının tahmini için kabul edilen bir yöntem vardır.

$$s_\tau = \sqrt{\frac{2(2N + 5)}{9N(N - 1)}}$$

Ayrıca *t*, $N \geq 10$ olduğunda yaklaşık olarak normal dağılır. Bu bize, Kendal'ın *t*'sunun örneklem dağılımı için normal dağılıma yaklaşımın kullanılabileceğini gösterir.

$$z = \frac{\tau}{s_\tau} = \frac{\tau}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}} = \frac{,345}{\sqrt{\frac{2(27)}{9(11)(10)}}} = \frac{,345}{,2335} = 1,48$$

Çift kuyruklu test için *p* = ,139 olduğundan manidar değildir.[3]

Kendall'ın $\tau$**'su**nun 0,2335 standart hata ile güven aralığı, $\tau$**'**nun normalliğini varsayarak, aşağıdaki gibidir:

$$GA = \tau \pm 1,96 s_\tau = \tau \pm 1,96\left(\sqrt{\frac{2(2N+5)}{9N(N-1)}}\right) = \tau \pm 1,96(,2335)$$

Örneğimiz için güven aralığı -,11 ≤ $\tau$ ≤ ,80 olarak elde edilir.

Kendall'ın $\tau$**'s**u genellikle Spearman'ın $r_S$'sine tercih edilir, çünkü evren parametresinin daha iyi bir kestiricisidir ve standart hatası bilinmektedir.

Ham puanlarda, verinin dağılımının normal olmaması durumunda ilgili evren parametresinin kestirilmesinde, Pearson'ın *r*'sinden daha iyi olduğuna dair kanıt vardır; bu sadece, oldukça aşırı düzeylerde doğru gibi görünüyor. Ham veri üzerinde ve kalıcı olarak, Pearson'ın

3 Küçük örneklemler için burada verinin bütün olası sıralamalarını kullanan daha kesin bir test vardır, bu nadiren gereklidir. (Ancak kitabın web sitesinde Kendall.R uygulanmaktadır.)

*r*'si genellikle tercih edilen katsayı olmuştur. (Bu veri kümesi için orjinal maliyet değerleri arasındaki Pearson korelasyonu $r = ,22$, $p = ,509$ olarak bulunmuştur.)

## 10.4 Sıralı Veriden Oluşan Çapraz Tabloların Analizi

Bölüm 6'daki ki-karede, bağımsız değişkenler, sıralı değişkenler olduğunda ortaya çıkan sorunlara değinildi. Geleneksel ki-kare analizi bu sıralamayı dikkate almaz ancak, doğru analiz için bu önemlidir. Bölüm 6'da belirttiğim gibi, bu bölüm, nezaketle kendi verisini bu örnek için kullanmama izin veren, Leicester Üniversitesi'ndeki Jennifer Mahon tarafından gönderilen bir soruyla tasarlanmıştır. Bayan Mahon, katılımcıların çocuklukta yaşadıkları travmatik olay sayısıyla ilgili yeme bozuklukları üzerine yaptığı çalışmasında, okulu bırakma olasılığı olup olmadığı sorusuyla ilgilenmiştir.

Bu çalışmadan elde edilen veri kümesi aşağıda gösterilmiştir. Veriyi hafifçe değiştirdim. Böylece küçük beklenen frekanslar sorunuyla uğraşmak zorunda kalmadım. Aynı zamanda verinin sıralı doğasının nasıl işlediğini göstermek için çalıştım. Değiştirilen veri hâlâ Mahon'ın bulduğu etkileri temsil etmektedir.

| | Travmatik Olay Sayısı | | | | | |
|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4+** | **Toplam** |
| **Okulu Bırakan** | 25 | 13 | 9 | 10 | 6 | 63 |
| **Kalan** | 31 | 21 | 6 | 2 | 3 | 63 |
| **Toplam** | 56 | 34 | 15 | 12 | 9 | 126 |

© Cengage Learning 2013

İlk bakışta bu veri kümesine standart ki-kare uygulamak cazip gelebilir. Hipotez testi, "okulu bırakma, çocukluk döneminde yaşanan travmatik olay sayısından bağımsızdır" şeklinde ifade edilebilir. Eğer bu yapılırsa, ki-kare 9,459 bulunur, 4 *sd*'de ilgili olasılık ,051'dir. Bu sonuç sıfır hipotezinin reddedilmesine izin vermez, buradan "travmatik olaylar okulu bırakma işlemiyle ilişkili değildir" sonucuna varabiliriz. Ancak, bu cevap fazlaca basitçedir.

Travmanın sıralı veriyi temsil ettiğine dikkat ediniz. 4 travmatik olay 3'ten, 3 travmatik olay 2'den vb. fazladır. Çocuklukta yaşadıkları travmatik olayların sayısının bir fonksiyonu gibi okulu bırakan katılımcıların yüzdesine baktığımızda, genel olarak monotonik olmasa da, okulu bırakma arttıkça travmatik olay sayısının da arttığı görülmektedir. Ancak bu durum, ki-kare değerinin hesaplanmasında herhangi bir rol oynamasına izin vermez. İstediğimiz bunu hesaba katan bir istatistiktir.

### Bir Korelasyonel Yaklaşım

İstediğimizi yapabilmenin birçok yolu vardır, ancak bunların hepsi bağımsız değişkenlerimize bir çeşit sıralı metrik değerler atamayla oluşur. Okulu bırakma ikili olduğundan bir problem oluşturmaz. Örneğin; okulu bırakanlar 1, kalanlar 2 veya okulu bırakanlar 1, kalanlar 0 vb. şeklinde puanlanabilir. Seçmiş olduğumuz değerler sonuçları etkilemeyecektir. Travmatik olayların sayısına geldiğimizde 0, 1, 2, 3 ve 4 'ü kullanabiliriz. Alternatif olarak 3 veya 4 travmatik olayın 1 veya 2 travmatik olaydan çok daha fazla önemli olaca-

ğını düşünüyorsak, 0, 1, 2, 4, 6 şeklinde de yazabiliriz. Pratikte, sayıları monotonik artan seçtiğimiz, çok büyük seçmediğimiz sürece sonuçlar bizim seçimlerimizin fonksiyonu gibi değişmeyecektir. Burada 0, 1, 2, 3 ve 4 kullandım.

Şimdi her bir değişken için bir metrik oluşturduk, burada gidebileceğimiz birçok farklı yol vardır. Sezgisel olarak ilgi çeken birisiyle başlayacağız. Sadece iki değişken ilişkili olacaktır.[4] Her katılımcının okulu bırakıp bırakmama için 0 ve 1 şeklinde, travma sayısı için de 0 ile 4 arasında puanı olacaktır. Bu iki ölçme arasındaki standart Pearson korelasyonu ,215 bulunmuştur, sıfır hipotezine göre ilgili olasılık ,016 olup hesaplanan korelasyon değeri manidardır ve bağımsızlık için sıfır hipotezini reddederiz.

"Travmatik olay sayısı" kesikli olduğundan, bazı insanlar bu durumunda Pearson'ın *r*'sinin kullanılması konusunda endişe duyabilirler. Aslında bu durum Pearson'ın *r*'si için bir problem değildir ve Agresti'den (2002) daha az bir otorite bile bu yaklaşımı önerir. Travmanın sıralı veri olduğunu kabul etmenize rağmen, bunun için belirli bir metrik belirleme fikri sizi muhtemelen mutsuz etmiştir. Eğer öyleyse Pearson'un *r*'si yerine Kendall'ın $\tau$'sunu hesaplayabilirsiniz. Değişkenin düzeylerinin artan olduğu varsayıldığında, $\tau$ travma düzeylerine atanan değerlerin herhangi kümesi için aynı olacaktır. Bu veri için $\tau$ = ,04 olasılıkla ,169 olacaktır. Bağımsız değişken(ler)in sıraları hakkında emin olsanız bile ilişki hâlâ manidar olacaktır. (Pearson'un *r*'si için olası yedek olarak Kendall'ın $\tau$'suna başvurmak, bu materyalin Bölüm 9 yerine burada kullanılmasının nedenidir.) Bununla beraber Agresti, hücre frekansları çok farklı ise, ya Kendall'ın Tau'sunu ya da Spearman'ın $r_S$'sini kullanarak olumsuz sonuçların olacağına işaret etmiştir. (Ben sadece *r* kullanmanızı şiddetle tavsiye ederim.)

Agresti (2002, s. 87), bizim korelasyondan ki-kare istatistiğini hesaplayabileceğimiz bir yaklaşımı göstermiştir ve aşağıdaki eşitliği vermiştir:

$$M^2 = (N - 1)r^2$$

Burada $M^2$, 1 serbestlik derecesindeki ki-kare istatistiğini, *r* Okulu Bırakma ve Travma arasındaki Pearson korelasyonunu, *N* ise örneklem büyüklüğünü gösterir. Örneğimiz için bu değer

$$M^2 = \chi^2(1) = (N-1)r^2$$
$$\chi^2(1) = 125(0,215^2) = 5,757$$

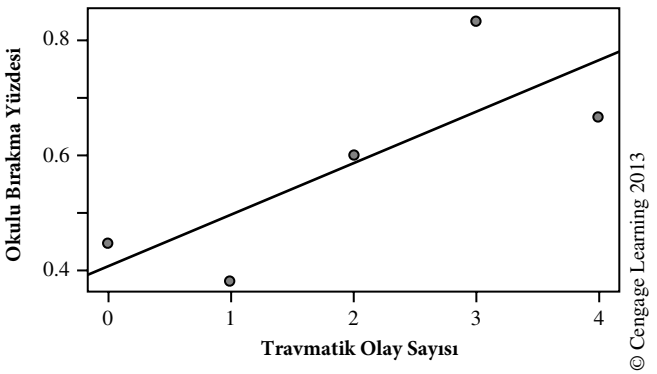
olup sıfır hipotezi altındaki ,016 olasılığına karşı gelir.

Olasılık değeri korelasyon testi tarafından verilmişti, dolayısıyla yeni değildir. Ancak bir adım daha ileri gidilebilir. Başlangıçta hesaplanan 4 *sd*'deki genel Pearson ki-kare değerinin 9,459 olduğunu biliyoruz. Biraz önce sadece iki değişken arasındaki doğrusal ilişkiye ait 1 *sd*'deki ki-karenin 5,757 olarak hesaplanmış olduğunu öğrendik. Bu doğrusal ilişki toplam ki-karenin bir parçasıdır ve elde edilen genel ki-kareden doğrusal bileşenleri çıkarmak istersek aşağıdaki tabloda yer alan değerleri elde ederiz.

| | *df* | Chi-square |
|---|---|---|
| Pearson | 4 | 9,459 |
| Doğrusal | 1 | 5,757 |
| Doğrusallıktan sapma | 3 | 3,702 |

© Cengage Learning 2013

4 Literatürde birçok makale sıralı veri kümeleriyle başa çıkmak için bir kaynak olarak Maxwell'i (1961) gösterir. Maxwell'in yaklaşımı burada savunulan bir yaklaşımdır, açıklamalarından bahsetmek zor da olsa, onun formülleri hesaplama kolaylılığı için seçildi.




**Şekil 10.2** Mahon'ın okulu bırakma verisine ait dağılımın grafiği

Sıfr hipotezi altında ,295 olasılıkla doğrusallıktan ayrılma ile ilgili ki-kare değeri 3 serbestlik derecesinde 3,702'dir. Bu veri altında yatan doğrusal eğilim dışında herhangi bir kanıt yoktur. Şekil 10.2'de görüldüğü gibi okulu bırakma ve travma arasındaki ilişki temelde doğrusaldır.

Agresti'de (2002, 2007), burada ele alınan yaklaşımın kusursuz bir tartışması vardır. Agresti, küçükten orta örneklem büyüklüklerine doğru ilgi çekici noktalara değinmiştir, standart Pearson ki-kare, küçük örneklemin olumsuz etkilerine karşı hesapladığımız, verinin tamamına ait ki-kareden daha duyarlıdır. Diğer bir deyişle, çapraz tablodaki bazı hücrelerin küçük olmasına rağmen, kendimi veri kümesinin tamamına ait ordinal (linear) 5,757'ye eşit ki-kare değeriyle 9,459'ye eşit Pearson ki-kare değerinden daha güvende buluyorum.

Doğrusallık için SPSS kullanarak ki-kare hesaplanabilir. Eğer istatistik iletişim kutusundan ki-kare istatistiğini seçerseniz, sonuç penceresinde Pearson ki-kare, ki-kare olabilirlik oranı (ki-kare Likelihood ratio chi-square), doğrusal-doğrusal ilişki (Linear-by-Linear Association) görünür. Mahon'ın verisi için SPSS çıktısı aşağıda gösterilmiştir. Tablodan, doğrusal-doğrusal ilişki ölçüsünün 5,757, $(N-1)r^2$ kullanarak hesaplanan $\chi^2$ değeriyle aynı olduğu görülmektedir.

| | Ki-Kare Testleri | | |
|---|---|---|---|
| | **Değer** | ***sd*** | **Asimp. Man. (2 yönlü)** |
| Pearson Ki-Kare | 9,459[a] | 4 | ,051 |
| Olabilirlik Oranı | 9,990 | 4 | ,041 |
| Doğrusal-Doğrusal | 5,757 | 1 | ,016 |
| Geçerli N Sayısı | 126 | | |

[a] 2 hücreye ait (%20) 5'ten daha az beklenen değere sahiptir. Minimum beklenen değer 4,50'dir.

Çapraz tablodaki sıralı değişkenler sorunu için başka yollar da vardır. Bazı durumlarda değişkenlerden biri sıralı, diğeri sınıflamalı olabilir. (İkili puanlanan değişkenlerle, analizi etkilemeden sıralı gibi işlem yapılabileceğini hatırlayınız.) Bazı durumlarda değişkenlerden biri bağımsız, diğeri bağımlı olabilmektedir. Bu metotların bazılarına ilişkin mükemmel tartışmalar Agresti, 2002 ve 2007'de bulunur.

## 10.5 Kendall'ın Konkordans Katsayısı (*W*)

**Kendall'ın Konkordans Katsayısı (*W*)**

Bu bölümde ilgili olan istatistiklerin hepsinde iki puan kümesi ($X$ ve $Y$) arasındaki ilişki ele alınmıştır. Ancak obje kümelerini iki hakemin sıralaması yerine, altı hakemin sıraladığını düşünelim. Altı hakemin uyuşma derecelerinin ölçüsüne ihtiyaç duyulur. Böyle bir ölçü **Kandall'ın Konkordans Katsayısı**yla (***W***) elde edilir.

Örnek olarak altı hakemden sekiz renkli kumaşı beğenme durumlarına göre sıralamaları istenmiş ve Tablo 10.5'teki veri kümesi elde edilmiş olsun. Eğer hakemlerin tümü en fazla B kumaşının renginden memnun kalırlarsa, hepsi buna 1. sırayı verecektir. Altı hakem karşısında o kumaş için sütun toplamı 6 olacaktır. Benzer şekilde A kumaşının rengi hepsi için 2. sırada olursa, bunun sütun toplamı 12 olacaktır. Son olarak tüm hakemler H kumaşına en yüksek sırayı verirlerse, H kumaşının sütun toplamı 48 olacaktır. Başka bir ifadeyle, sütun toplamları önemli değişkenlik gösterir.

Diğer yandan hakemler hiç uyum göstermezlerse her bir sütunda bazı yüksek sıralar ve düşük sıralar yer alacaktır ve sütun toplamları kabaca eşit olacaktır. Sütun toplamlarındaki değişkenlik hakemler arasındaki uyumsuzluğu (veya tesadüfî tepki) verecektir ve düşük olacaktır.

Kendall, sütun toplamlarındaki değişkenliği kullanarak kendi istatistiğini tanımlamıştır. Sütunlar arasındaki varyansı, sütun toplamlarının varyansının muhtemel maksimum varyansa oranı olarak almış ve bunu $W$ ile göstermiştir.

$$\mathrm{W} = \frac{\text{Sütun toplamlarının varyansı}}{\text{Sütun toplamlarının muhtemel maksimum varyansı}}$$

Sıralarla ilgilendiğimizden toplam maksimum varyansın ne olacağını biliyoruz. Biraz cebir ile, aşağıdaki tanımlamayı yapabiliriz:

$$W = \frac{12\sum T_j^2}{k^2N(N^2 - 1)} - \frac{3(N + 1)}{N - 1}$$

Bu eşitlikte $T_j$, sütun toplamları, $N$ = sıralanacak obje sayısı ve $k$ = sıralama yapan hakemlerin sayısıdır. Tablo 10.5'teki veri için

$$\sum T_j^2 = 11^2 + 9^2 + 18^2 + 30^2 + 28^2 + 36^2 + 45^2 + 39^2 = 7052$$

$$W = \frac{12\sum T_j^2}{k^2N(N^2 - 1)} - \frac{3(N + 1)}{N - 1}$$

$$= \frac{12(7052)}{6^2(8)(63)} - \frac{3(9)}{7} = \frac{84624}{18144} - \frac{27}{7}$$

$$= ,807$$

*W*'nun tanımından görebileceğiniz gibi, bu bir standart korelasyon katsayısı değildir. Benzer istatistikler cinsinden bir yorumu vardır. Bununla birlikte tüm muhtemel ikili yargılardan hesaplanan ortalama Spearman korelasyonunun fonksiyonu gibi görülebilir. Özellikle

$$\bar{r}_S = \frac{kW - 1}{k - 1}$$

**Tablo 10.5** Hakemlerin renkli kumaş parçalarını beğenme düzeylerine göre sıralaması

| | Renkli Kumaşlar | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Hakemler | A | B | C | D | E | F | G | H |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 2 | 1 | 5 | 4 | 3 | 8 | 7 | 6 |
| 3 | 1 | 3 | 2 | 7 | 5 | 6 | 8 | 4 |
| 4 | 2 | 1 | 3 | 5 | 4 | 7 | 8 | 6 |
| 5 | 3 | 1 | 2 | 4 | 6 | 5 | 7 | 8 |
| 6 | 2 | 1 | 3 | 6 | 5 | 4 | 8 | 7 |
| Σ | 11 | 9 | 18 | 30 | 28 | 36 | 45 | 39 |

© Cengage Learning 2013

Verimiz için,

$$\bar{r}_s = \frac{kW - 1}{k - 1} = \frac{6(,807) - 1}{5} = ,768$$

Böylece, eğer bütün olası ikili sıralamaları alıp her biri için $r_s$ hesaplansaydı, ortalama $r_s$ değeri ,768 olacaktı.

Hays (1981), *W*'nun raporlanmasını önerir, ancak yorumlanması için *W*'nun $\bar{r}_s$'ya dönüştürülmesini tavsiye eder. Bu tavsiyeye karşı çıkmak zordur çünkü, *W*'ya atfedilebilecek bir sezgisel anlam yoktur. *W*'nun 0 ile 1 aralığında olması avantajı olmasına karşın $\bar{r}_s$'da bu yoktur, fakat sütun toplamlarının varyansının toplam maksimum muhtemel varyansa oranın %80,7 olmasına çok fazla pratik anlam yüklemek zordur. $\bar{r}_s$'nın hatası ne olursa olsun tercih edilebilir görünüyor.

"Hakemler arasında herhangi bir uyum yoktur." sıfır hipotezinin testi belirli koşullar altında mümkündür. Eğer k ≥ 7 ise, miktar,

$$\chi^2_{(N-1)} = k(N - 1)W$$

$N - 1$ serbestlik derecesinde $\chi^2$ dağılımına yaklaşır. Böyle bir test nadiren kullanılır, ancak, uyum düzeyinin araştırıldığı ve manidarlık için belli bir minimum düzeyin istendiği araştırmalarda ve nadiren hakem sayısının yedi veya daha çok olduğu durumlarda *W*'yu genellikle hesaplarız.

## Anahtar Terimler

Korelasyonel ölçmeler (Giriş)
İlişkilerin ölçümleri (Giriş)
Geçerlik (Giriş)
İkili Puanlama (10.1)
Nokta çift serili korelasyon ($r_{nç}$) (10.1)
$\phi$ (fi) katsayısı (10.1)
Çift serili korelasyon katsayısı ($r_ç$) (10.2)
Tetrakorik korelasyon ($r_t$) (10.2)
Sıralı veri için Spearman korelasyon katsayısı ($r_S$)
Spearman ro (10.3)
Kendall'ın ro katsayısı (10.3)
Kendall'ın konkordans katsayısı (*W*) (10.5)

## Alıştırmalar

10.1 Bazı insanlar en iyi işlerini sabah yaptıklarını düşünürken bazıları da gece yaptıklarını iddia ederler. 20 tane ofis çalışanı, sabah ya da gece olarak ikiye ayrılmış (sabah = 0, gece = 1) ve bazı belirlenmiş sabahlarda üretilen işin kalitesini belirlemişlerdir. 100 puanlık bir ölçek üzerinden yapılan sıralamalar aşağıdaki gibidir:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Günün en yoğun zamanı:** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Performans değerlendirmesi:** | 65 | 80 | 55 | 60 | 55 | 70 | 60 | 70 | 55 | 70 |
| **Günün en yoğun zamanı:** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Performans değerlendirmesi:** | 40 | 70 | 50 | 40 | 60 | 50 | 40 | 50 | 40 | 60 |

a. Bu verinin grafiğini ve uyan regresyon doğrusunu çiziniz.
b. $r_{nç}$'yi hesaplayınız ve manidarlığını test ediniz.
c. Sonuçları yorumlayınız.

10.2 İş programlarındaki şanslı değişiklik nedeniyle, akşamki aynı görevler üzerindeki performans için Alıştırma 10.1'de belirtilen konular yeniden değerlendirilmiş. Buna ilişkin veri aşağıda belirtilmiştir.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Günün en yoğun zamanı:** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Performans değerlendirmesi:** | 40 | 60 | 40 | 50 | 30 | 40 | 50 | 50 | 20 | 20 |
| **Günün en yoğun zamanı:** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Performans değerlendirmesi:** | 40 | 50 | 30 | 30 | 50 | 50 | 40 | 50 | 40 | 60 |

a. Bu verinin grafiğini ve uyan regresyon doğrusunu çiziniz.
b. $r_{nç}$'yi hesaplayınız ve manidarlığını test ediniz.
c. Sonuçları yorumlayınız.

10.3 Alıştırma 10.1 ve 10.2'de elde ettiklerinizi karşılaştırınız. Ne karar verebilirsiniz?

10.4 Alıştırma 10.1 ve 10.2'deki veri kümesi üzerinde çift serili korelasyonu hesaplamak neden anlamlı olmaz?

10.5 Alıştırma 10.1'deki veriye $t$ testini uygulayınız ve $t$ ile $r_{nç}$ değerleri arasındaki ilişkiyi gösteriniz.

10.6 Bir lisansüstü okula kabul komitesi, başvuranların üniversite not ortalamaları ve bireylerin eğitim sonunda doktora derecesi için gereklilikleri tamamlayıp tamamlayamamaları arasındaki ilişki hakkında endişe duymaktadır. Onlar ilk başta 7 yıl önce programa kaydolan 25 öğrenciyi tesadüfî olarak seçerek, doktora programını tamamlayanlara 1, tamamlamayanlara 0 puanını vermişler ve elde ettikleri veriyi incelemişlerdir. Buna ilişkin veri kümesi aşağıda verilmiştir.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Not Ortalaması:** | 2,0 | 3,5 | 2,75 | 3,0 | 3,5 | 2,75 | 2,0 | 2,5 | 3,0 | 2,5 | 3,5 | 3,25 | 3,0 |
| **Doktora Derecesi:** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| **Not Ortalaması:** | 3,0 | 2,75 | 3,25 | 3,0 | 3,33 | 2,5 | 2,75 | 2,0 | 4,0 | 3,0 | 3,25 | 2,5 | |
| **Doktora Derecesi:** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |

a. Bu veri kümesinin grafiğini çiziniz.
b. $r_{nç}$'yi hesaplayınız.
c. $r_{ç}$'yi hesaplayınız.
d. Bu durumda $r_{ç.}$'ye bakmak mantıklı mıdır? Neden mantıklıdır veya neden mantıklı değildir?

10.7 10.6'daki veri için regresyon eşitliğini bulunuz. Bu eşitlikle tanımlanan doğrunun iki grubun ortalamalarından geçtiğini ispatlayınız.

10.8 Alıştırma 10.7'de bulduğunuz sabit ve eğim neyi belirtir?

10.9 Alıştırma 10.6'da yer alan komitenin not ortalamaları için 3,00 kesme puanının uygun olacağına karar verdiğini varsayalım. Başka bir ifadeyle, not ortalaması 3,00 veya üstünde olanlar kabul edilebilir ve 3,00'ın altında olanlar da kabul edilemez şeklinde sınıflandırıldığını düşünelim. Daha sonra komite, bunu doktora programını tamamlamayla ilişkilendirsin. Bu durumda,

a. Alıştırma 10.6'daki veriyi buna göre puanlayınız.

b. Korelasyonu hesaplayınız.

c. Korelasyon katsayısının manidarlığını test ediniz.

10.10 Alıştırma 10.9'daki veri kümesini çapraz tablo yardımıyla görselleştiriniz.

a. Bu tablo yardımıyla ki-kareyi hesaplayınız.

b. Ki-kare ve $\phi$ arasındaki ilişkiyi gösteriniz.

10.11 Bir araştırmacı, alkolizm ve çocukluk aşamalarındaki dikkat eksikliği bozukluğu (DEB) arasındaki ilişkiyle ilgilenmektedir. Aşağıda görülen veri kümesini toplamıştır. Bu kümede 1, ilgili sorunun varlığını göstermektedir.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DEB:** | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| **Alkolizm:** | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| **DEB:** | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| **Alkolizm:** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |

a. Bu iki değişken arasındaki korelasyon kaçtır?

b. Bu ilişki manidar mıdır?

10.12 Bir araştırmacı, dil becerilerinin gelişiminde görülen dil bozukluklarının sırası üzerine temellendirilmiş 15 maddeden oluşan bir ölçek düzenlemek istiyor. Becerilerin sıralamasının doğru olduğundan tamamen emin olamadığından, bir başka profesyonele bu 15 maddenin nasıl sıralanması gerektiğini soruyor. Bunlara ilişkin veri kümesi aşağıda verilmiştir.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Araştırmacı:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **Danışman:** | 1 | 3 | 2 | 4 | 7 | 5 | 6 | 8 | 10 | 9 | 11 | 12 | 15 | 13 | 14 |

a. Pearson'ın formülünü ($r$) kullanarak Spearman'ın $r_S$'ni hesaplayınız.

b. Sonuçların size söylediğine göre sıralama sürecini tartışınız.

10.13 Alıştırma 10.12'deki veri için,

a. Kendall'ın $\tau$'sunu hesaplayınız.

b. $\tau$'nun manidarlığını test ediniz.

10.14 Tanı süreçlerini içeren bir çalışmada, klinik yüksek lisansa giren öğrencilere çocuk davranışlarını içeren 20 dakikalık bir video kaset izlettirilmiştir ve kasette yer alan 10 davranışsal olayı önem sırasına göre (1 = en önemli) değerlendirmeleri istenmiştir. Daha sonra tüm sınıfa ait sıra düzenini sağlamak amacıyla yeni öğrencilerin ve deneyimli klinik uzmanlarının sıralamalarının ortalamaları alınmıştır. Aşağıdaki veri kümesi elde edilmiştir.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Olaylar:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Deneyimli klinisyenler:** | 1 | 3 | 2 | 7 | 5 | 4 | 8 | 6 | 9 | 10 |
| **Yeni öğrenciler:** | 2 | 4 | 1 | 6 | 5 | 3 | 10 | 8 | 7 | 9 |

Deneyimli kilnikçiler ile yeni öğrenciler arasındaki uyumu ölçmek için Spearman'ın $r_S$ katsayısını hesaplayınız.

10.15 Alıştırma 10.14'teki veriyi kullanarak Kendall'ın $\tau$ 'sunu hesaplayınız.

10.16 Alıştırma 10.14'te beş tane klinik öğrencinin olduğunu varsayalım. Onlardan elde edilen veri aşağıdaki gibidir:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Öğrenci 1:** | 1 | 4 | 2 | 6 | 5 | 3 | 9 | 10 | 7 | 8 |
| **Öğrenci 2:** | 4 | 3 | 2 | 5 | 7 | 1 | 10 | 8 | 6 | 9 |
| **Öğrenci 3:** | 1 | 5 | 2 | 6 | 4 | 3 | 8 | 10 | 7 | 9 |
| **Öğrenci 4:** | 2 | 5 | 1 | 7 | 4 | 3 | 10 | 8 | 6 | 9 |
| **Öğrenci 5:** | 2 | 5 | 1 | 4 | 6 | 3 | 9 | 7 | 8 | 10 |

Uyumun ölçümü olarak bu veri için Kendal'ın $W$ ve $\overline{r}_s$ katsayılarını hesaplayınız. Sonuçları yorumlayınız.

10.17 Sayfa 312'de, Rosenthal ve Rubin'in $r^2$'nin, 1024'e eşit değeriyle oldukça önemli etkisi olduğunu gösterdiğinden söz etmiştim. Rosenthal ve Rubin bunu, iki grubun her birinde 100'er bireyin olduğu 2 x 2'lik çapraz tabloda bir satırı için 34:66'ya ve diğer satırı için 66:34'e bölünerek elde edilen $\chi^2$ = 20,48'e ilişkilendirerek göstermişlerdir.

a. Kendi 2 x 2'lik tablonuz ile bu hesabı doğrulayınız.

b. Her bir grupta 100 birey olmak üzere $r^2$ değeri ,0512 olsaydı 2 x 2'lik tablo neye benzerdi? (2 x 2'lik tablolar üretilmesinde ve her biri için $\chi^2$ hesaplamasında bazı deneme yanılmaları gerektirebilir.)

10.18 Bu kitabın Web sitesindeki Mireault'un verisini (Mireault.dat) kullanarak Cinsiyet ve Depreseyon T puanları arasında nokta-çift serili korelasyonu hesaplayınız. Bu sorunun ilgili yönlerini Alıştırma 7.46'dan elde edilen sonuçlarla karşılaştırınız. (Kesim 10.1'deki "$r_{nç}$ ve $t$ arasındaki ilişki"ye bakınız.)

10.19 Alıştırma 7.48'deki Mireault'un verisini kullanlarak, ebeveynlerini kaybetmiş öğrencilerin tepkileri ile ebeveynlerini kaybetmemiş öğrencilerin tepkileri Global Belirti İndeksi T Puanları (GSIT) açısından diğer değişkenler arasında karşılaştırılmıştır. Alternatif bir analizde, GSIT üzerinden klinik açıdan manidar bir kesme puanı kullanarak o puanın üzerinde olanlar klinik vak'a (klinik olarak manidar semptom düzeyini gösteren) veya o puanın altında olanlar klinik vak'a olmayan şeklinde sınıflandırılacaktır. Derogatis (1983), bu kesme puanını 63 olarak önermiştir. (Örneğin, GSIT > 63 olursa, Klinik Vak'a (Clinase) = 1, değilse Klinik Vaka (ClinCase) = 0 şeklinde olacaktır.)

a. Derogatis tarafından tanımlandığı gibi, herhangi bir istatistiksel paket kullanarak Degoratis'in yaptığı gibi Clincase değişkenlerini oluşturunuz. Gruplara karşı Klinik Vak'a (Clincase) durumunun çapraz tablosunu oluşturunuz. Ki-kare ve Cramer'in $\phi_C$ değerlerini hesaplayınız.

b. Yukarıdaki (a) seçeneğinde elde edilen cevaplarla Bölüm 7'den elde edilen cevapları karşılaştırınız.

c. Neden Bölüm 7'de benimsenen yöntem üzerindeki bu yaklaşımı (vak'aya karşı vak'a olmayana bakınız) tercih edebilirsiniz?

(İpucu: SAS Ariel gerektirir. Ve SPSS Çapraz Tablo kullanacaktır. Uygun el kitapları (manueller) komutları ayarlamanıza yardımcı olacaktır.)

10.20 Çapraz Tabloda Cinsiyete Karşı Klinik Vaka'yı kullanarak, Alıştırma 10.19'daki analizi tekrar ediniz.

a. Cevaplarınızı Alıştırma 10.18'deki sonuçlarla karşılaştırınız.

b. Bu analiz, Alıştırma 10.18'dekinden nasıl farklıdır? Kabaca aynı soru mudur?

## Tartışma Soruları

10.21 Rosenthal ve arkadaşları (daha önce sözü edilen), küçük $r^2$ ile indekslendiği gibi küçük etkileri savunmuştu. Örneğin, bunun bazı durumlarda önemli olabileceğini belirtmişlerdi. Diğer durumlarda küçük etkilerin önemsiz olabileceğine hepimiz büyük olasılıkla katılırız.

a. Eğer yeterince büyük bir $r^2$ varsa, manidar olmayan bir etki daima önemli olabilir mi?

b. Bu etkinin önemi sorusuna, örneklem büyüklüğü katkısı nasıl olacaktır?



Bölüm 11

# Basit Varyans Analizi

## Hedefler

İki veya daha fazla ortalama arasındaki farkı test etmek için yöntem olarak varyans analizine giriş.

## Konular


**11.1** Bir Örnek
**11.2** Altta Yatan Model
**11.3** Varyans Analizinin Mantığı
**11.4** Varyans Analizinde Hesaplamalar
**11.5** Sonuçların Raporlanması
**11.6** Bilgisayar Çözümleri
**11.7** Eşit Olmayan Örneklem Büyüklükleri
**11.8** Varsayımlara Uyulmaması
**11.9** Dönüştürmeler
**11.10** Sabite Karşı Tesadüfî Modeller
**11.11** Deneysel Etkinin Büyüklüğü
**11.12** Güç
**11.13** Bilgisayar Analizleri

varyans analizi (ANOVA)

**VARYANS ANALİZİ (ANOVA)** psikolojik araştırmalarda en çok kullanılan istatistiksel teknik olma statüsüne sahiptir. Bu tekniğin popüler ve kullanışlı olması iki kaynağa atfedilebilir. Birincisi, varyans analizi, *t* gibi, örneklem ortalamaları arasındaki fark ile ilgilenir; *t*'den farklı olarak, ortalama sayıları ile ilgili bir kısıtlama zorunluluğu yoktur. İki ortalama arasındaki farkın olup olmadığını sormak yerine üç, dört, beş veya *k* ortalamanın farklı olup olmadığını sorabiliriz. Varyans analizi ayrıca, sadece her değişkenin ayrı ayrı temel etkilerini değil aynı zamanda iki veya daha fazla değişkenin ortak etkilerini sorgulayarak iki veya daha fazla bağımsız değişkenle aynı anda ilgilenmemize izin verir.

tek-yönlü varyans analizi

Bu bölüm, varyans analizinin altında yatan mantık ve sadece tek bağımsız değişkene dayalı deney sonuçlarının analizi ile ilgilidir. Ayrıca **tek-yönlü** (tek bağımsız değişkenli) **varyans analizi**nin içeriğindeki en kolay anlaşılan ilgili konuları inceleyeceğiz. Daha sonraki bölümlerde iki veya daha fazla bağımsız değişken içeren deney analizlerindeki tek örneklem ortalamaları ve her bireyde tekrarlı ölçümler yapılan desenlere sahip tek örneklem ortalamalarının karşılaştırılması üzerinde duracağız.

## 11.1 Bir Örnek

Varyans analizinin birçok özelliği en iyi şekilde basit bir örnekle açıklanabilir, bu nedenle Giancola ve Corman (2007) tarafından yapılan bir çalışma ile başlayacağız. Giancola ve Corman önemli ölçüde alkol tüketen bireylerin agresif davranışları üzerinde dikkat dağıtıcı görevin etkileriyle ilgilendiler. Alkolün saldırgan davranışlara yol açtığı iyi bilenmekteydi peki ama neden? Giancola ve Corman, daha az göze batan yasaklayıcılardan ziyade daha çok kışkırtıcı ipuçlarına odaklanarak alkolün saldırganlığı arttırdığı sayıltısı ile çalışmaya başladılar. Eğer deneklere dikkat dağıtıcı bir görev verirlerse, dikkatlerinin kışkırtıcı ipuçları yerine bu görevlere odaklanacakları ve saldırganlıklarının kısıtlanacağı şeklinde akıl yürüttüler. Ayrıca, eğer verilen görev çok karmaşık gelirse, ayırt edici etkilerin kaybolacağı ve saldırganlığın tekrar ortaya çıkacağını düşündüler. (Aslında eğer verilen görev zorsa, bu durum karmaşıklığa ve hayal kırıklığına yol açacak ve bu durum saldırganlığa dönüşecektir.)

Giancola ve Corman deneklerinden, kanlarındaki ortalama alkol oranını %10 düzeyine çıkaracak miktarda alkol tüketmelerini istedi. (Bu düzey hepsinde olmasa da birçok ülkede alkollü sürücü olarak nitelendirilir.) Sonra denekler kendilerine gösterilen, 3 x 3'lük aydınlatılmış bir matrisi sırasıyla ezberlemeleri gereken bir göreve katıldılar. Deneklerin hafızada tutmak zorunda oldukları kareler değiştirilerek görevin dikkat gereklilikleri çeşitlendirildi. Denekler, verilen görevdeki performanslarına bağlı olarak hem deneklere hafif düzeyde şok veren hem de denekler tarafından şok verilen hayalî rakiplere karşı oynadılar. Bağımlı değişken (saldırganlık), deneklerin fırsatını bulduklarında rakiplerine verdikleri şokun süresine ve şiddetine bağlıydı.

Çalışmada, verilen görevin güçlüğüne göre beş farklı grup vardı. Denekler hem 0, 2, 4, 6 ve 8 gibi kareleri hem de D0, D2, D4, D6 ve D8 gibi atandıkları grup desenlerini ezberlemek zorundaydı. Her grupta 12 denek vardı, ortalama ve standart sapmaların yer aldığı veri kümesi Tablo 11.1'de verilmiştir. (Veri gerçek çalışmada belirtilen ortalamalar ve standart sapmalara dayalıdır.)

**Tablo 11.1** Görev güçlüğünün bir fonksiyonu olarak uygulanan şok düzeyi

| | D0 | D2 | D4 | D6 | D8 | Toplam |
|---|---|---|---|---|---|---|
| | 1,28 | – 1,18 | – 0,41 | – 0,85 | 2,01 | |
| | 1,35 | 0,15 | – 1,25 | 0,14 | 0,40 | |
| | 3,31 | 1,36 | – 1,33 | – 1,38 | 2,34 | |
| | 3,06 | 2,61 | – 0,47 | 1,28 | – 1,80 | |
| | 2,59 | 0,66 | – 0,60 | 1,85 | 5,00 | |
| | 3,25 | 1,32 | – 1,72 | – 0,59 | 2,27 | |
| | 2,98 | 0,73 | – 1,74 | – 1,30 | 6,47 | |
| | 1,53 | – 1,06 | – 0,77 | 0,38 | 2,94 | |
| | – 2,68 | 0,24 | – 0,41 | – 0,35 | 0,47 | |
| | 2,64 | 0,27 | – 1,20 | 2,29 | 3,22 | |
| | 1,26 | 0,72 | – 0,31 | – 0,25 | 0,01 | |
| | 1,06 | 2,28 | – 0,74 | 0,51 | – 0,66 | |
| Ortalama | 1,802 | 0,675 | – 0,912 | 0,544 | 1,889 | 0,800 |
| Std. Sapma | 1,656 | 1,140 | 0,515 | 1,180 | 2,370 | 1,800 |
| Varyans | 2,741 | 1,299 | 0,265 | 1,394 | 5,616 | 3,168 |

© Cengage Learning 2013

## 11.2 Altta Yatan Model

Varyans analizi diğer tüm istatistiksel işlemler gibi bir modele göre inşa edilmiştir. Model üzerinde çok uzun durmayacağım ve yansımalarının hepsini tartışmayacağım ancak modeli genel olarak anlamak, varyans analizinin ne olduğunu ve daha sonraki ünitelerde yer alan daha karmaşık modelleri anlamada önemlidir. Yirmi beş yıl önce varyans analizinin altında yatan modeli konuşamayabilirdik. Fakat çeşitli işlemlerinin tartışıldığı doğrusal modelle tanışacaksınız ve bu örnek basit tek yönlü varyans analizi modeline bir giriş sağlayacaktır.

Giancola ve Corman'ın deneyindeki bir bireyin puanını tahmin etmek zorunda olduğunuzu düşünelim. Birinci açık tahmin evrendeki genel ortalama ($\mu$) olurdu. Bu tam tahmin olmayacaktır ancak diğer yapabileceklerinizden daha iyidir. Bu modeli (çok tatmin edici olmasa da) şu şekilde yazabiliriz.

$$X_{ij} = \mu$$

Fakat bireyin birinci grupta olduğunu söylediğimi farz edelim. Grup *j*'de olma etkisini $\tau_j$ olarak gösterelim. Burada $\tau_j$ basitçe grup *j*'nin ortalaması ile genel ortalama arasındaki farkı gösterir (örneğin, $\mu_j - \mu$). Bu durumda modelimiz

$$X_{ij} = \mu + (\mu_j - \mu)$$
$$X_{ij} = \mu + \tau_j$$

olacaktır.

Bu yaptığımız daha iyi bir tahmindir, çünkü bu bireyin puanını, grup ortalamasının genel ortalamadan ne kadar farklılaşacağını, genel ortalama artı veya eksi şeklinde kestirileceğini söyler. Bu durum sizin grup ortalamasını kestirdiğiniz anlamına gelse de, $\text{Grup}_1$'in etkisini ve genel ortalamayı ayırırsak daha iyi olacaktır.

Ancak bu sonuç cevap olmayacaktır. Eğer $Grup_1$'deki bir bireyseniz, muhtemelen grup ortalamasından bir şekilde farklılaşacaksınızdır. Diğer bir ifadeyle, sizin için size özel bir bileşen olacaktır ve bu bileşen $\varepsilon_{ij}$ şeklinde gösterilir ve $\varepsilon$ hatayı temsil eder. Bu durumda modelimiz aşağıdaki gibi olur.

$$X_{ij} = \mu + (\mu_j - \mu) + (X_{ij} - \mu_j)$$
$$X_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

**yapısal model**

Bu, varyans analizinin altında yatan **yapısal model**dir. $\mu$ genel ortalamayı, $\tau_j$ grup $j$'nin özel etkisini ve $\varepsilon$ bireysel hatayı, başka bir ifadeyle bireyin grup ortalamasından ne kadar saptığını gösterir. İlerleyen bölümlerde, temel fikir aynı kalacak şekilde $\tau_j$'den başka etkileri ekleyerek modeli daha karmaşık durumlara genişleteceğiz. Elbette yapısal modelde yer alan çeşitli parametre değerlerini bilmiyoruz ancak bu durum böyle bir model varsaymamızı engellemez.

## Sayıltılar

Bildiğimiz gibi, Giancola ve Corman 5 farklı dikkat dağıtıcı düzeyi altındaki saldırganlık düzeyi hakkında çalışıyorlardı. Bu koşulları, her bir koşul altında elde edilen evren puanlarının varyansını ve ortalamasını Şekil 11.1'deki gibi gösterebiliriz. Varyans analizi bu evrenler ve onların parametreleri hakkındaki temel sayıltılara dayalıdır. (Bu şekilde, bir dağılım diğerinin sağındadır gerçeği onun ortalamasının diğerlerinden farklı olup olmadığı hakkında herhangi bir şey söylemez.)

## Varyansın Homojenliği

Varyans analizinin altında yatan temel sayıltı, her evrenin aynı varyansa sahip olduğudur. Başka bir ifadeyle,

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_\varepsilon^2$$

$\sigma_\varepsilon^2$ veya sıklıkla $\sigma_e^2$ sembolleri, beş evren varyansı tarafından tutulan ortak değeri göstermede kullanılır. Bu durum **varyansın homojenliği** sayıltısı olarak olarak adlandırılır.

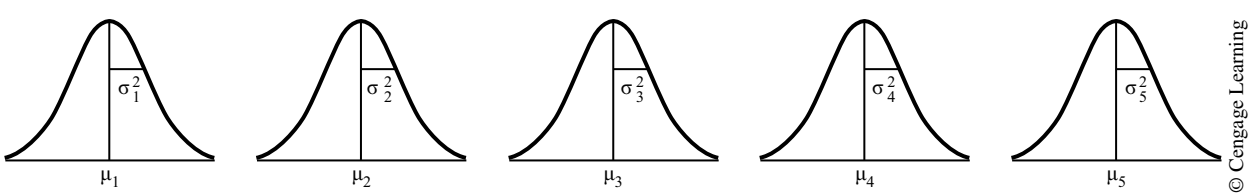
**varyansın homojenliği**

**hata varyansı**

"*e*" alt simgesi, hatayı temsil eder ve bu varyans **hata varyansı** -aynı koşullarda puan değişkenliğini gösteren herhangi bir grup farklılığıyla ilgisi olmayan varyans- olarak ifade edilir. Grup etkisinde her bireyin puanına bir sabit eklenirse, örneğin, sekiz uyarıcının pozisyonunu hatırlamaya çalışan her birey, saldırganlık ölçeği üzerinde diğerlerinden yüksek olanlardan fazladan herhangi bir puan ile puanlanırsa varyans homojenliğinin değişmemesi beklenir.

**varyans heterojenliği**

İlerleyen konularda göreceğimiz gibi, uygun koşullar altında, varyans homojenliği, çıkan sonucun anlamını değiştirme ihtimaline rağmen teste önemli ölçüde zarar vermeden esnetilebilir. Ancak **varyans heterojenliği** (*farklı* varyanslara sahip evrenler) sorunu olduğu




© Cengage Learning

**Şekil 11.1** Evrene ait saldırganlık puanlarının grafikle gösterimi

durumlar da vardır. Kullandığımız örnekte varyans heterojenliğinin olduğunu görebiliriz, ancak bu durumu şimdilik geçeceğim ve daha sonra bu duruma geri döneceğim.

### Normallik

Varyans analizinin ikinci sayıltısı, her koşuldaki saldırganlık puanları kendi ortalamaları etrafında normal dağılır. Başka bir ifadeyle, Şekil 11.1'deki her dağılım normaldir. Sayıltıyı yazmak için daha doğru yol, "hatanın, koşullar içindeki dağılımı normaldir" ifadesiyle yazmaktır, çünkü $\sigma_e^2$ her bireyin o koşulun puan ortalaması etrafındaki değişkenliğini gösterir. Böylelikle sık sık "hatanın normal dağılımı" sayıltısı terimiyle karşılaşırız. (Aynı anlama gelen "artıkların normal dağılımı" kavramı ile de karşılaşabilirsiniz.) Normallikten orta derecede ayrılış, genellikle ölümcül değildir. Varyans analizinin özel durumu olan iki bağımsız örneklem için *t* testine baktığımızda çoğunlukla aynı şeyi söyledik.

### Bağımsızlık

Üçüncü sayıltımız gözlemlerin birbirinden bağımsız olmasıdır. (Teknik olarak bu sayıltı, hata bileşenlerinin ($e_{ij}$) bağımsız olduğunu, ancak bunun basit desenler için geçerli olduğunu ifade eder, aynı şeyin miktarlarında olduğu gibi.) Bu yüzden deneysel işlem içindeki herhangi iki gözlem için, bu gözlemlerden herhangi birinin işlem (veya evren) ortalamasına göre konumlanma şeklinin, bize diğer gözlemler hakkında hiçbir şey açıklamadığını varsayarız. Bu, bireylerin gruplara neden seçkisiz olarak atandığının en önemli nedenlerinden biridir. Bağımsızlık sayıltısının gözardı edilmesi analiz için ciddî sonuçlara yol açabilir (bkz. Kenny ve Judd, 1986).

### Sıfır Hipotezi

Bildiğimiz gibi, Giancola ve Corman hatırlatma görevi verilerek dikkat dağıtıcı düzey ile saldırganlık düzeyinin değiştiğiyle ilgili araştırma hipotezini test etmeyle ilgilendiler. Bu hipotezi desteklemek, standart sıfır hipotezinin reddedilmesinden gelir,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

Sıfır hipotezi farklı sayılarda yanlış olabilir (örneğin; tüm ortalamalar birbirinden farklı olabilir, ilk ikisi birbirine eşit olabilir fakat diğer üç tanesi farklı olabilir vb.) fakat şimdilik sadece sıfır hipotezinin tamamen doğru veya tamamen yanlış olduğu durumlarla, Bölüm 12'de alt küme ortalamalarının eşit olup olmadığı problemiyle ilgileneceğiz.

## 11.3 Varyans Analizinin Mantığı

Varyans analizinin altında yatan mantık oldukça basittir ve onu anladığınızda tartışmanın geri kalan kısmı önemli ölçüde anlamlı olacaktır. Şimdi, normallik, varyansın homojenliği ve gözlemlerin (veya artıkların) bağımsızlığı gibi üç ana sayıltımızın etkilerini düşünelim: Sayıltıların ilk ikisini yaparak Şekil 11.1'de gösterilen beş dağılımın aynı şekle ve değişkenliğe sahip olduğunu söyledik. Sonuç olarak onları farklılaştırmak için geriye kalan tek yol ortalamalarıdır. (Normal dağılımın yalnızca $\mu$ ve $\sigma$ parametrelerine bağı olduğunu hatırlayınız.)

$H_0$ hakkında -doğru veya yanlış olabilir- hiçbir sayıltı yapmadan başlayacağız. Herhangi bir işlem için, gruptaki 12 puanın varyansı, puanların çekildiği evren varyansının kestirimi olurdu. Aynı zamanda ortak evren varyansının ($\sigma_e^2$) bir kestirimi olurdu çünkü tüm evrenlerin aynı varyansa sahip olduğunu varsaydık. Eğer bu durumu tercih ederseniz, "$\doteq$" simgesini "tarafından kestirilmiştir" şeklinde okuyarak

$$\sigma_1^2 \doteq s_1^2, \qquad \sigma_2^2 \doteq s_2^2, \qquad \cdots, \qquad \sigma_e^2 \doteq s_e^2,$$

eşitliklerini düşünebilirsiniz. Homojenlik sayıltısından dolayı tüm bunlar $\sigma_e^2$'ın kestirimlerdir. Güvenirliğin artması için ortalamalarını alarak beş farklı kestirimi bir araya toplayabiliriz. Eğer $n_1 = n_2 = \ldots = n_5$ ise ve $k$, işlem sayısı olduğuna göre (bu durumda beş oluyor).[1]

$$\sigma_e^2 \doteq s_e^2 \doteq \bar{s}_j^2 \doteq \sum s_j^2/k$$

$KO_{\text{hata(artık)}}$
$KO_{\text{grupiçi}}$

olur. Bu bize daha sonra $\boldsymbol{KO_{\text{hata(artık)}}}$ veya bazen $\boldsymbol{KO_{\text{grupiçi}}}$ olarak adlandırılacak olan evren varyansının kestirimini verir. Bu kestirimin $H_0$ hipotezinin doğruluğuna veya yanlışlığına bağlı olmadığını belirtmek gerekir, çünkü $s_j^2$ her örneklemden bağımsız hesaplanır. (1. gruptaki tüm gözlemlere 20 puan ekleyebilirsiniz ve bu durum değişkenliğini değiştirmeyecektir.) Giancola ve Corman'ın çalışmasındaki veri için, $\sigma_e^2$'nin birleştirilmiş kestirimi

$$\sigma_e^2 = (1,655^2 + 1,140^2 + 0,515^2 + 1,180^2 + 2,370^2)/5 = 11,313/5 = 2,263$$

olacaktır.

Şimdi $H_0$'ın doğru olduğunu varsayalım. Eğer durum böyle ise $n = 12$ olduğu durumdaki beş örneklemimiz aynı evrenden (veya eşdeğeri, beş özdeş evrenden) çekilen beş farklı bağımsız örneklem olarak düşünülebilir ve $\sigma_e^2$'nin bir diğer muhtemel kestirimini üretebiliriz. Aynı evrenden çekilen ortalamaların varyansının, örneklem büyüklüğüne bölünen evren varyansına eşit olduğunu ifade eden Bölüm 7'deki merkezî limit teoremini hatırlayınız. Eğer $H_0$ doğru ise, örneklem ortalamaları aynı evrenden (veya aynı miktara sahip eşdeğeri) çekilmiştir ve böylelikle beş örneklem ortalamasının varyansının kestirimi $\sigma_e^2/n$ olur. Her örneklemin boyutu n ise,

$$\frac{\sigma_e^2}{n} \doteq s_{\bar{X}}^2$$

olur. Genel sıralamayı koruyabiliriz ve $\sigma_e^2$'nin ikinci kestirimini elde etmek için örneklem ortalamalarımızın varyanslarını ($s_{\bar{X}}^2$) hesaplayabiliriz.

$$\sigma_e^2 \doteq n s_{\bar{X}}^2$$

$KO_{\text{işlem}}$

Bu ifade $\boldsymbol{KO_{işlem}}$($\boldsymbol{KO_{treatment}}$) veya sıklıkla kısaltılmış hali $KO_{\text{işlem}}$ veya $KO_{\text{grup}}$ olarak adlandırılır; bu konuya kısa zaman içinde geri döneceğiz.

Şu an elimizde iki evren varyans ($\sigma_e^2$) kestirimi bulunmaktadır Bu kestirimlerden biri ($KO_{\text{hata}}$), $H_0$'ın doğruluğu veya yanlışlığından bağımsızdır. Diğeri ($KO_{\text{işlem}}$), sadece $H_0$ hipotezi doğrulandığında $\sigma_e^2$'nin kestirimidir (sadece merkezî limit teoriminin koşulları karşılandığında, yani ortalamaların bir evrenden veya birçok özdeş evrenden çekildiği durumlarda). Böylece iki kestirim birbirleriyle örtüşüyorsa $H_0$'ın doğruluğunu desteklemiş olacağız ve eğer birbirleriyle örtüşmüyorsa $H_0$'ın yanlışlığını destekleyeceğiz.[2]

---

[1] Bu sadece ikiden fazla grup olmasına rağmen, $t$ testlerinin birleşik hata terimlerinin kullanımının bir uzantısıdır. Eğer örneklem büyüklükleri eşit değilse, beş kestirimin ortalamasını alırız ama bu durumda her kestirimi her örneklemin serbestlik derecesine göre ağırlıklandırmamız gerekir. Bu işlemin aynısını Bölüm 7'de de yapmıştık.

[2] Öğrenciler genellikle, gerçekte mantıksal olarak farklı örneklemlerden çekildiğini bildiğimiz halde, "ortalamalar aynı örneklemden çekilmiş" ifadesinin kullanılmasıyla ilgili problem yaşarlar. Erkeklerin ve kadınların iki farklı insan evreni olduğunu bildiğimiz halde, 'erkeklerle kadınların ortalamaları bir evrenden geliyor' demek oldukça saçma görünür. Eğer bayanların evren puanları, erkeklerin evren puanları ile tam olarak aynı ise, bu durumda, mantıksal olarak evren puanlarının aynı olduğu söylenebilir ve buna göre davranılabilir.

Önceki tartışmaya istinaden, varyans analizinin mantığından kısaca bahsedebiliriz. $H_0$'ı test etmek için birisi $H_0$'ın doğruluğu veya yanlışlığından bağımsız olan ve diğeri $H_0$'a bağımlı olan iki evren varyansı kestirimini hesaplayabiliriz. Eğer iki kestirim örtüşürse, $H_0$'ı reddetmek için hiçbir neden yoktur. Eğer iki kestirim yeterince örtüşmezse, grup farklılıklarının ikinci kestirimimize katkı sağladığı, onu şişirdiği ve birinciden farklılaşmasına yol açtığı sonucuna ulaşabiliriz. Bu yüzden $H_0$'ı reddederiz.

## Varyansın Kestirilmesi

**işlem etkisi**

Bu noktada kestirmekte olduğumuz iki değeri ispatsız ifade etmek faydalı olacaktır. Öncelikle $\tau_j$ olarak gösterilen **işlem etkisini** grup ortalamaları ($\mu_j$) ile genel ortalama ($\mu$) arasındaki fark ($\mu_j - \mu$) olarak tanımlayacağız ve $\theta_\tau^2$ ifadesini gerçek evren ortalamalarının ($\mu_1, \mu_2, ..., \mu_5$) varyansı olarak alacağız.[3]

$$\theta_\tau^2 = \frac{\sum(\mu_j - \mu)^2}{k - 1} = \frac{\sum \tau_j^2}{k - 1}$$

Ek olarak, bir istatisiğin beklenen değerini [E () olarak yazılır], limit kavramına başvurarak sınırsız sayıdaki deneme sonunda ulaşılacak değer olarak tanımlamıştık. Bu değer, alınacak tüm örneklemlerden elde edilen kestirimlerin ortalamasıdır. Bu kavramlarla,

$$E(KO_{\text{hata}}) = \sigma_e^2$$

$$E(KO_{\text{işlem}}) = \sigma_e^2 + \frac{n\sum \tau_j^2}{k - 1}$$

$$= \sigma_e^2 + n\theta_\tau^2$$

olduğunu ifade edebiliriz.

Şimdi, $H_0$ doğruysa ve $\mu_1 = \mu_2 = ... = \mu_5 = \mu$ ise, evren ortalamaları değişmez ve $\theta_\tau^2 = 0$ olur. Öyleyse,

$$E(KO_{\text{hata}}) = \sigma_e^2$$

ve

$$E(KO_{\text{işlem}}) = \sigma_e^2 + n(0) = \sigma_e^2$$

ve böylece

$$E(KO_{\text{hata}}) = E(KO_{\text{işlem}})$$

Bu değerlerin beklenen değerler olduğunu aklınızda tutun, pratikte kareler ortalamasına dayalı iki örneklem nadiren sayısal olarak eşit olur.

Eğer $H_0$ yanlışsa, $\theta_\tau^2$ sıfır olmayacak, pozitif bir sayı olacaktır. Bu durumda

$$E(KO_{\text{hata}}) < E(KO_{\text{işlem}})$$

Çünkü $KO_{\text{işlem}}$, $\mu_j$'ler arasındaki gerçek farkları temsil eden sıfırdan farklı terimler içerir.

---

[3] Gerçek parametreye ($\mu$) sahip olduğumuzdan ve $k - 1$ yerine $k$ ile bölmemiz gerektiğinden, teknik olarak, $\theta_\tau^2$ bir varyans değildir. Yine de ne yaptığımızı aklımızda tuttuğumuz sürece bunu varyans olarak düşünmemiz çok az kayba uğratacaktır. Birçok kitap, bu kitabın önceki baskısı da dahil, $\theta_\tau^2$'yi $\sigma_\tau^2$ olarak gösterir. Bu baskıda dürüst olmaya karar verdim $\theta_\tau^2$ kullandım.

# 11.4 Varyans Analizinde Hesaplamalar

Bu noktada varyans analizinde kullanılan hesaplamaları örneklendirmek için Giancola ve Corman'ın örneğini kullanacağız. Varyans analizini gerçekleştirmek için her zaman bilgisayar yazılımı kullanacağınızı düşünmenize rağmen, hesaplamaları hesap makinesi kullanarak yapmayı anlamak çok önemlidir. Öncelikle, bu temel işlemleri anlamanıza yardım eder. Ayrıca, bazı karmaşık durumları ve önerilen alternatif analizleri anlamayı kolaylaştırır. Son olarak, her istediğinizi yapacak bir bilgisayar programı yoktur ve duruma göre doğrudan hesaplamalara başvurmalısınız. Vaktimizi boşa harcadığımı düşünseniz bile sabırla hesaplamaları yapacağım.

## Kareler Toplamı

**kareler toplamı**

Varyans analizinde hesaplamaların çoğu **kareler toplamı** ile ilgilidir. Bölüm 9'da gördüğümüz gibi, kareler toplamı, sadece ortalamadan farkların karelerinin toplamıdır $\sum(X-\overline{X})^2$ veya daha sıklıkla onların çarpımıyla ilgili sapma kareler toplamıdır. Örneklem varyansını ilk kez tanımladığımızda,

$$s_X^2 = \frac{\sum(X-\overline{X})^2}{n-1} = \frac{\sum X^2 - (\sum X)^2/n}{n-1}$$

olduğunu gördük.

Burada, pay, $X$'in *kareler toplamı* ve payda da serbestlik derecesidir. Kareler ortalaması ve varyanslar yalnızca aynı sayıda serbestlik derecesine sahip olduğunda toplanabilirken, kareler toplamları toplanabilir olma avantajına sahiptir.

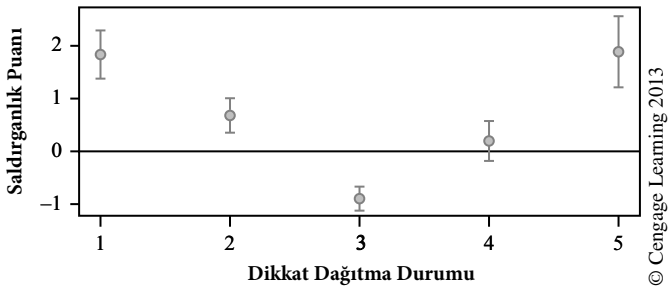
## Veri

Veri, Şekil 11.2'deki veri grafiği ve Tablo 11.3'teki hesaplamalarla birlikte Tablo 11.2'de yeniden üretilmiştir (Şekil 11.2'deki "hata çubukları" ± bir standart hata olarak etiketlenmiştir.

**Tablo 11.2** Görev güçlüğünün bir fonksiyonu olarak uygulanan şok düzeyi

| | D0 | D2 | D4 | D6 | D8 | Toplam |
|---|---|---|---|---|---|---|
| | 1,28 | −1,18 | −0,41 | −0,45 | 2,01 | |
| | 1,35 | 0,15 | −1,25 | 0,54 | 0,40 | |
| | 3,31 | 1,36 | −1,33 | −0,98 | 2,34 | |
| | 3,06 | 2,61 | −0,47 | 1,68 | −1,80 | |
| | 2,59 | 0,66 | −0,60 | 2,25 | 5,00 | |
| | 3,25 | 1,32 | −1,72 | −0,19 | 2,27 | |
| | 2,98 | 0,73 | −1,74 | −0,90 | 6,47 | |
| | 1,53 | −1,06 | −0,77 | 0,78 | 2,94 | |
| | −2,68 | 0,24 | −0,41 | 0,05 | 0,47 | |
| | 2,64 | 0,27 | −1,20 | 2,69 | 3,22 | |
| | 1,26 | 0,72 | −0,31 | 0,15 | 0,01 | |
| | 1,06 | 2,28 | −0,74 | 0,91 | −0,66 | |
| Ortalama | 1,802 | 0,675 | −0,912 | 0,544 | 1,889 | 0,800 |
| Std. Sapma | 1,656 | 1,140 | 0,515 | 1,180 | 2,370 | 1,800 |
| Varyans | 2,741 | 1,299 | 0,265 | 1,394 | 5,616 | 3,168 |

© Cengage Learning 2013




**Şekil 11.2** Giancola ve Corman'ın dikkat dağıtıcı düzeyin bir fonksiyonu olarak saldırganlık verisinin grafiği. Çubuklar ±1 standart hatayı temsil eder.

**Tablo 11.3** Tablo 11.2'deki veri için hesaplamalar

$$KT_{\text{toplam}} = \sum (X_{ij} - \overline{X})^2 = (1,28 - 0,80)^2 + (1,35 - 0,80)^2 + \ldots + (-0,66 - 0,80)^2$$
$$= 186,918$$
$$KT_{\text{işlem}} = n\sum (\overline{X}_j - \overline{X})^2 = 12\ (1,80\ -0,80)^2 + (0,675\ -0,80)^2 + \ldots\ + (1,889\ -0,80)^2)$$
$$= 12(5,205) = 62,460$$
$$KT_{\text{hata}} = KT_{\text{toplam}} - KT_{\text{işlem}} = 186,918 - 62,640 = 124,485$$

**Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| İşlemler | 4 | 62,460 | 15,615 | 6,90* |
| Hata | 55 | 124,458 | 2,263 | |
| Toplam | 59 | 186,918 | | |

* $p < ,05$

© Cengage Learning 2013

Bu durumu iyi bilmesi gereken birçok insan bu belirtmeyi atlar ve onların standart sapmaları veya standart hataları mı belirttiklerini veya onların bir birim mi iki birim mi, uzunlukta olduğunu veya güven aralıkları olup olmadığını bilemeyiz. Her zaman çubukların neyi temsil ettiğini belirtmeliyiz.) Hesaplamaları ve sonuçları ayrıntılı olarak tartışacağız. Gerçek veri noktaları farazî olduğundan (ortalamalar ve varyanslar farazi olmadığı halde), bireysel gruplardaki gözlemlerin dağılımlarını sınayarak kazanılabilecek küçük bir şey vardır - veri gerçek olarak normal dağılıma sahip bir evrenden çekilmiştir. Ancak gerçek verinin ciddi anlamda çarpık veya çift modlu olmadığından ve hatta daha önemlisi farklı yönlere çarpık olmadığından emin olmak için öncelikle bu dağılımların incelenmesi önemlidir. Hatta bu örnek için, varyansın homojenliği sayıltısını kontrol eder gibi bireysel grup varyanslarını incelemek faydalıdır. Varyanslar bizim istediğimiz kadar benzer olmamasına rağmen, standart varyans analizini uygulayacağız ve ardından varyansın heterejonliği konusuna geri döneceğiz. İlerde göreceğimiz gibi varyans analizi, eşzamanlı olarak bütün gruplara bakarak, özellikle her grupta aynı sayıda gözleme sahip olduğumuzda, sayıltıların ihlâline karşı güçlüdür.

Tablo 11.3 tek yönlü varyans analizini yapmak için gerekli hesaplamaları göstermektedir. Bu hesaplamaların ayrıntılandırılması gerekir.

## $KT_{toplam}$

$KT_{toplam}$

**$KT_{toplam}$** (toplam kareler toplamı şeklinde okunur), hangi işlemin ürettiğine bakılmaksızın tüm gözlemlerin genel ortalamadan sapmalarının kareleri toplamını gösterir. $\overline{X}$ genel ortalamayı gösterdiğine göre formülü aşağıdaki gibidir:

$$KT_{\text{toplam}} = \sum(X_{ij} - \overline{X}..)^2$$

Bu, çok önceden bir sayı kümesinin varyansını hesaplarken varyansın payında yer alan terimdir (payda serbestlik derecesiydi). Bu formül, onu izleyen diğerleri gibi, bu problem için elde hesaplama yapmak zorunda kaldığımızda kullandığımız formül değildir. Ancak bunlar kesin olarak doğru formüllerdir ve analiz hakkında normal düşünme yolumuzu gösterir. Geleneksel el ile hesaplama formüllerini tercih edenler, bu kitabın daha eski baskılarında bulabilirler.

## $KT_{işlem}$

$KT_{işlem}$

$KT_{işlem}$, grup ortalamalarının genel ortalamadan sapmaları yardımıyla aşağıdaki gibi tanımlanır:

$$KT_{\text{işlem}} = \sum(X_j - \overline{X}..)^2$$

**$KT_{işlem}$**'in, $n$ ile çarpılınca, evren varyansının bir kestirimini veren, genel ortalama etrafındaki işlem ortalamalarının sapma karelerinin toplamı olduğunu görebilirsiniz.

## $KT_{hata}$

$KT_{hata}$

Pratikte **$KT_{hata}$** çıkarma ile elde edilir.

$$KT_{\text{toplam}} = KT_{\text{işlem}} + KT_{\text{hata}}$$

Ayrıca aşağıdaki şekilde de gösterilebilir.

$$KT_{\text{hata}} = KT_{\text{toplam}} - KT_{\text{işlem}}$$

Bu Tablo 11.3'teki verilen işlemdir ve hesaplamalarımızı yapmayı kolaylaştırır.

$KT_{\text{hata}}$, ortalamadan sapmalarla aşağıdaki gibi yazılır:

$$KT_{\text{hata}} = \sum(X_{ij} - \overline{X}_j)^2$$

Burada $KT_{\text{hata}}$'nın, puanların kendi ortalamalarından sapma karelerinin gruplar üzerine toplamı olduğunu görebilirsiniz. Bu yaklaşım, her grup için kareler toplamını hesapladığım, aşağıdaki örnekte gösterilmektedir. Her grup içi verinin diğer gruplardan etkilenmesi durumunun kesinlikle olmadığında ve böylelikle sıfır hipotezinin doğrulanabilirliğinin veya yanlışlanabilirliğinin hesaplamalarla ilgisi olmadığına dikkat ediniz.

$$KT_{\text{grup içi}}D0 = \sum((1,28 - 1,802)^2 + (1,35 - 1,802)^2 + ... + (1,06 - 1,802)^2) = 30,148 \quad = \quad 30148$$

$$KT_{\text{grup içi}}D2 = \sum((1,18 - 0,675)^2 + (0,15 - 1,675)^2 + ... + (2,28 - 1,675)^2) \quad = \quad 14,291$$

$$KT_{\text{grup içi}}D4 = \sum((-0,41 - (-0,912)^2 + (-1,25 - (-0,912)^2 + ... + (-0,74 - (-0,912)^2) \quad = \quad 2,919$$

$$KT_{\text{grup içi}}D6 = \sum((-0,85 - 0,144)^2 + (0,14 - 0,144)^2 + ... + (0,51 - 0,144)^2) \quad = \quad 15,330$$

$$KT_{\text{grup içi}}D8 = \sum((2,01 - 1,889)^2 + (0,40 - 1,889)^2 + ... + (-0,66 - 1,889)^2) \quad = \quad 61,771$$

$$KT_{\text{hata}} = \quad = \quad 124,458$$

Bireysel terimleri topladığımızda 124,458 değerini elde ederiz ki bu değer, Tablo 11.3'ten elde ettiğimiz cevapla uyumludur.

## Özet Tablo

**özet tablo**

Tablo 11.3 aynı zamanda varyans analizi **özet tablo**sunu da göstermektedir. Birçok hesaplamayı özetlediği, verinin önerdiğini bir bakışta söylemeyi mümkün kılması gibi nedenlerle özet tablo olarak adlandırılır. Eski yayınlarda tüm özet tabloların gösterildiğini göreceksiniz. Son yıllarda, alandan tasarruf etmek amacıyla genellikle *F* istatistiği sonuçları ve serbestlik derecesi verilmektedir.

## Değişkenlik Kaynakları

Özet tablonun ilk sütunu varyansın kaynağını içerir ("değişkenlik" kelimesi, "kareler toplamı" ile eşanlamlı olarak kullanılmıştır). Tablodan da görülebileceği gibi, işlemden kaynaklı değişkenlik (işlem ortalamaları arasındaki değişkenlik), hatadan kaynaklı değişkenlik (gruplar içindeki değişkenlik) ve toplam değişkenlik olmak üzere üç farklı varyans kaynağı bulunmaktadır. Bu kaynaklar, biri gruplar içerisindeki değişkenlik ve diğeri birçok grup ortalaması arasındaki değişkenlik olmak üzere toplam kareler toplamının iki bölüme ayrıldığı olgusunu yansıtır.

## Serbestlik Derecesi

$sd_{\text{toplam}}$
$sd_{\text{işlem}}$
$sd_{\text{hata}}$

Tablo 11.3'teki serbestlik derecesi sütunu, iki değişkenlik kaynağı arasındaki toplam serbestlik derecesinin paylaşımını gösterir. Toplam 59 serbestlik derecesinden dört tanesi işlem ortalamaları arasındaki farklarla ve geriye kalan 55'i işlem grupları içindeki değişkenlikle ilgilidir. Serbestlik derecesinin hesaplanması muhtemelen işimizin en kolay bölümüdür. Toplam serbestlik derecesi ($\boldsymbol{sd}_{\textbf{toplam}}$), *N* toplam gözlem sayısı olduğuna göre, her zaman *N*-1'dir. İşlemler arası serbestlik derecesi ($\boldsymbol{sd}_{\textbf{işlem}}$), *k* grup sayısı olduğuna göre, $k - 1$ tanedir. Hatanın serbestlik derecesi ($\boldsymbol{sd}_{\textbf{hata}}$) $\boldsymbol{sd}_{\textbf{toplam}}$'dan $sd_{\text{grup}}$ çıkarılarak elde edilen ve en kolay hesaplanabilendir. Ancak $sd_{\text{hata}}$ her grup içi serbestlik derecesini toplayarak da doğrudan hesaplanabilir.

Bu durumu biraz farklı biçime koymak istersek, toplam değişkenlik *N* puan sayısına dayalıdır ve dolayısıyla $N - 1$ serbestlik derecesine sahiptir. İşlem ortalamalarının değişkenliği *k* ortalamaya dayalıdır ve böylece $k - 1$ serbestlik derecesine sahiptir. Herhangi bir işlem içindeki değişkenlik *n* puan sayısına dayalıdır ve $n - 1$ serbestlik derecesine sahiptir fakat işlem içinde bunlardan *k* tanesini topladığımız için $k(n - 1)$ serbestlik derecesine sahip oluruz.

## Kareler Ortalamaları

Şimdi Tablo 11.3'teki *KO* sütununa gideceğiz. (Burada *KT* sütunu hakkında küçük bir şey söylemek gerekir; *KT* sütunu hesaplamalar bölümünden elde edilen kareler toplamını içerir.) Kareler ortalaması sütunu $\sigma_e^2$'nin iki kestirimini içerir. Bu değerler kareler toplamına denk gelen serbestlik derecesine bölünerek elde edilir. Böylece 62,460/4 = 15,615 ve 124,458/55 = 2,263 bulunur. Genel olarak $KO_{\text{toplam}}$'ını hesaplamayız, çünkü ona ihtiyacımız yoktur. Bunu yapacak olsaydık, Tablo 11.3'te de görebileceğiniz gibi, bu terim 186,918/59 = 3,168'e eşit olacaktır, bu terim işlemlere bakılmadan *N* gözleme ait varyan-

sıdır. Kareler ortalamasının varyans kestirimi olması doğru olduğu halde, bu terimlerin hangi varyansı kestirdiğini akılda tutmak önemlidir. Böylece $KO_{\text{hata}}$, $H_0$'ın doğruluğuna veya yanlışlığına bakılmaksızın evren varyansının ($\sigma_e^2$) bir kestirimidir ve örneklem büyüklükleri eşit olduğunda her grup içindeki ortalama varyanstır.

$$KO_{\text{hata}} = (2{,}741 + 1{,}299 + 0{,}265 + 1{,}394 + 5{,}616)/5 = 2{,}263$$

Ancak $KT_{\text{işlem}}$ işlem ortalamalarının varyansı değildir, daha çok evren varyansının ($\sigma_e^2$) ikinci kestirimini üretmek için $n$ ile çarpılan ortalamaların varyansıdır.

$$s^2_{\text{işlem}} = \text{varyans}\ (1{,}802;\ 0{,}675;\ -0{,}912;\ 1{,}889) = 1{,}301$$
$$KO_{\text{işlem}} = \text{n}(s^2_{\text{işlem}}) = 12(1{,}301) = 15{,}615$$

## *F* İstatistiği

11.3'teki son sütun $F$ olarak adlandırılmıştır ve sıfır hipotezini test etmede en önemlilerden biridir. $F$, $\text{KT}_{\text{işlem}} / \text{KO}_{\text{hata}}$ eşitliği ile elde edilir. Bu oranın neden mantıklı olduğunu açıklamak için bir kesin yol, bir de özensiz yol vardır ve biz ikinci ile başlayacağız. Daha önce de belirttiğimiz gibi $KO_{\text{hata}}$, evren varyansının ($\sigma_s^2$) bir kestirimidir. Ayrıca $KO_{\text{işlem}}$, $H_0$ doğru olduğunda evren varyansının ($\sigma_s^2$) bir kestirimidir, $H_0$ doğru olmadığında değildir. Eğer $H_0$ doğru ise, $KO_{\text{işlem}}$ ve $KO_{\text{hata}}$'nın her ikisi de aynı şeyi kestiriyordur ve aslında hemen hemen eşit olurlar. Eğer durum öyle ise, birinin diğerine oranı 1'e yaklaşacaktır, örnekleme hatası için belli bir miktar alacak veya verecektir. Böylece bütün yapmamız gereken, oranı hesaplamak ve sıfır hipotezine destek göstermek için 1'e yeterince yakın olup olmadığını belirlemektir.

*F*'ye bakmak için birçok informel yol vardır. En doğru yaklaşım hata ve işlemler için *beklenen kareler ortalamalarıyla* başlamaktır. Bu bölümün baş kısımlarından aşağıdaki eşitlikleri biliyoruz:

$$E(KO_{\text{hata}}) = \sigma_e^2$$
$$E(KO_{\text{işlem}}) = \sigma_e^2 + n\theta_\tau^2$$

Şimdi aşağıdaki oranı oluşturabiliriz:

$$F = \frac{E(KO_{\text{işlem}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + n\theta_\tau^2}{\sigma_e^2}$$

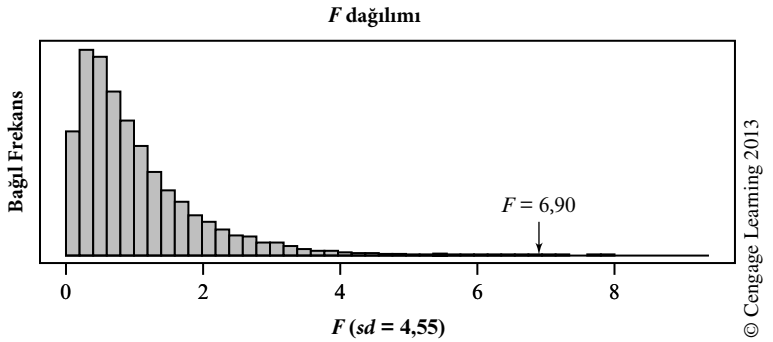
Bu oranın 1 olması, sadece $\theta_\tau^2 = 0$ yani $H_0$ doğru ve $\mu_1 = \ldots = \mu_5$ olduğunda gerçekleşir.[4] $\theta_\tau^2 > 0$ olduğunda beklenen değer 1'den büyük olacaktır.

Geriye kalan soru, beklenen değerleri kullanmadığımızda ancak veri kümesinden hesaplanan ve örneklem hatasına bağlı olan kareler ortalaması elde ettiğimizde hangi büyüklükte bir oranı $H_0$'ı reddetmeden kabul edeceğimizdir? Bu sorunun cevabı,

$$F = KO_{\text{işlem}}/KO_{\text{hata}}$$

---

4 Her ne kadar $\theta_\tau^2 = 0$ iken $\frac{E(KO_{\text{işlem}})}{E(KO_{\text{hata}})} = 1$ olsa da $H_0$ altında ***F***'nin beklenen değeri tam olarak 1 değildir. $H_0$ altında, tam olarak $E(F) = \frac{sd_{\text{işlem}}}{sd_{\text{hata}} - 2}$ olur. Bütün pratik amaçlar için, $F$'nin $H_0$ altında 1 ve $H_1$(alternatif hipotez) altında 1'den büyük olmasını beklemek herhangi bir kayıp yaratmaz.




**Şekil 11.3** 4 ve 55 serbestlik derecelerinde F dğılımı

oranının $k$-1 ve $k(n$-1) serbestlik derecesinde $F$ gibi dağıldığını gösterebileceğimiz durumda yatar. Bu durum iki varyans kestirim oranını test etmede kullanılan $F$ dağılımı ile aynıdır (burada yaptığımız durum ile aynı). Serbestlik derecesi sırasıyla pay ve paydaya ait serbestlik derecesini belirtir.

Öncelikle 4 ve 55 serbestlik derecesine sahip olduğumuzda ve sıfır hipotezi doğru olduğunda $F$ dağılımının nasıl göründüğüne bakacağız. Şekil 11.3'teki grafiği oluşturmak için her birinin ortalaması 0,800 (örneğimizdeki genel ortalama), birbirine eşit ve standart sapmaları Giancola ve Corman tarafından bulunanlara eşit, tek normal bir evrenden 5 örneklem çektim. Tüm örneklemler aynı ortalamaya sahip evrenlerden çekildiği için sıfır hipotezimiz tanımına göre doğrudur. Örneklemleri çektikten sonra $F$ istatistiğini hesapladım ve kaydettim. Daha sonra bu işlemi 9999 defa tekrarladım, böylelikle sıfır hipotezi altında 10000 $F$ değeri ürettim. (Bu hesaplamaları gerçekleştirmek için gereken R programı kitabın Web sayfasında bulunmaktadır.) Sonucun dağılımı Şekil 11.3'te görülmektedir. Mevcut veriden hesapladığımız $F$'nin, sıfır hipotezini reddetmemizi sağlayan yani $H_0$'ın beklendiği gibi doğru olması durumuna çok yakın olduğunu görebilirsiniz. Sıfır hipotezi altında F(4,55) = 6,90'nın gerçek olasılığı ,00014'tür.

Şekil 11.3'teki gibi bir örnekleme çalışmasıyla kesinlikle çalışmak istemeyiz, her zaman $F$'nin manidar olup olmadığını bilmek isteriz. Neyse ki $F$ tablolaştırılmıştır ve kritik değerlere basitçe bakabiliyoruz. Örneğimiz için $F$ = 6,90'dır. Pay için 4 ve payda için 55 serbestlik derecesine sahibiz ve bu değerlerle $F$ tablosuna (*EK F*) girebiliriz. Bir bölümü Tablo 11.4'te gösterilen *EK F*, $\alpha$ = ,05 ve ,01 için kritik değerleri verir. Bu özel durumumuz için 4 ve 55 serbestlik derecesine ve doğrusal eklentiyle $F_{,05}$ (4,55) = 2,54'tür. Eğer $\alpha$ = ,05 ile çalışmayı seçseydik, $H_0$ hipotezini ret edecektik ve işlem ortalamaları arasında manidar farklılıklar vardır sonucuna ulaşacaktık.

## Sonuçlar

$F$ değerinin manidarlığının temelinde, evrendeki işlem ortalamalarının arasındaki farkın sıfır olduğu hipotezini reddederiz. Kesin konuşacak olursak, bu sonuç evren ortalamalarından en az birinin, diğerlerinin ortalamasından farklı olduğu sonucunu gösterir; fakat hangi ortalamanın diğer ortalamalardan farklı olduğunu bilemeyiz. Bu konuya Bölüm 12'de devam edeceğiz. Dikkat dağıtıcı düzeyinin *belirli bir noktaya* kadar artırılmasının saldırganlık davranışını azalttığı Şekil 11.2'deki kutu grafiği incelendiğinde açık bir şekilde görülmektedir. Ancak bu noktayı aşacak şekilde dikkat dağıtıcı düzey artırılırsa, nokta ters etki yapar. Bu durum bize alkolün saldırganlık davranışındaki rolü hakkında en azından bir şeyler söylemektedir.

**Tablo 11.4** Ek F'nin kısaltılmış versiyonu. $\alpha$ = ,05 olduğunda F dağılımının kritik değerleri

| | Pay İçin Serbestlik Derecesi | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| sd. payda | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 161,4 | 199,5 | 215,8 | 224,8 | 230,0 | 233,8 | 236,5 | 238,6 | 240,1 | 242,1 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,35 | 19,37 | 19,38 | 19,40 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,89 | 8,85 | 8,81 | 8,79 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,77 | 4,74 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,64 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,35 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,14 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 |
| 12 | 4,75 | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 |
| 13 | 4,67 | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 |
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 |
| 26 | 4,23 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 |
| 28 | 4,20 | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 |
| 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 |
| 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 |
| 50 | 4,03 | 3,18 | 2,79 | 2,56 | 2,40 | 2,29 | 2,20 | 2,13 | 2,07 | 2,03 |
| 60 | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 |
| 120 | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,18 | 2,09 | 2,02 | 1,96 | 1,91 |
| 200 | 3,89 | 3,04 | 2,65 | 2,42 | 2,26 | 2,14 | 2,06 | 1,98 | 1,93 | 1,88 |
| 500 | 3,86 | 3,01 | 2,62 | 2,39 | 2,23 | 2,12 | 2,03 | 1,96 | 1,90 | 1,85 |
| 1000 | 3,85 | 3,01 | 2,61 | 2,38 | 2,22 | 2,11 | 2,02 | 1,95 | 1,89 | 1,84 |

© Cengage Learning 2013

## 11.5 Sonuçların Raporlanması

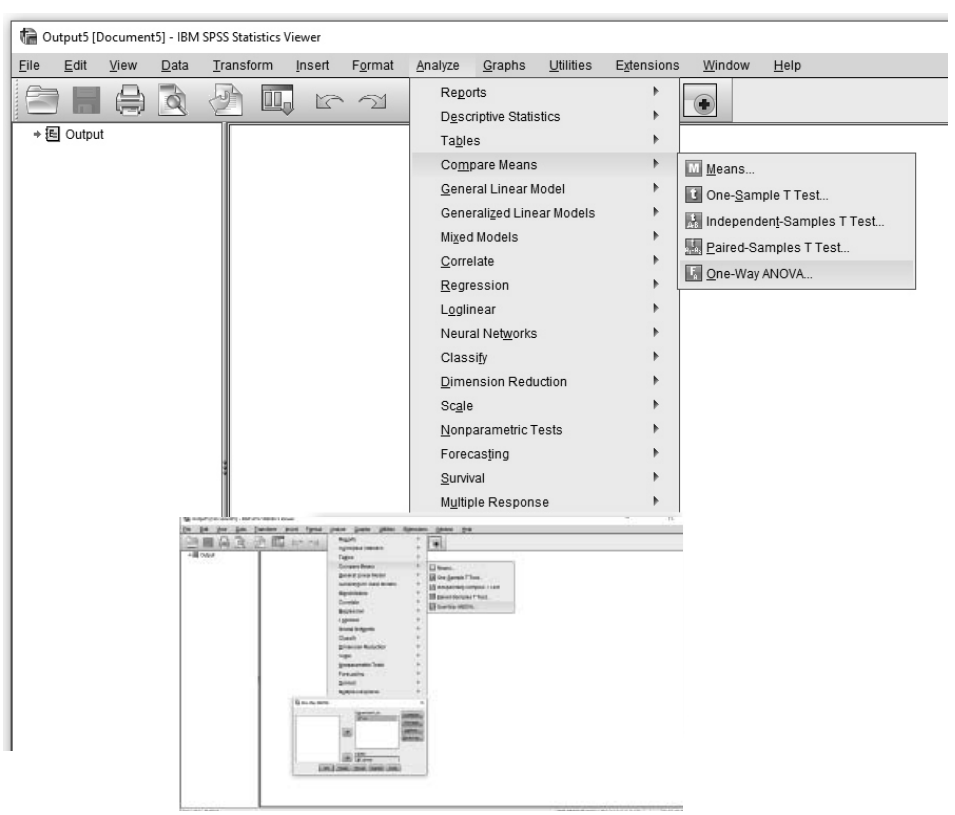
Varyans analizinin sonuçlarının raporlanması *t* testinin sonuçlarının raporlanmasına göre biraz daha karmaşıktır. Bunun nedeni sadece *F*'nin manidar olup olmadığını göstermek istememiz değil, aynı zamanda bireysel ortalamalar arasındaki farklılıklar hakkında da cümleler kurmak istememizdir. Bir sonraki bölüme kadar bireysel ortalamalar üzerindeki testleri tartışmayacağız, bu yüzden bu örnek yetersiz gelecektir. Bu konuya Bölüm 12'de geri döneceğiz. Sonuçlar hakkındaki cümlelerin sadeleştirilmiş hali aşağıdadır:

> "Alkol, daha az belirgin engelleyici ipuçlarından ziyade daha kışkırtıcı olanlara dikkatini verme eğilimindedir." hipotezinin testinde, Giancola ve Corman saldırganlık davranışı üzerinde dikkat dağıtıcı uyarıların etkisini inceledi. Gruplar, yarışma görevinde sağlanan dikkat dağıtıcı düzeyde farklılaştılar. Alkolü tükettikten sonra, deneklere 3 x 3'lük matrisin hücrelerinde yer alan 0, 2, 4, 6 veya 8 uyarıcı tekrarlı olarak verildi ve hangi uyarıcının geldiğini sırayla rapor etmeleri istendi. Deneklere hayali

rakipten daha hızlı cevapladıklarında, rakibe elektrik şoku verecekleri anlatıldı. Eğer daha yavaş olurlarsa, rakip onlara şok verecekti. Katılımcılara herhangi bir dikkat dağıtıcı uyarıcı verilmediğinde, deneyim edindikleri düşük düzeydeki dikkat dağıtıcıdan daha yüksek yoğunlukta ve daha uzun süreli şok yaşadılar. Buna ek olarak, dikkat dağıtma işleminin karmaşıklığı orta düzeyin üstünde arttığında, şok yoğunluğundaki azalma ve süre tersine doğru değişecektir. Tek yönlü varyans analizi beş grubun ortalamaları arasında manidar farklılıklar olduğunu gösterdi ($F(4,55) = 6,90$; $p < 0,05$). Grup ortalamalarının görsel incelemesi, uygulanan şok düzeyinin dikkat dağıtıcının artması ile azaldığı, fakat dikkat dağıtma işi daha karmaşık hale geldikçe kuramsal olarak tahmin edildiği gibi tekrar bir artışın olduğunu gösterdi. (Not: Bu farklılıklarla ilgili daha başka tartışmalar için Bölüm 12'nin beklenmesi gerekir. İlgili veri burada rapor edilmemesine rağmen, alkolün tüketilmediği kontrol koşul davranışı dikkat dağıtma düzeyinde manidar farklılıklar olmadığını gösterdi.)

## 11.6 Bilgisayar Çözümleri

Birçok varyans analizi standart bilgisayar yazılımlarınca yapılır ve Gösterim 11.1 SPSS program çıktısından bir örneği içerir. Diğer istatistik yazılımları benzer sonuçlar üretecektir. (Bu sonuçları R yazılımında üretmek için kitabın Web sayfasında yer alan GiancolaExample.R örneğine bakınız.) SPSS çıktısını üretmek için şunları yaptım: Sıklıkla kullanılan "General Linear Model/Univariate" yerine "Compare Means" menüsünden "One-Way" seçimini kullandım. Bu sonucu elde etmek için bağımlı ve bağımsız değişkenleri belirtmemiz ve daha sonra "descriptives" ve "Means Plot" seçmek için "Options" linkini tıklamamız gerekir.




**Gösterim 11.1** Tekyönlü SPSS çıktısı

Niteleyiciler

dv

| | N | Ortalama | Standart Sapma | Standart Hata | Ortalama İçin %95 Güven Aralığı Alt Sınır | Ortalama İçin %95 Güven Aralığı Üst Sınır | Minimum | Maksimum |
|---|---|---|---|---|---|---|---|---|
| 1 | 12 | 1,8025 | 1,65550 | ,47790 | ,7506 | 2,8544 | −2,68 | 3,31 |
| 2 | 12 | ,6750 | 1,13981 | ,32904 | −,0492 | 1,3992 | −1,18 | 2,61 |
| 3 | 12 | −,9125 | ,51515 | ,14871 | −1,2398 | −,5852 | −1,74 | −,31 |
| 4 | 12 | ,5442 | 1,18051 | ,34078 | −,2059 | 1,2942 | −,98 | 2,69 |
| 5 | 12 | 1,8892 | 2,36971 | ,68408 | ,3835 | 3,3948 | −1,80 | 6,47 |
| Toplam | 60 | ,7997 | 1,77992 | ,22979 | ,3399 | 1,2595 | −2,68 | 6,47 |

dv

| Levene İstatistiği | sd1 | sd2 | Manidarlık |
|---|---|---|---|
| 3,769 | 4 | 55 | ,009 |

ANOVA

dv

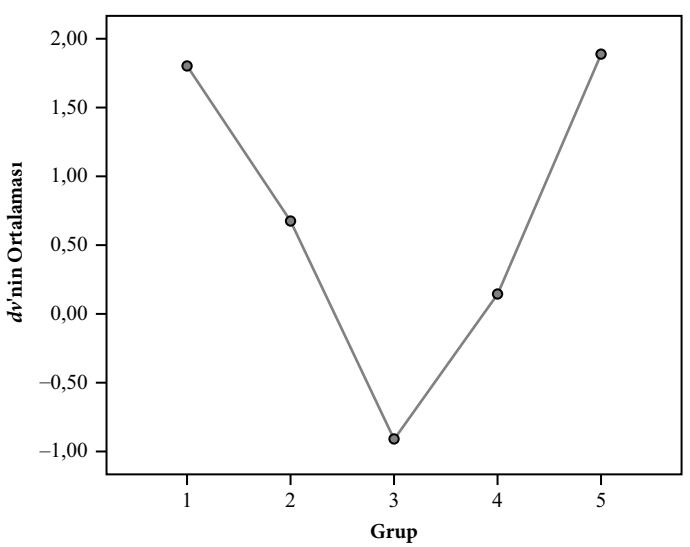
| | Kareler Toplamı | sd | Kareler Ortalaması | F | Manidarlık |
|---|---|---|---|---|---|
| Gruplar Arası | 62,460 | 4 | 15,615 | 6,901 | ,000 |
| Gruplar İçi | 124,458 | 55 | 2,263 | | |
| Toplam | 186,918 | 59 | | | |

Ortalamaların Eşitliği İçin Dayanıklı Testler

dv

| | İstatistik[a] | sd1 | sd2 | Manidarlık |
|---|---|---|---|---|
| Welch | 14,228 | 4 | 25,432 | ,000 |
| Brown-Forsythe | 6,901 | 4 | 32,943 | ,000 |

[a] Asimptotik dağılmış F.




Adapted from output by SPSS, Inc.

**Gösterim 11.1** *(devamı)*

Buradaki çıktılar bizim hesapladığımız değerlere benziyor. Fazladan bilgi içermesine rağmen özet tabloyu bu ünitenin sonuna kadar tartışmayacağım, menülerden "Analyze / General Linear Model / Univariate" kısmını seçseydiniz aynı sonucu elde edecektiniz. SPSS varyans homojenliğini test etmiştir (Bölüm 7'de tartışılan Levene testi) ve sonucun manidar olduğuna dikkat etmişsinizdir. Bu iyiye işaret değildir, bu durumu ünite içinde ilerledikçe tartışacağız. Ayrıca "Ortalamaların Eşitliğinin Dayanıklı Testleri" başlıklı bölüme dikkat edeceksiniz. Kısaca bu konuyu tartışacağız.

## 11.7 Eşit Olmayan Örneklem Büyüklükleri

**dengelenmiş desen**

Birçok deney her grup için aynı sayıda gözlem toplama fikri ile tasarlanır. (Bu tür desenler genellikle **dengelenmiş desen** olarak adlandırılır.) Ancak sıklıkla işler bu yönde gerçekleşmez. Denekler teste gelemez veya verilen görevleri yerine getiremediğinden dolayı elenirler. Hayvanlar deney sırasında tedavi edilemeyecek hastalıklara yakalanabilirler. Hayvanın, sürekli olarak deneycinin elini ısırmasından dolayı çalışmadan çıkarıldığı bir mezuniyet sonrası okulundaki örneği hatırlatmak isterim (Sgro ve Weinstock, 1963). Ayrıca okul sınıfları gibi eksiksiz gruplarda yürütülen çalışmaların, bu grupların boyutlarının hemen hemen her zaman değişebileceği gerçeği ile de mücadele etmesi gerekir.

Eğer örneklem büyüklükleri eşit değilse, daha önce tartıştığımız analizin değiştirilmesi gerekir. Bir bağımsız değişkenin olduğu durum için bu değişim oldukça küçüktür. (Regresyon desenleri ve çeşitli varyans analizi için gruplardaki kayıp veri ile ilgili daha fazla tartışmalar Howel'da (2008) veya daha kolay bir şekilde http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html adresinden bulunabilir.)

Daha önceden işlemlere ait kareler toplamını aşağıdaki eşitlikle tanımladık:

$$KT_{\text{işlem}} = n\sum(X_j - \overline{X}..)^2$$

Sapmaları $n$ ile çarpabilirdik, çünkü $n$ tüm işlemler için ortaktı. Eğer örneklem büyüklükleri farklılaşırsa ve $n_j$'yi $j$'nci işlemdeki ($\sum n_j = N$) birey sayısı olarak tanımlarsak bu ifadeyi,

$$KT_{\text{işlem}} = \sum\left[n_j(\overline{X}_j - \overline{X}..)^2\right]$$

biçiminde yazabiliriz. Tüm $n_j$ değerleri eşit olduğunda orjinal eşitliğe ulaşılır. Bu ifade, eşit olmayan $n$'ler durumunda her işlem ortalamasının genel ortalamadan sapmasının örneklem büyüklüğüyle ağırlıklandırıldığını gösterir. Böylece bütün diğer değerler eşit olduğunda, örneklemin büyüklüğü diğerlerine göre ne kadar fazla ise genel ortalamaya ($\overline{X}..$) ve $KT_{\text{işlem}}$'e o kadar katkı sağlar.

**tesadüfî kayıp**

Bu tür basit desenlerde bile kayıp veriye karşı biraz daha dikkati olmanız gerekir. Eğer insanların tesadüfî olarak boş bıraktığı veya diğerlerine göre daha az mevcuda sahip okuldaki bazı sınıflara ait veri **tesadüfî kayıp** olarak adlandırılırsa, bu durum çok problem olmayacaktır. Hepsinde kayıp veriye sahipseniz bu durum deneyinizin gücünü düşürür. Ancak puanların eksik olmasında sistematik nedenler varsa, bu farklı bir durumdur. Davranış problemleri için üç farklı gruba sahip olduğunuzu ve bu gruplardan bir tanesinin daha ciddî problemlere sahip insanlar için etkili olmadığını farz ediniz. Ciddî problemlere sahip birisi muhtemelen boş bırakacak, çünkü tedavi ona yaramayacaktır, o grup için ortalama, hafif probleme sahip insanların ortalamasıdır, boş bırakma problemi yaşamayan iki grup her tür probleme sahip insanların ortalamaları olacaktır. Üç ortalamayı karşılaştırmak gerçekten adil değildir, çünkü farklı türden insanlara dayalıdır. Problem eşit olmayan örneklem büyüklüklerinden birisi kadar değildir, fakat gruplar böyle olsa da karşılaştırılamaz.

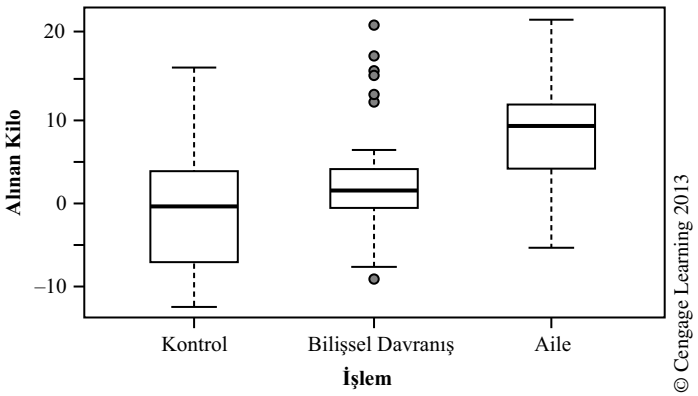
## İştahsızlık (Anoreksiya) İçin Etkili Terapiler

Aşağıdaki örnek iştahsız kızlarda kilo almada iki terapi koşulunun ve bir kontrol koşulunun etkilerini karşılaştıran Everitt'in çalışmasından alınmıştır. Veri, Hand ve arkadaşları (1994) tarafından rapor edilmiştir. Everitt herhangi bir müdahale almayan kontrol koşulunu, bilişsel-davranışsal tedavi koşulunu ve aile terapi koşulunu kullanmıştır. Burada analiz edilen bağımlı değişken sabit zaman aralıklarındaki ağırlığı elde etmekti. Bu örnekte boş bırakma problemi yoktu, bu yüzden bazı örneklemlerin diğerlerinden büyük olma durumu hakkında kaygılanacak bir duruma sahip değiliz. Veri kümesi Tablo 11.5'te verilmiş ve Şekil 11.4'te grafiği çizilmiştir. Veride, Bilişsel-Davranışsal Terapi grubu için biraz çift modluluk eğilimi varsa da, bu eğilim büyük olasılıkla sonuçlarımızı bozmaya yeterli ol-

**Tablo 11.5** Everitt'in küçük kızlardaki iştahsızlık işlemleri üzerindeki veri kümesi

| | Kontrol | Bilişsel Davranış Terapisi | Aile Terapisi | Toplam |
|---|---|---|---|---|
| | −,5 | 1,7 | 11,4 | |
| | −9,3 | ,7 | 11,0 | |
| | −5,4 | −,1 | 5,5 | |
| | 12,3 | −,7 | 9,4 | |
| | −2,0 | −3,5 | 13,6 | |
| | −10,2 | 14,9 | −2,9 | |
| | −12,2 | 3,5 | −,1 | |
| | 11,6 | 17,1 | 7,4 | |
| | −7,1 | −7,6 | 21,5 | |
| | 6,2 | 1,6 | −5,3 | |
| | −,2 | 11,7 | −3,8 | |
| | −9,2 | 6,1 | 13,4 | |
| | 8,3 | 1,1 | 13,1 | |
| | 3,3 | −4,0 | 9,0 | |
| | 11,3 | 20,9 | 3,9 | |
| | ,0 | −9,1 | 5,7 | |
| | −1,0 | 2,1 | 10,7 | |
| | −10,6 | −1,4 | | |
| | −4,6 | 1,4 | | |
| | −6,7 | −,3 | | |
| | 2,8 | −3,7 | | |
| | ,3 | −,8 | | |
| | 1,8 | 2,4 | | |
| | 3,7 | 12,6 | | |
| | 15,9 | 1,9 | | |
| | −10,2 | 3,9 | | |
| | | ,1 | | |
| | | 15,4 | | |
| | | −,7 | | |
| Ortalama | –0,45 | 3,01 | 7,26 | 2,76 |
| Std. Sapma | 7,989 | 7,308 | 7,157 | 7,984 |
| Varyans | 63,819 | 53,414 | 51,229 | 63,738 |
| *n* | 26 | 29 | 17 | 72 |

© Cengage Learning 2013




**Şekil 11.4** Everitt'in üç grubundaki kilo alma

mayacaktır. (Bir non-parametrik test (Bölüm 18) benzer sonuçlar üreten çift modluluktan etkilenmeyecektir.)

Varyans analizinin hesaplanması aşağıdadır ve eşit olmayan örneklem büyüklüğünün bulunmasıyla gerekli değişikliğin çok küçük olduğunu görebilirsiniz. Karmaşık desenlere geldiğimizde, eşit olmayan örneklem büyüklüklerinin kolaylıkla çıkarılmayacağını göstermek için hızlanmalıyım, fakat tek yönlü desenlerde böyle bir problem yoktur.

$$KT_{\text{işlem}} = \sum(\overline{X}_j - \overline{X}..)^2 = [(-0{,}5 - 2{,}76)^2 + (-9{,}3 - 2{,}76)^2 + ... + (10{,}7 - 2{,}76)^2]$$
$$= 4525{,}386$$

$$KT_{\text{işlem}} = \sum\left[n_j(\overline{X}_j - \overline{X}..)^2\right] = 26(-0{,}45 - 2{,}76)^2 + 29(3{,}01 - 2{,}76)^2 + 17(7{,}26 - 2{,}76)^2)$$
$$= 614{,}644$$

$$KT_{\text{hata}} = SS_{\text{toplam}} - SS_{\text{işlem}} = 4525{,}286 - 614{,}644$$
$$= 3910{,}742$$

Bu analiz için özet tablo aşağıdadır.

| Kaynak | *sd* | KT | KO | F |
|---|---|---|---|---|
| **İşlemler** | 2 | 614,644 | 307,322 | 5,422* |
| **Hata** | 69 | 3910,742 | 56,677 | |
| **Toplam** | 71 | 4525,386 | | |

* $p < ,05$

Özet tablodan, işlemin manidar bir etkisinin olduğunu görebilirsiniz. Bu etkinin varlığı, iki grubun önemli ölçüde kilo aldığını gösterirken, kontrol grubunda önemli ölçüde artışın olmadığını gösteren Şekil 11.4'te açıktır. Bilişsel-davranış grubunun ve Aile Terapisi grubunun manidar derecede farklı olup olmadığını, aynı zamanda her ikisinin kontrol grubundan farklılaşıp farklılaşmadığını henüz bilmiyoruz ancak bu problemi diğer bölüme kadar saklayacağız.

## 11.8 Varsayımlara Uyulmaması

Daha önce de gördüğümüz gibi, varyans analizi normallik ve varyansların homojenliği sayıltılarına dayanır. Ancak pratikte, varyans analizi dayanıklı bir istatiksel işlemdir ve sıklıkla bazı sayıtlılar göreceli olarak küçük etkilerle ihlâl edilebilmektedir. Bu durum özellikle normallik sayıltısı için geçerlidir. Bu problemle ilgili yapılan çalışmalar için Box (1953, 1954a, 1954b), Boneau (1960), Bradley (1964), ve Grissom'a (2000) bakabilirsiniz.

Sonraki referans diğerlerine göre biraz daha karamsardır, ancak özellikle belirli grup karşılaştırmalarından çok genel sıfır hipotezine odaklandığımızda normalliğin hayati bir sayıltı olmadığına ve varyans homojenliğinin çok kötü sonuçlar doğurmadan ihlâl edilebileceğine inanmak için hâlâ nedenimiz vardır.

Genel olarak, evrenlerin simetrik olduğu veya en azından benzer şekillere sahip olduğu varsayılırsa ve en büyük varyans en küçükten dört kattan fazla büyük değilse, varyans analizi büyük ihtimalle geçerli olacaktır. Bununla birlikte varyans heterojonliği ve eşit olmayan örneklem büyüklüklerini karıştırmamak gerektiğini söyleyebiliriz. Eşit olmayan varyansları tahmin edecek nedenleriniz varsa, örneklem büyüklüklerini mümkün olduğunca eşit tutmaya çalışınız. Bu ciddî bir durumdur ve insanlar eşit olmayan örneklem büyüklüğünün, testleri, varyans heterejonliğine karşı kayda değer şekilde daha güçsüz yaptığını unutma eğilimindedir.

Bölüm 7'de varyans heterojenlği için Levene testini (1960) göz önüne aldım ve Brown ve Forsythe (1974) ve O'Brien (1981) tarafından geliştirilen benzer testlerden bahsetmiştim. İlk ikisi aslında gözlemlerin örneklem ortalaması veya medyandan sapmaları (mutlak veya karesi) üzerine tanımlanan *t* testleridir. Eğer bir grup diğerinden daha yüksek varyansa sahipse, puanların ortalama veya medyandan sapmaları da aynı zamanda küçük varyansa sahip olan gruptan daha büyük olacaktır. Böylece, sapma değerlerinin mutlak değerleri üzerindeki manidar bir *t* testi, grup varyansı üzerindeki bir testi temsil eder. Bu testler açık yollarla ikiden fazla grubun olduğu durumlara kolayca genişletilebilir. Tek fark çoklu gruplarla sapmalara ilişkin *t* testi, sapmalarla ilgili varyans analizi ile yer değiştirecektir. Levene testinin ikisinden daha zayıf olduğunu önerecek ipuçları vardır, ancak Levene testi birçok istatistiksel yazılım tarafından rapor edilmektedir. Wilcox (1987b) bu testin daha tutucu göründüğünü belirtir.

Verinizde normallik sayıltısını ve varyans heterojenliğinin olmasını göz ardı etmek istemiyorsanız, problemi ele almanın farklı yolları vardır. Öncelikle alternatif olarak *F* testini, ardından veri dönüşümlerini değerlendireceğiz.

Yıllar önce Box (1954a), eşit olmayan varyanslarla elde edilen *F* dağılımını $F_{\text{elde edilen}}$ ile karşılaştırmak için serbestlik derecesi düzeltilmiş normal *F* ile aynı olduğunu göstermiştir. Eğer *F*'nin gerçek kritik değerini (varyans için düzeltilmiş) $F'_a$ ile tanımlarsak Box aşağıdaki eşitsizliklerin doğruluğunu ispatlamıştır.

$$F_\alpha(1, n - 1) \geq F'_\alpha \geq F_\alpha[k - 1, k(n - 1)]$$

Başka bir ifadeyle, *F*'nin gerçek kritik değeri, *F*'nin 1 ve (n – 1) serbestlik derecesindeki kritik değeri ile *F*'nin $(k - 1)$ ve $k(n - 1)$ serbestlik derecesindeki kritik değer arasında bir yerde yer alır. Sonraki sınır normallik ve varyansların homojenliği sayıltıları karşılandığında kullandığımız kritik değerdir. Box, $F_{\text{elde edilen}}$'i $F_\alpha(1, n - 1)$ ile karşılaştırarak daha tutucu bir test tavsiye etmiştir. Eğer bu manidar bir sonuca ulaşırsa, ortalamalar, varyansların eşitliğine veya eşitsizliğine bakılmadan manidar derecede farklılaşır. Box'ın yaklaşımı ile ilgili tek güçlük çok tutucu olmasıdır. Welch (1951) tarafından, bir sonraki bölümde değerlendireceğimiz ve kullandığımız birçok istatistiksel yazılım tarafından uygulanan, farklı bir yaklaşım öne sürülmüştür.

Wilcox (1987b) pratikte, varyansların sıklıkla dört faktörden daha fazlasıyla farklılaştığını tartışmıştır. Bu dört faktör, varyans heterojonliği hakkında mantıklı limit olarak düşünülür. Özellikle örneklem büyüklükleri eşit olmadığında Welch'in farklı varyanslara sahip örneklemlerle ilgili işlemlerini önermektedir.

## Welch Yöntemi

Kohr ve Games (1974), Keselman, Games ve Rogan (1979) ve Tomarken ve Serlin (1986) heterojen varyansa sahip örneklem gruplarına alternatif yaklaşımları incelemiş ve Welch

(1951) tarafından ileri sürülen işlemlerin normal evrenden çekilen örnekleme yapıldığında hem güçlü hem de I. Tip hataya karşı korumalı olduğunu göstermiştir. Formül ve hesaplamalar biraz biçimsizdir ancak çok zor değildir ve ne zaman varyans heterejonliğinden şüphelenseniz özellikle eşit olmayan örneklem büyüklüğüne sahip olduğunuzda bu testi kullanabilirsiniz. SPSS ve SAS Welch testini bir seçenek olarak sunmaktadır.

Tanım,

$$w_k = \frac{n_k}{s_k^2}$$

$$\overline{X}_{.}' = \frac{\sum w_k \overline{X}_k}{\sum w_k}$$

Buradan

$$F'' = \frac{\dfrac{\sum w_k(\overline{X}_k - \overline{X}_{.}')^2}{k-1}}{1 + \dfrac{2(k-2)}{k^2-1}\sum\left(\dfrac{1}{n_k - 1}\right)\left(1 - \dfrac{w_k}{\sum w_k}\right)^2}$$

olup bu istatistik (F″), $(k - 1)$ ve

$$sd' = \frac{k^2 - 1}{3\sum\left(\dfrac{1}{n_k - 1}\right)\left(1 - \dfrac{w_k}{\sum w_k}\right)^2}$$

serbestlik derecesinde hemen hemen $F$ gibi dağılmıştır.

Bu formüllerin düzensiz olduğu açıktır, fakat kullanılması imkânsız değildir. Öncelikle tüm terimleri (örneğin, $w_k$) toplarsanız, ardından problemler hakkında sistematik olarak çalışırsanız, hiçbir güçlük yaşamayacaksınız. (Bu gibi formülleri herhangi bir hesap tablosu programı kullanarak kolay bir şekilde uygulayabilirsiniz.) Sadece iki grubunuz olduğunda, Bölüm 7'de yer alan yaklaşımı (aynı zamanda Welch'e dayandırarak) kullanıp heterojen varyanslar için $t$ testinden yardım almak muhtemelen daha kolaydır. Örneğimizde, 4 ve 25,43 serbestlik derecesinde $F'' = 13{,}217$'dir. Sıfır hipotezi altında bu sonucun olasılığı 0,000006'dır. Bu hesaplamaları yapmak için R'de yazılan program kitabın Web sayfasında Welch.R ismi ile mevcuttur.

## FAKAT!

Heterojen varyanslarla nasıl başa çıkıldığını gösterdikten sonra, grup ortalamaları üzerinde varyans analizi testi yapmak homojenlik sayıtlısına uyulamamasına karşı güçlüdür, Bölüm 7'de söylediğim bir noktayı tekrarlamam gerekir. Welch gibi testlerimizin olması gerçeği bizi heterojen varyanslardan uzaklaştırmaz. Sadece ortalamalarla ilgili varyans analizini korur. Varyansın heterojenliğinin kendisi akla uygun bir bulgudur. Bu özel durumda, bilişsel/davranışsal terapinin alışılmadık şekilde etkili olduğu bir grup insanın, bu grupta biraz daha çift durumlu kazanımlara neden olduğu görünür. Bu önemli bir noktadır. Fakat grubun geri kalanları için, terapi en azından mâkul bir şekilde etkilidir. 10 pound'tan daha büyük kilo aldıran veriyi kırpsaydık böylece bilişsel/davranışsal terapi altında nadiren puanlanan bu katılımcıların kaldırılması ile $F$ sonucu manidar olacaktı ($F(2,52) = 4{,}71$, $p < 0{,}05$). Bu sonuçların tanımı, en azından bilişsel/davranışsal terapi koşullarındaki büyük varyansın bahsi olmaksızın eksik olacaktı. Sıklıkla heterojen varyanslar, burada olduğu gibi, bir tedavinin bazıları için etkili, diğerleri için oldukça etkisiz olduğu durumları yansıtırlar.

## 11.9 Dönüştürmeler

Bir önceki bölümde varyansın heterojenliği problemine bir yaklaşım olarak heterojen veri kümelerinde $F''$ değerini hesapladık ve uygun serbestlik derecesindeki genel $F$ dağılımına göre değerlendirdik. Bu işlemin, örneklemler normal dağılıma sahip evrenden çekildiğinde iyi çalıştığını gösterdik. Fakat dağılımı normal olmayan evrenlerde davranışı hakkında çok az bilgi vardır. Alternatif yaklaşım ise veri kümesini homojen varyanslar üreten formlara dönüştürmek, ardından dönüştürülmüş değerler üzerinden standart varyans analizini uygulamaktır. Stres çalışmasındaki semptom puanlarını kullanarak Ünite 9'da benzer bir çalışma yapmıştık.

Birçok insan veriyi dönüştürme fikrini kabul etmekte zorlanmaktadır. Elinizdeki veriyi beğenmediğinize karar vermek ve böylece onu daha iyi veya daha kötü veriye dönüştürmek, bazılarını çalışmadan çıkarmak, onları hiç toplamamış gibi yapmak dürüst olmayan bir durum olarak görülmektedir. Bu konu hakkında düşündüğünüzde, veriyi dönüştürme ile ilgili olağan dışı hiçbir şey yoktur. Sıklıkla veriyi dönüştürürüz. Bazen bir farenin geçite gitme *süresini* ölçeriz, ancak ardından koşma *hızları* arasındaki farka bakarız (doğrusal olmayan dönüştürme). Fiziksel enerji ile ilgili sesi ölçeriz, ardından onları desibel türünden belirtiriz (logaritmik dönüştürme). Denekten, benzer uyarıcıların boyutlarını karşılaştırmak için test uyarıcısının boyutunu ayarlamasını isteriz ve ardından bağımsız değişkenimiz olarak test-düzeltme ayarlarının yarıçapını alırız, fakat *yarıçap,* düzeltme *alan*ının karekökünün fonksiyonudur ve bu alanı doğal olarak bağımlı değişkenimiz gibi kullanabiliriz. Bazı testlerde öğrencilerin doğru cevapladığı maddelerin sayısını hesaplarız fakat ardından yüzdelikli olarak rapor ederiz (doğrusal olmayan dönüştürme). Hızın zamandan daha iyi bir ölçme olduğunu, desibelin enerji düzeyinden, çevrenin alandan veya yüzdeliğin doğru sayısından daha iyi olduğunu kim söyleyebilir ki? Conti ve Musty'nin (1984), farelerin psikomotor hareketliliğine THC'nin (marihuanadaki en psikoaktif madde) etkisini araştırdığı çalışmayı ele alalım. Conti ve Musty hareketliliği voltaj türünden gösteren dönüştürücüden kafesin hareketliliğini okuyarak etkinliği ölçmüştür. Test odası titreşiminin dönüştürülmüş ölçümü, elektriksel olarak bir hareketliliğin doğal ölçümü olarak adlandırılır? Daha önemlisi, enjeksiyon sonrası işlemini enjeksiyon öncesi işlemin bir yüzdesi olarak, bağımlı değişken gibi almışlardır, fakat dönüştürme işlemi yaptıkları için koltuğunuzdan fırlayıp "hata" diye ağlar mıydınız? Elbette hayır, fakat her şeye rağmen dönüşüm gerçekleşmiştir.

Giancola ve Corman'ın saldırganlık ve alkol ile ilgili çalışmasına geri dönelim: Örnekte kısmen küçük örneklemlere sahip olmamıza rağmen D8 koşulundaki varyans D4 koşulundaki varyanstan 20 kattan daha büyüktür. Levene testi istatistiği 4 ve 55 serbestlik derecesinde 3,69'dur. Bu, varyans heterojenliği için güçlü bir kanıttır. Heterojen varyansları da hesaba katan Welch testinin sıfır hipotezini reddettiğini gördük. Fakat alternatif yaklaşımla veri dönüştürülünce varyanslar artık heterojen olmayacaktı. Giancola ve Corman'ın durumunda, bağımlı değişken ilk olarak hem yoğunluk hem de süre puanlarını $z$ puanına çevirip ardından toplayarak oluşturuldu. Bu mâkul bir yaklaşım gibi görünüyor, ancak bence ölçmede doğal hiçbir şeyin olmadığını kabul etmeniz gerekiyor. Logartiması veya karekökü sadece doğal gibi görünüyor. Tukey muhtemelen bu hesaplamaları "dönüştürme"den çok "yeniden ifade etme" olarak adlandırmakta haklıydı. Sadece verinin başka terimlerle söylediğini yeniden ifade ediyoruz.

Kitabın ilk bölümlerinde bahsettiğim gibi, bağımlı değişkenlerimiz çalışmayı arzu ettiğimiz temel değişkenlerin sadece uygun ve eksik göstergeleridir. Herhangi bir duyarlı araştırmacı "stresli yaşam olaylarının sayısı" konusu gibi ciddî bir araştırmaya katılımının belirttiği "stresli olayların sayısı" olarak işe koyulmadı. Bu tür deneylerin her zaman amacı *stres*i incelemek olmuştur ve bildirilen olay sayısı sadece *stresin* uygun ölçümüdür. Aslında stres, büyük olasılıkla olayların sayısıyla doğrusal bir şekilde değişmez, bunun üssel olarak değişmesi oldukça mümkündür – uzun adımlarda birkaç stresli olay alabilirsiniz, ancak bir kez tabağınızda birkaç tane bulunursa, sonrakiler çok daha büyük etkiye

sahip olmaya başlarlar. Bu doğruysa, bazı güçlerle yükseltilmiş olayların sayısı –örneğin, $Y$ = (olay sayısı)$^2$ – daha uygun bir değişken olabilir.

Bunları söyledikten sonra, dönüştürülmüş verinin kullanıldığı sonuçların her zaman orijinal ölçümlere düzgünce aktarılmadığının farkına varılması önemlidir. Grissom (2000), aslında dönüştürülmüş değişkenlerin anlamlarının bazen, orijinal değişkenlerin anlamlarını tersine çevirebileceğini bildiriyor. Bu rahatsız edicidir ve ne yaptığının anlamı hakkında düşünmek önemlidir; ama sadece bu değildir, kendi içinde, bir neden dönüşümlerin kullanılmasını reddeder.

Bunu kabul etmeye istekliyseniz, bir dizi ölçümü başka bir dizi ölçüme dönüştürmeye izin verilir –örneğin; $Y_i = \log(X_i)$ veya $Y_i = \sqrt{X_i}$– daha sonra birçok olasılık, istatistiksel testlerin altında yatan sayıltılara daha yakından uyacak şekilde verimizde değişiklik yapmak için kullanılabilir hale gelir. Bu dönüştürmelerin çoğu hakkındaki güzel bir şey, bir sayıltıyı karşılamak için veriyi dönüştürdüğümüzde, diğer varsayımları karşılamamız da çoğunlukla yakınlaşmaktadır. Bu durumda, bir karekök dönüşümü sadece grup varyanslarının eşitlenmesine yardım etmeyebilir, çünkü bu dağılımın üst ucunu dağılımın alt ucundan daha fazla sıkıştırır, aynı zamanda pozitif çarpık dağılımları normale daha yakın yapma etkisi olabilir.

Dönüştürülmüş veri kümesinin raporlanması hakkında bir ifade vardır: Dönüştürülmüş değerler üzerindeki bir istatistik testini, varyans analizi gibi çalıştırmak meşru ve doğru olmasına rağmen, dönüştürülmemiş ölçek birimlerindeki anlamları genellikle rapor ederiz. Orijinal birimler, özünde anlamlı olduğunda bu özellikle doğrudur. Bununla birlikte, analizin dönüştürülmüş veri kümesi üzerinde yürütüldüğünün okuyucuya bildirilmesi gerekir.

Bir örnek, farklı takımlardan beyzbol oyuncularının maaşları olsun. Maaş miktarlarıyla çalışan kişiler hep log(maaş) üzerinde kendi analizlerini gerçekleştirirler. Ancak, log(maaş) çoğumuz için anlamlı bir ölçüm değildir. Daha iyi bir yaklaşım, bütün veriyi logaritmaya (bir logaritmik dönüştürme kullanmayı seçtiğinizi varsayarsak) dönüştürmek; bu log değerlerinin ortalamasını bulmak ve sonra antilog almak olacaktır, bunun anlamı orijinal birimlere geri dönmektir. (Alternatif olarak, maaş gibi değişkenler için sıklıkla tercih ölçüsü olan medyan değerini de raporlayabilirsiniz.) Bu dönüştürülmüş ortalama büyük olasılıkla orijinal değerlerin ortalamasına eşit olmayacaktır, ancak bu dönüştürülmüş ortalamalar rapor edilmelidir. Fakat ben hem dönüştürülmüş hem de dönüştürülmemiş ortalamalara bakılmasını ve onların aynı temel hikâyeyi anlattıklarından emin olunmasını teşvik ediyorum. Standart sapmayı dönüştürmeyin – eğer bunu denerseniz ciddî haksızlık yaparsınız. Ve ne yaptığınızı okuyucularınıza göstermek için emin olunuz.

Bu bölümde sadece en yaygın dönüştürmeleri göz önünde bulunduruyoruz, çünkü sizin için en yararlı olacak olanlar bunlardır. Dönüştürmelere ilişkin tüm yaklaşımların iyi bir tartışması Tukey (1977), Hoaglin, Mosteller ve Tukey (1983) ve Grissom'da (2000) bulunabilir.

## Logaritmik Dönüştürmeler

Logaritmik dönüştürmeler, standart sapma, ortalamayla orantılı olduğunda yararlıdır. Aynı zamanda veri belirgin şekilde pozitif çarpık olduğunda da yararlıdır. Bu ifadelerin her ikisinin de niçin doğru olduğunu anlamanın en kolay yolu logaritmanın ne olduğunu hatırlamaktır. [Logaritmanın güç olduğunu hatırlayın – $\log_{10}(25)$, 10'nun 25'i verecek şekilde yükseltilmesi gereken üssüdür. Bu nedenle, $\log_{10}(25) = 1{,}39794$'tür, çünkü $10^{1{,}39794} = 25$'tir. Karşıt olarak, 1,39794'ün anti logaritması $10^{1{,}39794} = 25$'tir. İstatistikte genellikle "doğal log" olarak adlandırılan, $e = 2{,}71828$ tabanına göre logaritmayı kullanırız, bu durumda $\log_e(25) = \ln(25) = 3{,}21887$, $e^{3{,}21887} = 25$ olur. Tutarlı olduğunuz sürece hangi tabanı kullandığınızın önemi yoktur.] Eğer 10, 100 ve 1000 sayılarını alırsak, bunların $\log_{10}$'ları 1, 2 ve 3'tür. Bu durumda log birimi olarak 10 ile 100 arasındaki uzaklık, 100 ile 1000 arasındaki uzaklığa

eşittir. Başka bir ifadeyle, logaritma alarak dağılımın sağ tarafı (daha pozitif değerler) sol tarafından daha fazla sıkıştırılmış olacaktır. (Bu beyzbol oyuncularının maaşlarının iyi bir örnek olarak verilmesinin nedenidir.) Bu sadece "pozitif çarpık dağılımların logaritmik dönüşümler altında simetrikleşme eğiliminde olması" anlamında değildir. Aynı zamanda, görece büyük sayılardan oluşan bir kümenin standart sapmasının, küçük sayılardan oluşan bir kümenin standart sapmasından büyük olması durumunda, logaritma almanın, büyük sayılardan oluşan örneklemin standart sapmasını, küçük sayılardan oluşan örneklemin standart sapmasına göre daha çok düşüreceği anlamındadır.

Tablo 11.6, Conti ve Musty (1984) tarafından yapılan esrardaki aktif madde olan THC uygulamasının ardından farelerdeki etkinlik düzeyi çalışmasındaki örneği içerir. (Enjek-

**Tablo 11.6** Conti ve Musty'nin (1984) orjinal ve dönüştürülmüş veri kümeleri

**(a) Orijinal Veril**

| | Kontrol | 0.1 $\mu$g | 0.5 $\mu$g | 1 $\mu$g | 2 $\mu$g | |
|---|---|---|---|---|---|---|
| | 130 | 93 | 510 | 229 | 144 | |
| | 94 | 444 | 416 | 475 | 111 | |
| | 225 | 403 | 154 | 348 | 217 | |
| | 105 | 192 | 636 | 276 | 200 | |
| | 92 | 67 | 396 | 167 | 84 | |
| | 190 | 170 | 451 | 151 | 99 | |
| | 32 | 77 | 376 | 107 | 44 | |
| | 64 | 353 | 192 | 235 | 84 | |
| | 69 | 365 | 384 | | 284 | |
| | 93 | 422 | | | 293 | |
| Ortalama | 109,40 | 258,60 | 390,56 | 248,50 | 156,00 | |
| | | | | | | $r = ,88$ |
| Std. Sapma | 58,50 | 153,32 | 147,68 | 118,74 | 87,65 | |
| Varyans | 3421,82 | 23,506,04 | 21,809,78 | 14,098,86 | 7682,22 | |

**(b) Log Veri**

| | Kontrol | 0.1 $\mu$g | 0.5 $\mu$g | 1 $\mu$g | 2 $\mu$g | |
|---|---|---|---|---|---|---|
| | 2,11 | 1,97 | 2,71 | 2,36 | 2,16 | |
| | 1,97 | 2,65 | 2,62 | 2,68 | 2,04 | |
| | 2,35 | 2,60 | 2,19 | 2,54 | 2,34 | |
| | 2,02 | 2,28 | 2,80 | 2,44 | 2,30 | |
| | 1,96 | 1,83 | 2,60 | 2,22 | 1,92 | |
| | 2,28 | 2,23 | 2,65 | 2,18 | 2,00 | |
| | 1,50 | 1,89 | 2,58 | 2,03 | 1,64 | |
| | 1,81 | 2,55 | 2,28 | 2,37 | 1,92 | |
| | 1,84 | 2,56 | 2,58 | | 2,45 | |
| | 1,97 | 2,62 | | | 2,47 | |
| Ortalama | 1,981 | 2,318 | 2,557 | 2,353 | 2,124 | |
| | | | | | | $r = -,33$ |
| Std. Sapma | 0,241 | 0,324 | 0,197 | 0,208 | 0,268 | |
| Varyans | 0,058 | 0,105 | 0,039 | 0,043 | 0,072 | |

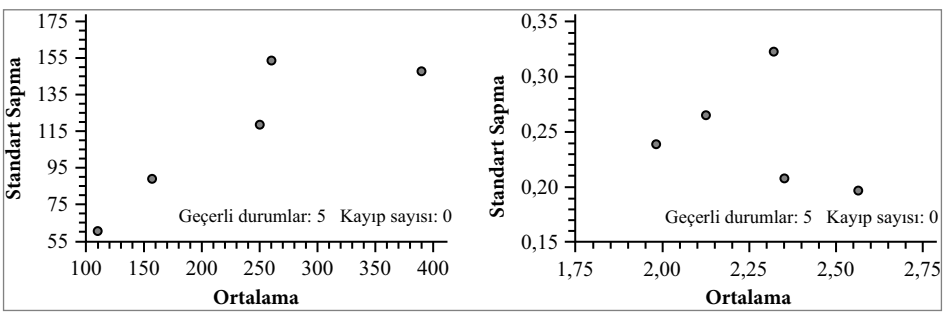
© Cengage Learning 2013

siyon sonrası 10 dakikalık süre boyunca her bir hayvan için, etkinlik birimleri -keyfi bir ölçek üzerinde- raporlanmıştır, oysa Conti ve Musty, başlangıç etkinlik yüzdesi olarak enjeksiyon sonrası etkinliği raporlamıştır.) Tablo 11.6a'daki veri kümesinden varyansların eşit olmadığını görebilirsiniz. En büyük varyans en küçüğünün yaklaşık yedi katıdır. Bu kısmen, ilâçların, ortalamanın yanı sıra değişkenliği arttırma eğilimi gerçeğine rağmen, iyi kurulmuş bir fonksiyondur. Sadece varyansların eşit olmaması değil, ayrıca standart sapmaların ortalamalarla orantılı olduğu görülmektedir. Bu ordinat üzerinde standart sapmaların, apsis üzerinde ortalamaların çizildiği Şekil 11.5a'da kolaylıkla görülmektedir. Bu iki istatistik arasında doğrusal bir ilişkinin olduğu açıkça görülmektedir ($r$ = ,88). Bu doğrusallık logaritmik dönüştürmenin yararlı olabileceğini akla getirir. Tablo 11.6b'de veri kümesi 10 tabanına göre logaritmaya dönüştürülmüştür. (Herhangi bir taban kullanabilir ve aynı etkiye sahip olabilirdim. 10 tabanını seçtim çünkü bunun aşinalığı daha fazladır, gerçi çoğu istatistiksel çalışmalarda, teknik nedenlerden dolayı $e$ tabanına göre log ($\log_e$ veya ln) tercih edilmiştir.) Şekil 11.5b'den de görülebileceği gibi artık ortalama ve standart sapma ilişkili değildir (manidar olmayan $r$ = –,33). Ortalama ve standart sapma arasındaki orantılılığı kırdık ve şimdi en büyük grup varyansı en küçük grup varyansının üç katından daha azdır.

Varyans analizi artık bu dönüştürülmüş veri kümesiyle yapılabilir. Bu durumda Levene testi, 4 ve 42 *sd*'de ,204 olasılıkla, manidar olmayan 1,557 $F$ değeri üretir. Bu, varyans heterojenliğinin artık bir sorun olmadığını doğrulamaya yardımcı olur. Varyans analizi için $F(4,42)$ = 7,2, açıkça manidardır. (Fark ayrıca Welch testiyle de manidardır.) Conti ve Musty, daha önce söylediğim gibi, hem teorik nedenlerle ve hem de bu araştırmanın kendi alanında standart bir uygulama olmasından, oran ölçmelere ilişkin varyans analizini yapmayı seçti. Bununla birlikte, özellikle ya dağılımların şekilleri ya da varyans heterojenliğiyle ilgili sorunlar oluşursa, gelecekteki analizler için orijinal birimlerin logaritmik dönüştürmelerinin daha uygun olabileceği bir örnek yapılmış olur.

Daha önce de belirttiğim gibi, logaritmik dönüştürme için kullandığınız taban fark etmez ve çoğu istatistikçi $\log_e$ kullanma eğilimdedir. Ancak, taban ne olursa olsun orijinal değerler ($X_i$) negatif ya da sıfıra yakın olduğunda problemler vardır, çünkü log sadece pozitif sayılar için tanımlanmıştır. Bu durumda log almadan önce tüm $X$ değerlerini pozitif hale getirmek için bir sabit eklemeniz gerekir. Genel olarak, sıfıra yakın değerlere sahip olunduğunda, log ($X_i$) yerine log ($X_i$ +1) kullanmalısınız. Eğer sayıların kendileri -1'den daha küçük ise, hepsini sıfırdan daha büyük yapmak için gerekli sabit ne olursa olsun ekleyin.

Logaritmik dönüştürmenin kullanım alanlarından biri sayma sayılarından oluşmuş veri kümeleriyle ilgilenmesidir. Örneğin, sınıftaki otistik çocuklarla ilgilenebilirsiniz ve bağım-




a) Ham veri grafiği b) Log-dönüştürülmüş veri grafiği

**Şekil 11.5** Tablo 11.6'daki orijinal ve dönüştürülmüş veri kümeleri için ortalamalar ve standart sapmalar arasındaki ilişki

© Cengage Learning 2013

lı değişkeniniz bu çocukların 20 dakikalık süre içerisinde sergiledikleri uygunsuz davranışların sayısı olabilir. Bunlar 0 ve 5 arasındaki sayılar olabilir. Hiçbir sayma sayısı negatif olmadığından bu verinin muhtemelen normal dağılmayacağını biliyorsunuz. Sayma veri kümeleri genellikle Poisson dağılımına uyar ve standart yaklaşım log(sayma sayısı)'nın, sayma sayısının kendisi yerine modellenmesidir.

## Karekök Dönüştürmesi

Ortalama, standart sapma yerine varyansla orantılı olduğunda karekök dönüştürmeleri kullanılır. Bu durumda, varyansı sabitlemek ve çarpıklığı azaltmak için $Y = \sqrt{X}$ kullanışlıdır. Eğer $X$'in değeri oldukça küçükse (10'dan daha küçük), sonra varyansı sabitlemek için $Y = \sqrt{X + 0,5}$ veya $Y = \sqrt{X} + \sqrt{X + 1}$ genellikle daha iyidir. Problem, eğer varyans ile korelasyon çok daha büyük veya standart sapmayla korelasyon daha küçükse, bunu sizin veya benim bilmemin genellikle çok zor olmasıdır. Bu yüzden ikisini denemek iyi fikirdir. Conti ve Musty verisi için, ortalama, standart sapmayla olduğu gibi varyansla da yaklaşık olarak iyi korelasyon verir. Eğer değer aralığı büyük değilse standart sapmaların ve varyansların kendileri de büyük ölçüde ilişkilidir (bu durumda $r_{s.s^2} = ,99$).

## Ters Dönüştürme

Eğer pozitif kuyrukta çok büyük değerlere sahip bir dağılım varsa, bu uç değerlerin etkisini, ters dönüştürmeler, logaritmik dönüştürmelerden belirgin olarak daha iyi azaltabilir. Örneğin, labirentte ya da düz sokakta hayvanlar genellikle kendi işlerini unuturlar ve tüm fotosellerdeki koklamayı durdururlar ve yollarını öyle bulurlar. Bir hayvan 30 saniye boyunca aparata sokulduktan sonra, koşuyu tamamlamak için başka bir 300 saniye alırsa, bu bizim için önemli değildir. Bölüm 2'de başvurulan bir yaklaşım; günde birkaç deneme varsa, günlük ortanca zamanı ölçüm olarak alabilirsiniz. Alternatif bir yaklaşım, tüm veriyi zamanın tersini alarak kullanmaktır (örneğin, hız), çünkü bunun neredeyse uzun zamanları eşitleme etkisi vardır. Aşağıdaki süreleri elde ettiğimizi düşünelim,

[10, 11, 13, 14, 15, 45, 450]

Bu sürelerin tersleri şu şekildedir:

[0,100; 0,091; 0,077; 0,071; 0,067; 0,022; 0,002]

Uzun süreler arasındaki farklılıkların orijinal birimlerindekinden çok daha düşük olduğuna dikkat ediniz. Oysa normal dağılıma sahip veri kümelerinde çarpıklık = 0'dır, dönüştürülmemş veri kümesinin çarpıklığı 2,62'dir, ikinci küme için ise bu -0,91'e düşürülmüştür. Ayrıca uç değerlerin, standart sapmanın büyüklüğü üzerindeki etkisi, dönüştürmeden önceki etkisinden oldukça az olacaktır. Reaksiyon sürelerine ters dönüştürme yapıldığında benzer türde etkiler bulunur, burada uzun reaksiyon sürelerinin bilgi işleme hızları hakkında gösterdikleri, onların yaptıkları hakkında gösterdiklerinden muhtemelen daha azdır, aslında denek anlık dikkat etmemiş veya vurması gereken tepki anahtarını kaçırmıştır.

## Arcsin Dönüştürmesi

Bölüm 5'te binom dağılımı için, $\mu = Np$ *ve* $\sigma^2 = Npq$ olduğunu görmüştük. Bu durumda ortalama ve varyansın her ikisi $p$'ye bağlı olduğu için, varyans, ortalamanın doğrudan fonksiyonu olacaktır. Bazı deneyler için bağımlı değişkenin, her denek tarafından doğru olarak

hatırlanan maddelerin oranı olduğunu varsayalım. Sonra her madde doğru olma olasılığı $p$ olan Bernoulli denemesi (yanlış olma olasılığı 1 - $p$) ve madde takımının tamamı Bernoulli denemelerinin serisi olarak düşünülebilir. Başka bir ifadeyle, varyansın ortalamaya bağlı olduğu yerde, sonuçlar binom dağılımına sahiptir. Eğer bu doğruysa, farklı ortalamalardaki gruplar mutlaka farklı varyanslara sahip olacaktır ve bu bizim bir problemimiz olacaktır. Bu durum için arcsin dönüştürmesi genellikle yararlı olur. Bu dönüştürmenin genel formu $Y = 2 \arcsin\sqrt{p}$'dir. Bu durumda $p$ doğru oranıdır ve $Y$ de $p$'nin karekökününe ait sinüs açısının iki katı olacaktır.[5] Arcsin dönüştürmesi, radyan ile çalışmasını ayarladığınız birçok hesap makinesiyle elde edilebilir ($\sin^{-1}$ olarak etiketlenmiştir). ("arcsin(,50)" için bir Google araması yapıp sonucu 0,523598776 olarak almak, buna göre daha kolaydır. Google'ın arkadaşınız olduğunu söylemiştim).

Varyans ortalamayla orantılı olduğunda, karekök ve arcsin dönüştürmelerinin her ikisi de uygundur. Ancak aralarında bir fark vardır. Karekök dönüştürme, dağılımın üst kuyruğunu sıkıştırırken, arcsin dönüştürme ortaya göre her iki kuyruğu dışa uzatır. Eğer oranlarla ilgileniyorsanız, normalde arcsin dönüştürme daha yararlıdır.

## Kırpılmış Örneklemler

**aşırı kuyruklu dağılımlar**

Varyansların homojenliğini veya normalliğini elde etmek için her bir ham puanı dönüştürmek yerine, **aşırı kuyruklu dağılımlar** (kuyruklarında alışılmadık sayıda gözlem olan nispeten düz dağılımlar ) ile bir alternatif yaklaşım kırpılmış örneklemleri kullanmaktır. Bölüm 2'de *kırpılmış örneklem*, her bir kuyruğundan uç değerlerinin sabit bir yüzdesi çıkarılmış bir örneklem olarak tanımlanmıştır. Böylece 40 vaka ile bir %5'lik kırpılmış örneklem, her kuyruğundan çıkarılan iki gözlemden oluşan örneklem olacaktır. Varyans analizinde olduğu gibi, çeşitli gruplar karşılaştırılırken her bir örneklem aynı miktarda kırpılacaktır. Kırpılmış örneklemler çok uzun bir süredir istatistikçilerin etrafında olmasına rağmen, son zamanlarda bunlara ilgi artmıştır, çünkü bunların nadir uç değerleri olan dağılımlarla ilgilenildiği zaman faydalılığı dikkat çekmeye başlamıştır. Muhtemelen gelecekte bunları daha fazla göreceksiniz.

**pencerelenmiş (winsorized) örneklemler**

Kırpılmış örneklemlerle yakından ilgili bir örneklem **pencerelenmiş (winsorized) örneklemler**dir. Bu örneklemede, kırpmadan sonra en uçta kalan değer, kırpılmış değerler yerine yazılır. %10'luk bir durumda,

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 7 | 12 | 15 | 17 | 17 | 18 | 19 | 19 | 19 |
| 20 | 22 | 24 | 26 | 30 | 32 | 32 | 33 | 36 | 50 |

en düşük iki değer (3 ve 7) yerine 12, en yüksek iki değer (36 ve 50) yerine 33 yazılacaktır.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 15 | 17 | 17 | 18 | 19 | 19 | 19 |
| 20 | 22 | 24 | 26 | 30 | 32 | 32 | 33 | 33 | 33 |

[Bu örneklem üzerinden hesaplanacak varyans veya herhangi bir test istatistiği ($N - 1 - 4$) *sd'*de olacaktır, çünkü dört değer kırpıldı ve bunlar sözde değerlerle değiştirildi. Bu sözde veriye gerçek veri gibi davranmak gerçekten mâkul değildir.] Alışılmadık sayıda uç değerler içeren örneklemlerde, kırpılma veya pencerelemekten faydalanılabilir. Kırpılmış veri üzerinden varyans analizi yaptığınızda, kırpılmış örneklemin varyansına değil, buna karşı gelen pencerelenmiş örneklemin varyansına ait $KO_{\text{hata}}$ kullanılmalıdır. Kırpılmış ör-

5 arcsin dönüştürmesi, bu özelliğinden dolayı, çoğu zaman "açısal" dönüştürme olarak anılır. $p$, 0'a veya 1'e yakın olduğunda genellikle $2\arcsin = \sqrt{p \pm \frac{1}{2n}}$ alınır, burada $p$ 0'a yakınsa artı, 1'e yakınsa eksi işareti kullanılır.

neklemde uygulanan *t* testinin etkililiğine (ve bunun genişletilmişi olan varyans analizine) dair okunabilir bir çalışma Yuen ve Dixon (1973) tarafından yapılmıştır; bu tür analizleri yapmadan önce okumalısınız. Ayrıca Wilcox (1993 ve 1995) tarafından yazılan raporlara da bakmalısınız. Wilcox uzun süre kırpılmış örneklemlerin savunucusu olmuştur. Bölüm 12'de çoklu karşılaştırmalara gelindiğinde Keselman, Holland ve Cribbie (2005, s. 1918-1919) yararlı bir referans olacaktır.

## Ne Zaman Dönüştürme Yapılır ve Dönüştürme Nasıl Seçilir?

Dönüştürmelerin, bütün veri kümelerine rutin olarak uygulanması gerektiği izlenimi edinilmemelidir. Genel bir kural olarak, "kırık değilse, bunu düzeltmeyin". Veriniz mâkul bir şekilde dağılıyorsa (yani, aşağı yukarı simetrikse ve uç değer yoksa veya sadece birkaç uç değer varsa) ve varyanslar homojense, bir dönüştürme yapılarak kazanılacak bir avantaj muhtemelen yoktur. Bununla birlikte, eğer veriniz belirgin olarak çarpıksa ve varyanslar heterojense, bazı dönüştürme türleri yararlı olabilir. Ayrıca, varyans veya şekil için gerekli değişiklikleri yapan bir dönüştürme için araştırma yapmak mantıklıdır. Logaritmik dönüştürme istediğinizi yapmazsa (varyansı dengelemek, şekli iyileştirmek), karekök (veya küp kök) dönüştürmeyi düşününüz. Eğer sıfıra yakın değerleriniz varsa $Y = \sqrt{X + 0,5}$ çalışmazsa, $Y = \sqrt{X} + \sqrt{X + 1}$'i deneyiniz. *Yapmamanız* gereken tek şey bütün dönüştürmeleri denemeniz, anlamlı gelen bir tanesine bakmanızdır. (Elde edilen *F* için değil, veriyi iyi duruma getirmek için çalışıyorsunuz.) İstatistiksel teste başlamadan önce dönüştürmenizi seçiniz. Son olarak, dönüştürmeleri kullanmayı düşünüyorsanız, Tukey (1977), Hoaglin, Mosteller ve Tukey'ye (1983) bakmak iyi bir fikir olur.

Osborne'da (2008), dönüştürmeler ve bunların kullanımının mükemmel bir tartışması bulunmaktadır. O, en azından karekök, logaritmik ve ters dönüştürmeler yapıldığında, verinin minimum değeri 1,0 olduğunda dönüştürmenin etkisinin en büyük olacağını belirtmiştir. Sonuç olarak Osborne, bu dönüştürmelerin birini uygulamadan önce, genellikle veriye sabit ekleyerek ya da çıkararak veriyi minimum 1,0'a taşımak için ilk veri dönüştürmeyi önerir. Ayrıca, bu dönüştürmelerin en iyi pozitif çarpık veri kümelerinde işlediğini; fakat kümedeki bütün veriyi önce -1 ile çarptıktan sonra uygun bir sayıyı toplayıp çıkararak negatif çarpık veriye de uygulanabileceğine dikkat çekmiştir.

## Yeniden Örnekleme

**yeniden örnekleme istatistikleri**

Çok fazla destek kazanmaya başlayan, istatistiksel hipotez testinde eski ama çok değerli bir yaklaşım, **yeniden örnekleme istatistikleri** olarak bilinir. Bölüm 18'de bu yaklaşım hakkında önemli işlerden söz ettim, fakat sayıltıların ihlâli yöntemlerini kapatmadan önce, yeniden örnekleme yöntemlerinin varyans analizinde gerekli bazı sayıltıları önlemek için bir fırsat sunduğundan en azından bahsetmeliyiz. Bu yöntemler, esas olarak aynı sayıda gözlemlerle elde edilen veri kümelerinin birleştirilmiş dağılımlarına tam olarak benzeyen aynı sayıda gözlemlerle evren oluştururlar. Ardından bilgisayar, ölçümleri gruplara yerlerini değiştirerek tesadüfî olarak sanki sıfır hipotezi doğruymuş gibi yerleştirir[6] ve daha sonra test istatistiklerini hesaplar, *F* gibi. Bu süreç daha sonra, gerçek yokluk hipotezi ile beklenen *F* değerinin tam dağılımını üretmek için çok büyük sayıda tekrar edilir. Daha sonra kaç tane $F_{\text{örneklem}}$ değerinin sizin verinizden daha farklı olduğunu, sonuçlara bağlı olarak reddetmeyi ya da reddetmemeyi hesaplamak kolaydır. Bununla ilgilenen öğrenciler Bölüm 18'e atlayabilirler, aradaki bölümleri okumadan bunun anlaşılması zor değildir.

---

6 Diğer bir yol karma olabilir. Bireylerin isimlerini kartlara yazarsınız, desteyi kararasınız sonra ilk 10 kartı Grup 1'e ikinci 10 kartı grup 2'ye vb. atarsınız.

## 11.10 Sabite Karşı Tesadüfî Modeller

Bağımsız değişkenin düzeylerinin nasıl seçileceği hakkında henüz bir şey konuşmadık, sadece "işlemler" hakkında konuştuk. Aslında, bunu düşündüğünüzde, işlem değişkeninin düzeyi en az iki farklı şekilde elde edilir. Bunları bilinçli olarak seçebiliriz, genellikle böyle yaparız veya tesadüfî olarak örnekleyebiliriz. Çalışmamızdan çıkaracağımız genellemeler için düzeylerin etkileri ortaya çıkarılabilir.

**tekrarlamak**

Doğu Aurora'da New York altıncı sınıf çocuklarının o bölgedeki beşinci sınıf çocuklarından manidar olarak daha iyi okuyup okumadıklarının belirlenmeye çalışıldığını düşünelim. Doğu Aurora özellikle büyük bir köy değildir, bu yüzden beşinci ve altıncı sınıfın her birinden onar öğrenci çekilerek onlara test uygulanmıştır. Bu deneyi birkaç kez **tekrarlarsak**, her zaman aynı sınıflardaki farklı çocukları kullanırız, çünkü bunlar şehirdeki tek beşinci ve altıncı sınıftır. Yani bir tekrarlamadan diğerine, kullanılan sınıflardaki değişkenliğe atfedilebilecek bir değişkenlik yoktur. "Sınıf" sabit değişken olarak isimlendirilir çünkü özellikle sınıfları kullanmak için seçtik. Burada önemli olan nokta, sınıf düzeyleri aslında sabit anlamındadır, bunlar çalışmadaki bir tekrardan diğerine tesadüfî olarak değişmemektedir. Böyle bir deneyin analizi **sabit-model varyans analizi** olarak adlandırılır.

**sabit-model varyans analizi**

Fakat, bunun yerine, aynı sorunun New York eyaletinin bütünündeki okullar için sorulmuş olduğunu düşünelim. Muhtemelen okullardan bir grup tesadüfî olarak, bu okulların içinden de tesadüfî olarak sınıflar seçilecektir. Bu durumda birden fazla çoklu tekrarı düşündüğümüzde, biliyoruz ki sonuçlar sadece örnekleme hatasına göre değişmeyecek, aynı zamanda değişkenlik, kullandığımız sınıflara (ve okullara) bağlı olacaktır. Böylece hem sınıflar hem de okullar tesadüfî değişken olarak adlandırılacaktır. Burada sınıf düzeyleri tesadüfî sürecin sonucudur ve sınıf bakımından ilgilendiğimiz evren oldukça büyüktür. Çünkü, işlem düzeylerinden elde edilen süreç nedeniyle, işlemler tesadüfî değişken ve analiz de **tesadüfî-model varyans analizi** olarak isimlendirilir.

**tesadüfî-model varyans analizi**

Sabit ve tesadüfî modeller hakkında Bölüm 13 ve 14'te daha fazla söyleyeceğimiz şeyler vardır. Bunlar davranış bilimlerindeki araştırma analizlerinde geniş rol oynar ve bunları anlamaya ihtiyaç duyarsınız. Burada önemli olan nokta, bir sabit modelde işlem düzeyleri bilinçli seçilmiştir ve bir tekrardan diğerine sabit kalacaktır. Tesadüfî modelde, işlem düzeyleri tesadüfî süreçle elde edilmektedir ve tekrarlamalar arasında farklılık göstermesi beklenir. Tek yönlü varyans analizi için ayrım çok önemli değildir ancak sadece tesadüfî değişkenlerle uğraşmak zorunda olmadığımız, ama genellikle "iç içe/yuvalanmış (nested) değişkenler" olarak adlandırdığımız, daha kompleks desenleri kullandığımızda bu oldukça önemli olabilir. Daha kompleks modellerdeki bağımsız değişken olan tesadüfî değişken çoğu zaman kendi başına büyük önem taşımaz. Bu genellikle başta çalışmamızın genellenebilirliğini artırmak içindir. Ancak bunun varlığı elde edilen *F* değerlerini önemli ölçüde etkileyebilir.

## 11.11 Deneysel Etkinin Büyüklüğü

Varyans analizinin manidar bir *F* üretmesi gerçeği bize basitçe orada hataya atfedilemeyen işlemlerin ortalamaları arasında farklılıkların olduğunu söyler. Bu farklılıkların herhangi bir pratik önemi olup olmadığı hakkında bir şey söylemez. Bu nedenle, farkın "önemini" yansıtan ek ölçümleri tanımlamak için *F* değerinin ötesine bakmak gerekir. Daha önceki bölümlerde, ortalamalar arasındaki farklarla doğrudan ilişkili olan *d*-ailesi ölçümleri, bağımlı ve bağımsız değişkenler arasındaki korelasyona dayalı olan *r*-ailesi ölçümleri arasında bir ayrım yapılmıştı. Geniş kapsamlı *F*'yi dikkate aldığımızda, bütün ortalamalar

arasında herhangi bir fark ararız, *d*-ailesi ölçümleri uygun olabilir veya olmayabilir, buna rağmen *d*-ailesi ölçümlerini kısaca tartışacağız. Bölüm 12'deki bireysel karşılaştırmalar tartışıldığında onlar çok uygun olacaktır. *r*-ailesi ölçümleri genellikle bütün ortalamaların hepsinin birden testi için önerilir ve ilk odaklanılacak şey budur. Ancak itiraf etmek zorundayım ki *r*-ailesi ölçümlerini özellikle ilgi çekici bulmuyorum, çünkü bu ölçümler için büyük veya küçük olan değerin ne olduğunun bilinmesi zordur. Bazı durumlarda, değişkenliğin %5'inin açıklanması çok önemli olabilecekken, diğerlerinde %5 önemsiz olabilir. (Ancak hangi durumun içerisinde olduğumuzu nasıl bileceğiz?)

Bu baskıda genel analizdeki etki büyüklüğüne çekilmiş dikkati anlamlı bir şekilde azaltıyorum. Bunu yapıyorum, çünkü tüm ortalamalar arasında belirli karşılaştırmalar için dikkati etki büyüklüğüne odaklamak istiyorum. Belirli karşılaştırmalardaki etki büyüklüğü hesaplamaları bu etkilere dikkat çekmektedir ve daha az kullanışlı genel etkiden uzaktır.

Burada tartışılan ölçme takımları genellikle "etkinin büyüklüğü" ölçümleri şeklinde sınıflandırılır ve $r^2$ ile ilgilidir. Bunlar bağımlı değişkendeki genel değişkenliğin ne kadarının işlem etkisine atfedilebileceğini temsil ederler. Benim son sayıma göre orada **deneysel etkinin büyüklüğü**nün en az altı ölçümü vardır – tümü farklı ve çoğu diğer bazı ölçümlerden daha az yanlı olduğunu iddia eder. Bu kısımda en yaygın iki ölçüme ($\eta^2$ ve $\omega^2$) odaklanılacaktır çünkü onlar dikkatimizi çekecek en güçlü iddiaya sahiptir.

**deneysel etkinin büyüklüğü**

## Eta-Kare ($\eta^2$)

Eta–kare deneysel etkinin gücünün ölçümünün muhtemelen en eski ve en basit ölçümüdür. Varyans analizi özet tablosunu düşünürsek $KT_{\text{toplam}}$ tüm veri kümesindeki puanların birinden diğerine nasıl farklı olduğunun indeksidir. Ayrıca $KT_{\text{işlem}}$, gözlemler arasındaki farkın ne kadarının farklı işlemlere atfedilebileceğini gösterir. $KT_{\text{işlem}}$'in $KT_{\text{toplam}}$'a bölünmesindeki oran, işleme atfedilebilecek tüm değişkenliğin yüzdesini verecektir. Bu şu şekilde verilir:

$$\eta^2 = \frac{KT_{\text{işlem}}}{KT_{\text{toplam}}}$$

**eta-kare ($\eta^2$)**
**korelasyon oranı**

Bu istatistik **eta – kare ($\eta^2$)** olarak adlandırılır, buna bazen **korelasyon oranı** da denir. Tablo 11.2'deki Giancola ve Corman verisine $\eta^2$ uygulandığında,

$$\eta^2 = \frac{KT_{\text{işlem}}}{KT_{\text{toplam}}} = \frac{62,460}{186,918} =,334$$

elde edilir.

İki kareler toplamına erişemediğiniz durumlarda bir alternatif hesaplama şekli aşağıdaki gibidir:

$$\eta^2 = \frac{1}{1 + \frac{sd_{\text{hata}}}{Fxsd_{\text{işlem}}}}$$

Giancola ve Corman verisi için, saldırganlık puanlarındaki değişkenliğin %33,4'ü, dikkat dağınıklığı görevinin düzeylerindeki farklılıklara atfedilebilir ve böylece, muhtemelen deneklerin dikkatini kışkırtıcı uyaranlardan uzaklaştıran göreve odaklanma derecesidir. Bu, dikkat dağıtan görevin etkililiğini yansıtır, açıklanan değişkenliğin mâkul bir miktarıdır.

$\eta^2$'ye bakmanın bir başka yolu, bir önceki formülden doğrudan çıkarılır ve son zamanlarda etki büyüklüğünün herhangi bir ölçümünün arzu edilen bir özelliği olarak in-

**Hatada Yüzde Azaltma (HYA)**

celenmiştir. Judd ve McClelland (1989) tarafından $\eta^2$ **Hatada Yüzde Azaltma (Percent Redaction in Error (PRE))** olarak terminolojide popülerleştirilmiştir. Grup üyeliklerini dikkate almadığımızda, 60 gözlemin bütününün değişimi, hatırlama puanlarıyla ilişkili hata, $KT_{\text{toplam}}$ olurdu. Fakat bireylerin hangi grupta yer aldığı bilindiğinde, kestirimle ilgili hatamız $KT_{\text{hata}}$'ya düşer. Ancak $KT_{\text{toplam}}$ ve $KT_{\text{hata}}$ arasındaki fark $KT_{\text{işlem}}$'dir. Böylece $KT_{\text{işlem}}$'in $KT_{\text{toplam}}$'a bölümü, kestirim hatası grup üyeliği dikkate alınarak düşürülmüş yüzdedir. Bizim örneğimiz açısından, kestirimle ilgili hatamız, grup üyeliğine katılma olmadan 190,907 hata birimi ve grup üyeliğine katılma olduktan sonra 124,458 hata birimi bulunmuştur. Böylece, hatayı 186,918 - 124,458 = 62,460 veya 62,460 / 189,918 = %33,4 azaltmış oluruz.

$\eta^2$'nin gerçek regresyon doğrusunun bireysel işlem ortalamalarından geçtiğini varsaydığını fark etmek önemlidir. Veri kümesi evren olarak kabul edildiğinde sayıltı doğrudur. Ancak veri kümesi geniş bir evrenden alınmış örneklem olarak kabul edildiğinde, yanlılık oluşur. Çünkü bu ortalamalar gerçekten örneklem ortalamalarıdır, onlar örnekleme hatasına maruz kalırlar ve $\eta^2$ yukarı doğru yanlı olur - gerçek regresyon doğrusu muhtemelen her örneklem ortalamasından geçmeyecektir, fakat evren ortalamalasından geçer. Tartıştığımız bütün ölçmeler yanlı olmasına rağmen, $\eta^2$ bu sayıltı nedeniyle en yanlı olanıdır. $\eta^2$'nin sadelik avantajı ve sezgisel çekiciliği olmasına rağmen, değişkenlere ilişkin genel açıklamalar yapmak istediğinizde, daha az yanlı kestiriciyi kullanmayı tercih edebilirsiniz. Verinin sadece belirli kümesi hakkında açıklama yapmak veya etki büyüklüğü hakkında kaba bir fikir elde etmek istediğimizde, $\eta^2$ gayet iyi bir ölçüdür. Bundan başka $\eta^2$ ve diğerlerinde, örneklem büyüklüğü arttıkça ölçmeler birbirine yaklaşır, daha az yanlı olurlar.

## Omega-Kare ($\omega^2$)

**omega-kare ($\omega^2$)**

Dengelenmiş ($n$'ler eşit) veya yaklaşık olarak dengelenmiş desenler ile deneysel etkilerin büyüklüğünü ölçme yöntemi, **omega-kare ($\omega^2$)'**dir. $\omega^2$ birçok amaç için daha iyi bir alternatiftir. $\omega^2$'nin türetilmesi, daha önce görülen, alta yatan yapısal modele dayalıdır, $\omega^2$ için bağımsız değişkenin sabit veya tesadüfî olmasına bağlı olarak iki farklı formül bulunur. Tek yönlü desenlerde tesadüfî bağımsız değişken nadiren görülür, bu yüzden burada bu versiyon göz ardı edilecek; Bölüm 13 ve 14'te daha kompleks desenlerde anlamlı hale gelecektir.

Sabit model varyans analizi için, deneysel etkinin büyüklüğünün nispeten yansız kestirimi aşağıdaki eşitlikle verilmiştir:

$$\omega^2 = \frac{KT_{\text{işlem}} - (k-1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}}$$

Bu formülü Tablo 11.2'deki veriye uyguladığımızda, aşağıdaki değeri elde ederiz:

$$\omega^2 = \frac{KT_{\text{işlem}} - (k-1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{62,460 - 4(2,26)}{187,658 + 2,26} = \frac{53,42}{189,918} =, 285$$

Bu durumda $\omega^2$'nin kestirimi (,285), $\eta^2$ = ,334'ten daha düşük olur, ikincisi gerçeği daha yanlı yansıtır.

Lane ve Dunlap (1978), ölçümlerin büyüklüğünün düzenli olarak raporlanması hakkında bazı önemli çekinceleri belirtmişlerdir ve onların yorumlanması ışığında dergiler ağırlıklı olarak manidar sonuçlanan çalışmaları yayımlamıştır. Cohen (1973) hesaplamadaki bazı önemli hususları ve ölçümlerin büyüklüğünün yorumlanmasını özetler. Cohen

öncelikle faktöriyel desenlere (Bölüm 13'te tartışılacak) endişe duymasına rağmen, yaptığı yorumlar arkasındaki felsefe burada bile geçerlidir. Adı geçen tüm sayfalar net ve okunabilir, onları öneririm.

## *d*-Ailesi Etki Büyüklüğünün Ölçümü

**kökü-alınmış-ortalama-karenin-standartlaştırılmış etkisi (KAOKSE)**

Bir sonraki bölümde etki büyüklüğünün *d*-ailesi ölçümleri hakkında söyleyecek çok fazla sözüm olacaktır, ama burada kısaca Steiger (2004) tarafından tercih edilen, **kökü-alınmış ortalama-karenin-standartlaştırılmış etkisi (KAOKSE) (root-mean-square standardized effect RMSSE)** olarak adlandırılan bir ölçmeden kısaca söz edeceğim. Bu, grup farklılıklarının mantıksal ölçümüne dayanmaktadır ve çoklu gruplara da uygulanır. Ayrıca güç hesaplanmasında kullanılacak etki büyüklüğüyle neredeyse eşdeğerdir.

Belirli bir grup ortalamasının genel ortalamadan ne kadar saptığının ölçüsü aşağıdaki eşitlikle verilmiştir:

$$d_j = \frac{(\mu_j - \mu)}{\sigma}$$

Bunun basitçe, belirli bir ortalama ve genel ortalama arasındaki standartlaştırılmış fark olduğuna dikkat ediniz ve Bölüm 7'de görülen *d* ile tamamen aynı olmadığı halde benzerdir. Tüm gruplar üzerindeki bu ortalama mantıklıdır, ancak öncelikle bunların karesi alınmalıdır, yoksa ortalama sıfır çıkacaktır. Bu bize aşağıdaki gibi yazılabilecek bir ölçü verir:

$$d = \sqrt{\left(\frac{1}{k-1}\right)\Sigma\left(\frac{\mu_j - \mu}{\sigma}\right)^2}$$

Ortalamayı almak için *k* yerine (*k* - 1)'e bölüyoruz, çünkü sabit bir değişkenle ortalamadan sapmaların toplamı 0 olacaktır. Gruplardan sadece (*k* - 1) *sd* vardır ve bu bizim bölenimizdir. Bu istatistik için isim biraz garip görünüyor ama aslında bu çok doğru bir açıklamadır. Her bir grup için, onun ortalaması ve genel ortalama arasındaki farkı alınır ve bu standart sapmaya bölünür. Bu şekilde standartlaştırılmış bir etkiye sahip olunur. Daha sonra bunun karesi alınır ve tüm gruplar üzerinde ortalama alınır; böylece ortalama karesi alınmış standarlaştırılmış etkiye sahip olunur. Daha sonra bu sonucun karekökü alınır, bu bize, karesi alınmış ortalama standartlaştırılmış etkiyi verir. (Gerçekte karesi alınmış ortalama kare kök anlamına gelirken, neden "kökü alınmış ortalama kare" denildiğini bilmiyorum, ama daha etkileyici geliyor sanırım.)

KAOKSE etki büyüklüğünün mantıksal bir ölçümüdür, çünkü bu standart sapmaya bölünmesiyle standartlaştırılmış grup ortalamaları farkının doğrudan ölçümüdür.

Giancola ve Corman'ın alkol ve saldırganlık çalışması için ortalamalar şöyledir:

| D0 | d2 | d4 | d6 | d8 | Genel Ortalama |
|---|---|---|---|---|---|
| 1,802 | 0,675 | –0,912 | 0,544 | 1,889 | 0,800 |

Genel ortalamadan sapmanın karesel toplamı 5,5345 ve $KO_{\text{hata}}$'nın kare kökü 2,26'dır. Böylece,

$$\text{d} = \sqrt{\left(\frac{1}{\text{k - 1}}\right)\sum\left(\frac{\mu_\text{j} - \mu}{\sigma}\right)^2} = \sqrt{\left(\frac{1}{\text{k - 1}}\right)\frac{\sum(\mu_\text{j} - \mu)^2}{\sigma^2}} = \sqrt{\frac{\sum(\overline{\text{X}}_\text{j} - \overline{\text{X}}_{..})^2}{\text{KO}_{\text{hata}}}}$$

$$= \sqrt{\frac{5,205}{4(2,26)}} = \sqrt{,5758} = ,759 \quad \text{bulunur.}$$

Böylece grup ortalamaları, ortalama olarak ve yaklaşık olarak genel ortalamadan ¾ standart sapma farklılık gösterir, bu da önemli bir farklılıktır. (Bir sonraki kısımda tanıtılan G*Power yazılımını kullanıyorsanız, hemen hemen aynı etki büyüklüğünü elde edersiniz. Bu yazılımın yazarları KAOKSE (RMSSE) hesaplamayı işaret etmektedirler.)

Steiger (2004), bu ölçüm üzerinde güven aralıkları ayarlarını tartışmış ve ücretsiz yazılım sağlamıştır (bu makaleye bakın). Kelley (2008), R için fonksiyon olarak hatta daha fazlası için yazılan benzer programlar sağlamıştır. (ClonRMSSE.R olarak adlandırılan örnek için Web sitesine bakınız.) Glancola ve Corman deneyi için bu yazılımları kullanarak KAOKSE'nin (RMSSE) güven aralıklarını gösterebiliriz ,378 < $\delta$ < 1,038 (Bir parametre kestirildiği için bu ifadede $\delta$ kullandım). Bu sonuçtaki önemli bir husus, %95 olasılıklı güven aralığındaki alt sınırın yaklaşık ,40 olduğu, bu alt sınırın, ortalama grup farkı olarak, bir standart sapmanın yaklaşık yarısı olduğu anlamına gelmesidir. Bu durum, gruplar arasında önemli bir farktan bahsettiğimizi gösterir. Bir sonraki bölümde böyle ölçmeler hakkında daha fazla konuşacağız, ancak bu durumda tüm ortalama takımı yerine ortalama çiftlerine odaklanacağız. Bu bölümde aynı anda birçok grup arasındaki farklarla ilgili ölçmelere odaklandım. Kendi araştırmalarınızda en önemli olanın belirli grup farklılıkları olduğunu ve bu durumda *d*–ailesi ölçümlerinin ayrı avantajı olduğunu bulacağınızı düşünüyorum.

## 11.12 Güç

**merkezî olmayan *F* dağılımı**

Varyans analizinde gücün kestirimi, *t* için güç analizinin basit bir uzantısıdır, gösterimi farklı olmakla birlikte, tabloları aynıdır. *t* gibi, $\mu_j$'ler arasında beklenen farklılıklara dayalı bir istatistik ($\phi'$) istatistiğidir. Daha sonra çıkarılan ikinci istatistik phi ($\phi$) istatistiğidir ve bu *n* ve $\phi'$'nün fonksiyonudur ve sonuç olarak **merkezi olmayan *F* dağılım** tablosundan güç hesaplanır. Güç ile ilgili işlemlerin tamamı Cohen (1988) ve Koele'de (1982) bulunur.

Geçmişteki tartışmalarda hemen hemen her zaman merkezî olmayan *F* dağılımı tablosundan güç hesapladık, fakat yazılımın büyük kısmı hazırdı, şu anda büyük ölçüde bu tabloları göz ardı edebiliriz ve gücü doğrudan hesaplayabiliriz. Ama bunu yapmak için hâlâ doğrusal olmayan parametreleri ve ilgili istatistikleri anlamamız gerekir. Aşağıda, tabloları içeren hesaplamaları gösterilmiş ve iki farklı istatistik programına (SPSS ve G*Power) ve yeniden örnekleme tekniklerine odaklanılmıştır.

### Güç İçin Yeniden Örnekleme

Tek-yönlü varyans analizi için gücü anlamanın en kolay yollarından birisi basit bir yeniden örnekleme çalışması yapmaktır. Belki de 10.000 kere çekilen bir örneklem için "basit" kelimesini kullanmak garip gelebilir ama aslında bu basittir. SPSS, SAS veya R kullanarak bunu yapabilirsiniz.[7] Bu yaklaşımı tartışmamım nedeni, aslında buna ek olarak gücün neyi temsil ettiğini açık hale getirmesi, bir şekilde yeniden örneklemeyi kullanan daha fazla makale bulmaya çalışıyor olmamdır. Ve zaman geçtikçe bu kullanım artacaktır.

### Bir Örnek

Daha önce Conti ve Musty (1984) tarafından farelerin davranışları üzerinde THC'nin etkisini inceledikleri çalışmayı düşündük. Tablo 11.6'da verinin orijinal birimlerini kullandık, çünkü bu dönüştürmelere iyi bir örnek oluşturuyor. Fakat Conti ve Musty enjeksiyon öncesi et-

7 Bu problemi çözmek için gerekli olan R kodunu kitabın Web sayfasında ResamplingPower.R dosyasında bulabilirsiniz.

kinliğin bir yüzdesi olarak THC enjeksiyonundan sonrası etkinlik seviyesini kullanmıştır. Burada bu veriyi kullanacağız. Web'deki Tablo 11.6 veri dosyası da "DV" etiketli bir değişken içerir, biz bunu bağımlı değişken olarak kullanacağız.

Her bir koşul için ortalamalar aşağıda verilmiştir:

| Kontrol | 0,1 mg | 0,5 mg | 1 mg | 2 m2 | GM |
|---|---|---|---|---|---|
| 34,00 | 50,80 | 60,33 | 48,50 | 38,10 | 46,346 |

Ortalama örneklem varyansı ($KO_{\text{hata}}$) 240,345 idi, bu bize her grubun standart sapmasının kestirimini = $\sqrt{240,345}$ = 15,503 olarak verir.

Basit bir yeniden örnekleme çalışması kurabiliriz, evrenden 5 grup için tesadüfî veri kümesi, yani grup ortalaması ve standart sapması = $\sqrt{KO_{\text{hata}}}$ = 15,503 olacak şekilde veri kümeleri çekebiliriz. Başka bir ifadeyle, örneklem istatistikleri, parametrelere karşılık gelen kesrtirimler olduğunda onları alabiliriz. *F* için kritik değerin 2,579 olduğunu biliyoruz. Bu süreç 10.000 kere tekrar edilir, her seferindeki *F* sonuçlarını kaydederiz. Bu deneyin gücü şimdi tekrar etme yüzdesidir, 10.000 üzerinden elde edilen *F* değeri 2,579'u geçer. Bu neredeyse gücün tanımıdır. Bu durumda örnekleme 1 dakika 12 saniye sürdü ve ortaya çıkan olasılık ,9074 oldu. Evren ortalaması ve standart sapması Conti ve Musty'nin bulduklarına tam olarak eşitse bu bizim yeni çalışmamızın gücüdür. Güç için bu değer diğer ortalamalarla bulunan değerlere oldukça yakındır. (Çalışmayı her grupta tam olarak 10 denek olacak şekilde değiştirseydik, kestirilen güç ,9206 olurdu.)

Biz sadece gücü düşündük ve onu hesaplamanın bir yolunu gördük. Başkaları da var. Stephane Champely tarafından yazılan ve "pwr" olarak isimlendirilen R paketini cran.r-project.org Web adresinden indirebiliriz ve eşit örneklem büyüklüklerini varsayarak varyans analizi için güç hesaplayabilirsiniz. Her grupta 10 deneğin yer aldığı bu örnek için, bu program gücü ,92065 olarak kestirmiştiir. Bundan çok daha yakın olamazsınız. Gücü hesaplamak için bir başka yol SPSS için bir eklenti paketi kullanmaktır, bu sizin üniversitenizde mevcut olabilir, benimkinde mevcut değil veya Conti ve Musty verisini alarak SPSS'te varyans analizini çalıştırın ve "observed power (gözlenen güç)" seçeneğini tıklayın. "Gözlenen" kelimesinin güç aramak için yanlış isim olduğunu düşünüyorum; ortalamalar ve standart sapmalar tam olarak kestirilecekse, bu sonraki deneyin gücüdür. Bu amaç için G*Power'ın nasıl kulanılacağını kısaca göreceğiz.

## Biraz Daha Teori

Sadece ortalamaları, standart sapmaları ve örneklem büyüklüklerini bilerek hesaplamaları çalıştırabilir olsanız da, etki büyüklükleri ve merkezî olmayan parametreleri bilmeniz gerekir. Daha önceden bildiğimiz gibi,

$$\frac{E(KO_{\text{işlem}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + n\sum \tau_j^2/(k-1)}{\sigma_e^2}$$

$H_0$ doğruysa, $\sum\tau^2 = 0$ ve oran $F = KO_{\text{işlem}}/ KO_{\text{hata}}$ olacaktır. Bu alışılmış (merkezî) *F* dağılımı gibi dağılacaktır. Bu dağılımın ortalaması $sd_{\text{hata}}/ (sd_{\text{hata}} - 2)$ olacaktır, mâkul örneklemler için 1,00'e çok yakındır (bkz: s. 336, dipnot). $H_0$ yanlış ise, bu oran aşağıdaki gibi olacaktır,

$$\frac{E(KO_{\text{islem}})}{E(KO_{\text{hata}})} = 1 + \frac{n\sum \tau_j^2/(k-1)}{\sigma_e^2} = 1 + \frac{\lambda}{(k-1)}$$

Burada,

$$\lambda = \frac{n\sum \tau_j^2}{\sigma_e^2}$$

**lamda ($\lambda$)**

**merkezî olmama parametresi**

olup **lamda (lambda)** veya **merkezî olmama parametresi** olarak adlandırılır.

Merkezi olmama parametresinin, *F* dağılımının yerleşimini, basitçe, evren ortalamasından olan gerçek farkları 1'den pozitif yöne doğru olacak şekilde değiştirdiğini görebilirsiniz. Merkezî olmama parametresi *F* dağılımının ne kadar değiştiğini söyler ve bu güç kestirimleri üretmek için birçok bilgisayar yazılımı tarafından kullanılır. Her grupta 10 hayvanın olduğu bu deneyde her grup için merkezî olmama parametresi

$$\lambda = \frac{n\sum \alpha_j^2}{KO_{\text{hata}}} = \frac{10\left[(34,00-46,346)^2 + \ldots + (38,10-43,346)^2\right]}{240,345} = 18,329$$

olacaktır.

## Etki Büyüklüğü

Güç hesaplamak için, çoğu yaklaşım bir etki büyüklüğü ölçümü ile başlar (yazara bağlı olarak $\phi'$ veya *f* şeklinde gösterilir). Bu ölçü basitçe, beklenen evren ortalamalarının, $KO_{\text{hata}}$ olarak kestirdiğimiz hücreler-içi varyansa bölümüdür. Ancak Conti ve Musty gücü hesaplamak için sanki evren varyansıymış gibi örneklem varyansını işlediler, böylece standart sapmanın hesaplanmasında grupların *sd*'leri yerine ortalamaların sayısına bölünmüştür. Bu formül aşağıdaki gibidir:

$$\phi' = f = \sqrt{\frac{\sum \alpha_j^2/k}{KO_{\text{hata}}}} = \sqrt{\frac{\left[(34,00-46,346)^2 \ldots (38,10)-43,346)^2\right]/5}{240,345}} = \sqrt{\frac{440,534/5}{240,345}}$$

$$= \sqrt{0,3666} = ,605$$

## Gücün G*Power Kullanılarak Hesaplanması

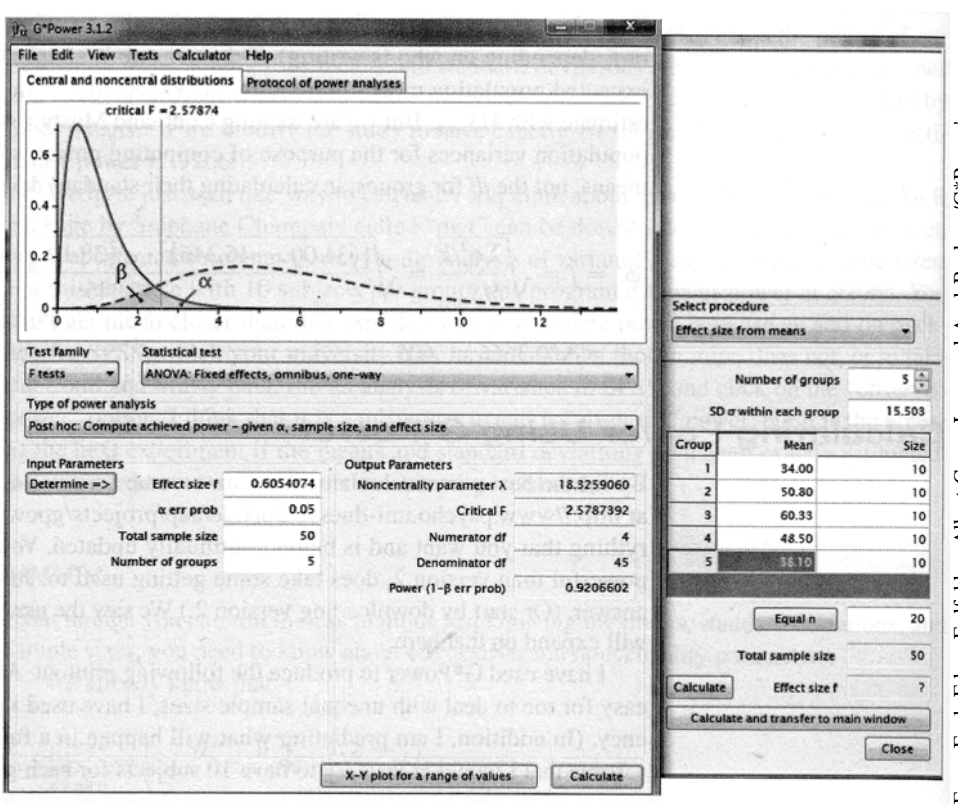
Bugüne kadar rastladığım en iyi güç hesaplayıcısı http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/ internetten ücretsiz olarak kullanılabilir. Bu istediğiniz her şeyi yapacaktır ve sürekli güncellenmektedir. Sürüm 3.0 sürüm 2'den daha güçlü olsa da, alışmak biraz zaman alıyor. Doğru cevabı bulana kadar sadece oynayın (veya verisyon 2'yi indirerek başlayın). Bölüm 8'de G*Power'ın kullanımını gördük, ancak burada bu konu genişletilecektir.

Aşağıdaki çıktıyı üretmek için G*Power kullandım. Yazılım benim için eşit olmayan örneklem büyüklükleri ile başa çıkmayı kolaylaştırmasına rağmen, tutarlılık için örneklem büyüklüklerini 10 kullandım. (Buna ek olarak, gelecekteki bir deneyde ne olacağını kestiriyorum ve en azından her bir grup için 10 deneğin var olacağını varsayıyorum.) G*Power kendi hesaplamalarını her durumda ortalama örneklem büyüklüğü üzerine dayandırır. Gösterim 11.2'deki ekranın sağ tarafı etki büyüklüğü hesaplanmasının sonucunu göstermektedir. Varyans analizi için güç istediğimi belirttim ve beş grup için ortalamalar ile örneklem büyüklüklerini girdim. "Calculate and transfer" düğmesine bastığımda program otomatik olarak etki büyüklüğünü hesapladı. Bu durum için 0,6054 olarak elde etti. Bu daha önceden *f* veya $\phi$' olarak etiketlediğimiz değer için hesapladığımızın aynısıdır. Bölüm 13'te yer alan faktöriyel varyans analizine gelirsek, *f*'yi doğrudan hesaplamamız

gerekir. (Etki büyüklüğü gözlem sayısından bağımsızdır, bu da örneklem büyüklüğünün eşit olmamasını önemsizleştirir.) Daha sonra gücü hesaplamak için pencerenin solunda yer alan "Calculate" düğmesine bastım. Conti ve Musty'nin örneklem ortalamalarını ve hata terimlerini kullandığımdan çoklu karşılaştırma (Post-Hoc) gücü için de güç isteğinde bulundum.[8]

Dikkat ederseniz hesaplanan güç daha önce hesapladığımız değer olan ,92066 ile aynıdır.

Koele (1982), tesadüfî modellerde güç hesaplaması için bazı yöntemler sunmuştur. Tesadüfî modeller, tek yönlü düzenlemeler için yaygın olmasa da yüksek mertebeden desenler için daha yaygındır ve gücün az olmasından kaynaklı bazı problemleri vardır. Bu modeller için, iki tesadüfî süreç gerekir-katılımcıların tesadüfî ve işlem düzeylerinin tesadüfî örneklenmesi. Koele bunu "Her düzeyde çok sayıda gözlem olmasının yanında, her işlem için de birçok düzey olmalıdır (bağımsız değişken olarak) ifadesiyle belirtmiştir. İki veya üç düzeyli tesadüfî faktör içeren deneyler, iki veya üç gözlem içeren örneklemlerde *t* testi yapmak kadar saçmadır" (s. 516) ifadesini kullanmıştır. Bu, tesadüfî modeller için akılda tutulması gereken önemli bir tavsiyedir. Bunun hakkında Bölüm 13'te daha fazlasını söyleyeceğiz.




Franz Faul, Edgar Erdfelder, Albert-Georg Lang and Axel Buchner/G*Power'in nazik izinleriyle

**Gösterim 11.2** Conti ve Musty deneyinin G*Power güç kestirimi

8 Eğer ben bu kitabı yazıyor olsaydım burada "çoklu karşılaştırma gücü (post - hoc power)" deyimini kullanmazdım çünkü bu farklı kişilerde farklı anlaşılmalara yol açar. Gerçekte yaptığım, daha önceki çalışmalardan parametre kestirimi yapmak ve bunları kullanarak güç kestirmi yapmaktır. Bu, benim gibi çoklu karşılaştırma gücü (post-hoc power) terimini kullanmak istemeyenler için oldukça geçerli bir yaklaşımdır.

Güç ve desen hakkında üzerinde durulması gereken son bir nokta daha vardır. McClelland (1997) ikna edici bir biçimde, bağımsız değişken olarak kullandığımız sabit değişkenlerin genellikle ihtiyacımız olandan daha fazla düzeyde alındığını iddia etmiştir. Örneğin, eğer McClelland, Giancola ve Corman'ın (2007) alkol ve saldırganlık üzerine olan deneyini yapıyor olsaydı, McClelland'ın yalnız üç grup (D0, D4, and D8) kullanacağından şüphelenirdim. McClelland beş grup kullanmanın dört serbestlik derecesi ile etkiyi azaltacağını iddia ederdi. Benzer olarak Conti ve Musty (1984) çalışmasında büyük ihtimalle 0,05 *mg* ve 2 *mg* gruplarını kullanırdı. Bunu da diğer iki koşuldakiler birleştirildiğinde oluşan örneklem büyüklüğü kadar deneği 0,05 *mg*'lik gruba yerleştirerek yapardı. Gücün maksimize edilmesi ve iyi deneysel desenler hakkında olan bu çalışmayı okumanızı tavsiye ederim. Oldukça önemli ve kolay okunabilecek türden bir çalışmadır.

## 11.13 Bilgisayar Analizleri

Gösterim 11.3 Everitt'in iştahsız kızlarla ilgili işlemlere ait verinin SPSS analiz çıktılarını kapsamaktadır. Analyze/Compare Means/One-Way Anova yerine Analyze/General Linear Model/Univariate prosedürünü tercih ettim. (Menü seçimlerini göstermedim, ama bunlar kanıt olabilir.) Bu, Bölüm 13 ve Bölüm 14'te kullanacağımız prosedür olacaktır ve bu yol tek yönlü yöntem ile aynı sonuçları üretecektir. Bu aynı zamanda, sizin aşina olmadığınız ve aşağıda açıklanan birtakım çıktılar üretebilir.

**Betimsel İstatistikler**

Bağımlı Değişken: WTGAIN

| İşlem Grubu | Ortalama | Std. Sapma | N |
|---|---|---|---|
| Kontrol | −,450000 | 7,988705 | 26 |
| Bilş. Dav. | 3,006897 | 7,308504 | 29 |
| Aile | 7,264706 | 7,157421 | 17 |
| Toplam | 2,763889 | 7,983598 | 72 |

**Katılımcılar Arası Etkilerin Testleri**

Bağımlı Değişken: WTGAIN

| Kaynak | Tip III Kareler Toplamı | sd | Kareler Ortalaması | F | Mnd. | Eta Kare | Merkezî Olmayan Parametre | Gözlenen Güç[a] |
|---|---|---|---|---|---|---|---|---|
| Düzeltilmiş Model | 614,644[b] | 2 | 307,322 | 5,422 | ,006 | ,136 | 10,845 | ,830 |
| Sabit (Intercept) | 732,075 | 1 | 732,075 | 12,917 | ,001 | ,158 | 12,917 | ,943 |
| GRUP | 614,644 | 2 | 307,322 | 5,422 | ,006 | ,136 | 10,845 | ,830 |
| Hata | 3910,742 | 69 | 56,677 | | | | | |
| Toplam | 5075,400 | 72 | | | | | | |
| Düzeltilmiş Toplam | 4525,386 | 71 | | | | | | |

[a] Alfa =,05 kullanılarak hesaplanmıştır.
[b] R kare = ,136 (Düzeltilmiş R kare = ,111)

Adapted from output by SPSS, Inc.

**Gösterim 11.3** Everitt'in iştahsızlık üzerine yaptığı çalışmaya ait veri için SPSS'te genel doğrusal model analizi

Özet tabloya dikkat ederseniz ilk satır "Düzeltilmiş Model" olarak adlandırılmıştır. Eğer orada iki veya daha fazla bağımsız değişken olsaydı (Grup ve Cinsiyet gibi) bu durumda bu satır bu değişkenlerin birleştirilmiş etkisini gösterecekti. Sadece bir bağımsız değişken olduğundan Model ve Grup etkileri tam olarak aynı olacaktır.

"Sabit (Intercept)" olarak adlandırılan satır, evrende genel ortalamanın 0'a eşit olduğu yokluk hipotezinin test edilmesi içindir. Bu durum için çalışmada, "kızlar, gruplar boyunca ağırlık kazanmış gibi mi" önemli sorusu sorulsa da, bu testi nadiren ciddiye alırız.

"Grup", "Hata" ve "Düzeltilmiş toplam" Kesim 11.1.'de gördüğümüz gibidir.

Grup etkisi için olan her eta-karenin, tablonun en altında yer alan "R-kare" ile aynı sonucu verdiğine dikkat ediniz. Bu bize ağırlık kazanımındaki %14'lük değişimin işlemler arası farka atfedilebileceğini söyler.

SPSS daha sonra elde edilen ortalamalara parametre gibi ve $KO_{\text{hata}}$'ya evren varyansının uygun kestiricisiymiş gibi davranarak "gözlenen güç (observed power)"ü hesaplar. Bu örnekte eşit olmayan örneklem büyüklükleri olduğundan değerleri tam olarak aynı olacak şekilde üretmekte problem yaşayabilirsiniz.

## Anahtar Terimler

Varyans analizi (ANOVA) (Giriş)
Tek-yönlü varyans analizi (Giriş)
Yapısal model (11.2)
Varyansın homojenliği (11.2)
Heterojenlik (11.2)
Hata varyansı (11.2)
Varyansın heterojenliği (11.2)
Varyansların heterojenliği (11.2)
$KO_{\text{hata}}$ (11.3)
$KO_{\text{grupiçi}}$ (11.3)
$KO_{\text{işlem}}$ (11.3)
İşlem etkisi (11.3)
Beklenen değer (11.3)
Kareler toplamları (11.4)
$KT_{\text{toplam}}$ (11.4)
$KT_{\text{işlem}}$(11.4)
$KT_{\text{hata}}$ (11.4)
Özet tablo (11.4)
$sd_{\text{toplam}}$ (11.4)
$sd_{\text{işlem}}$ (11.4)
$sd_{\text{hata}}$ (11.4)
Dengelenmiş desenler (11.7)
Tesadüfî kayıp (11.7)
Aşırı-kuyruklu dağılımlar (11.9)
Pencerelenmiş (Winsorize edilmiş) örneklemler (11.9)
Yeniden örnekleme istatistikleri (11.9)
Tekrarlama (11.10)
Sabit-model varyans analizi (11.10)
Tesadüfî-model varyans analizi (11.10)
Deneysel etkilerin büyüklüğü (11.11)
Eta-kare ($\eta^2$) (11.11)
Korelasyon oranı (11.11)
Hatada yüzde azaltma (11.11)
Omega kare ($\omega^2$) (11.11)
Karekökü alınmış ortalama kare standartlaştırılmış etki (KAOKSE) (11.11)
Merkezi olmayan $F$ dağılımı (11.12)
Lamda ($\lambda$) (11.12)
Merkezî olmayan parametre (11.12)

## Alıştırmalar

11.1 Eysenck (1974) yılında, katılımcıların, bir kelime listesinden, kelime hatırlamalarının gerektiği bir çalışma yapmıştı. Bu çalışmada, bir kelimedeki harfleri basitçe sayma, kafiyeli bir kelimeyi düşünme, uygun bir sıfatı düşünme, kelimenin görsel imajını yaratma veya daha sonra hatırlamak için kelimeleri çalışma gibi farklılaşan durumlar yer almaktadır. Bağımlı değişken, test denemesinde hatırlanan kelime sayısıdır. Bu veri kümesi aşağıda gösterilmiştir.

| | Sayma | Kafiye | Sıfat | İmgeleme | Kasıtlı | Toplam |
|---|---|---|---|---|---|---|
| | 9 | 7 | 11 | 12 | 10 | |
| | 8 | 9 | 13 | 11 | 19 | |
| | 6 | 6 | 8 | 16 | 14 | |
| | 8 | 6 | 6 | 11 | 5 | |
| | 10 | 6 | 14 | 9 | 10 | |
| | 4 | 11 | 11 | 23 | 11 | |
| | 6 | 6 | 13 | 12 | 14 | |
| | 5 | 3 | 13 | 10 | 15 | |
| | 7 | 8 | 10 | 19 | 11 | |
| | 7 | 7 | 11 | 11 | 11 | |
| Ortalama | 7,00 | 6,90 | 11,00 | 13,40 | 12,00 | 10,06 |
| Std. Sapma | 1,83 | 2,13 | 2,49 | 4,50 | 3,74 | 4,01 |
| Varyans | 3,33 | 4,54 | 6,22 | 20,27 | 14,00 | 16,058 |

© Cengage Learning 2013

Bu veri kümesine tek-yönlü varyans analizi uygulayınız.

11.2 Eysenck'ın (1974) çalışmasının başka bir yönü, Alıştırma 11.1'e bakılırsa, Genç ve Yaşlı deneklerin, daha sonra hatırlamak için ezberlemeleri gerektiği talimatlar doğrultusunda, materyali hatırlama yeteneklerini karşılaştırmıştır - Amaçlı grup. (Büyük ihtimalle bu görev yüksek düzeyde işlem gerektirmektedir.) Aşağıdaki veri kümesi, bağımlı değişken olan hatırlanan öğe sayısını vermektedir.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Genç: | 21 | 19 | 17 | 15 | 22 | 16 | 22 | 22 | 18 | 21 |
| Yaşlı: | 10 | 19 | 14 | 5 | 10 | 11 | 14 | 15 | 11 | 11 |

a. Bu iki grubun ortalamalarını karşılaştıran varyans analizi yapınız.

b. Veri kümesi üzerinde bağımsız gruplar için *t* testi yapınız ve sonuçları (a)'da elde ettiğiniz sonuçlarla karşılaştırınız.

11.3 Eysenck'ın (1974) çalışmasındaki veri kümesini incelemenin başka bir yolu da deneklere ait dört grubu karşılaştırmaktır. Bir grup, düşük işlem düzeyindeki bir durumda hatırlamaları için kelime sunulan Genç deneklerden oluşmaktadır. İkinci grup, yüksek işlem düzeyindeki (Alıştırma 11.2'de olduğu gibi) bir durumda hatırlamaları için kelime verilen Genç deneklerdir. Diğer iki grup da düşük ve yüksek işlem düzeylerindeki durumlarda kelime sunulan Yaşlı deneklerin oluşturduğu gruplardır. Veri kümeleri aşağıdaki gibidir.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Genç/Düşük: | 8 | 6 | 4 | 6 | 7 | 6 | 5 | 7 | 9 | 7 |
| Genç/Yüksek: | 21 | 19 | 17 | 15 | 22 | 16 | 22 | 22 | 18 | 21 |
| Yaşlı/Düşük: | 9 | 8 | 6 | 8 | 10 | 4 | 6 | 5 | 7 | 7 |
| Yaşlı/Yüksek: | 10 | 19 | 14 | 5 | 10 | 11 | 14 | 15 | 11 | 11 |

a. Veri üzerinde tek-yönlü varyans analizi yapınız.

b. 1. ve 3. işlemleri birleştirerek ($n = 20$) ve 2. ve 4. işlemleri birleştirerek bu iki yeni grup için tek yönlü varyans analizi yapınız. Hangi soruya cevap veriyorsunuz?

c. (b) için cevabınızı yorumlaması neden bu kadar zor?

11.4 Alıştırma 11.3'e bakınız. Şimdi 1. ile 2. işlemleri ve 3. ile 4. işlemleri birleştirerek bu iki yeni grup için tek-yönlü varyans analizi yapınız. Hangi hipotezleri test ediyoruz?

11.5 Alıştırma 11.2'ye bakınız. Ek veri topladığımızı varsayalım ve Genç gruba puanları 13 ve 15 olan iki katılımcı daha ekleyelim.

a. Tekrar varyans analizi yapınız.

b. Varyansları birleştirmeden bağımsız gruplar için *t* testi yapınız.

c. Varyansları birleştirdikten sonra bağımsız gruplar için *t* testi yapınız.

d. (b) ve (c)'deki *t* değerlerinden hangisi (kareleri alındıktan sonra), (a)'da hesaplanan *F* değerine karşılık gelir?

11.6 Alıştırma 11.2'deki veri için $\eta^2$ ve $\omega^2$'yi hesaplayınız. Sabit-modeli mi, yoksa tesadüfî-modeli mi kabul ettiniz?

11.7 Alıştırma 11.3'teki veri için $\eta^2$ ve $\omega^2$ hesaplayınız.

11.8 Foa, Rothbaum ve Murdock (1991) tecavüz kurbanları için dört farklı terapi türü önerdikleri bir çalışma yapmışlardır. Stres aşılama terapisi (The Stres Inoculation Thrapy SIT)) grubuna stresle başa çıkma ile ilgili talimatları verilmiştir. Uzun süreli maruz kalma (The Prolongued Expoasure (PE)) grubu ise zihinlerinde defalarca aynı olayı tekrarlamışlardır. Destekleyici danışmanlık (Supportive counseling (SE)) grubuna genel problem teknikleri öğretilmiştir. Son olarak Bekleme listesine (Waiting list (WL)) kontrol grubu olarak hiçbir terapi uygulanmamıştır. Aşağıdaki veri kümesinde, bağımlı değişken, semptomların bir dizi önem derecesi olarak verilmiştir.

| Grup | *n* | Ortalama | Std. Sap. |
|---|---|---|---|
| SIT | 14 | 11,07 | 3,95 |
| PE | 10 | 15,40 | 11,12 |
| SC | 11 | 18,09 | 7,13 |
| WL | 10 | 19,50 | 7,11 |

a. Varyans heterojenliği ile ilgili herhangi bir problemi yok sayarak varyans analizi yapınız ve sonuçları doğru kabul ederek açıklayınız.

b. Heterojen varyanslar için Welch prosedürünü uygulayınız. Bu sizin sonuçlarınızı etkiledi mi?

c. Dört grubun ortalamalarını gösteren bir grafik çiziniz.

d. $H_0$'ın reddedilmesi bu durumda ne anlama geliyor?

11.9 Alıştırma 11.8'deki veri için $\eta^2$ ve $\omega^2$'yi hesaplayınız ve sonuçları yorumlayınız.

11.10 Alıştırma 11.8a'daki örneklem büyüklüğü iki katı olsaydı ne olurdu? Tüm istatistikler aynı kalır mıydı?

11.11 Giancola çalışması için verilen Tablo 11.1'deki veri kümesini logaritmik dönüştürme kullanarak tekrar analiz ediniz. Dönüştürme nasıl etkilere yol açmaktadır?

11.12 Alıştırma 11.8'deki veri Ex 11.12.dat Web sitesinde de bulunmaktadır. Analizi SPSS veya başka bir yazılım programı kullanarak yapınız ve testleriniz varyans heterojenliği sağlamak için Welch modifikasyonu içersin. Sonuçları Alıştırma 11.8b'deki cevaplarınızla karşılaştırınız.

11.13 Alıştırma 11.1 için uygun bir istatistiksel model yazınız.

11.14 Alıştırma 11.2 için uygun bir istatistiksel model yazınız.

11.15 Alıştırma 11.3 için uygun bir istatistiksel model yazınız. Bölüm 13'te kullanmak üzere kaydediniz.

11.16 *F,* 1'den küçük olduğunda, genelde gerçek değer yerine "<1" yazarız. *F*'in 1'den küçük olması ne anlama gelir? 1'den küçük olan bir *F* değeri için "önemli ölçüde" mantıklı konuşabilir miyiz? Cevabınıza *E*(*KO)*'nı ekleyiniz.

11.17 Howell ve Huessy (1981), çocukları, dikkat eksikliği bozukluğu (attention deficit desorder (ADD)) gösteren ve göstermeyen olarak sınıflandırmışlardır - ikinci, dördüncü ve beşinci sınıftaki çocukların davranışlarıyla ilgili olarak. Denekler bu davranışı gösterdikleri yıla göre sıraya konulmuşlardır. Daha sonra, bu çocukların lisedeki genel başarı ortalamalarına (GBO) bakılmıştır. Aşağıdaki veri kümesinde her grup için GBO görülmektedir.

| | ADD hiç görülmedi | Sadece ikincide | Sadece dördüncüde | İkincide ve dördüncüde |
|---|---|---|---|---|
| Ortalama | 2,6774 | 1,6123 | 1,9975 | 2,0287 |
| Std. Sap. | 0,9721 | 1,0097 | 0,7642 | 0,5461 |
| *n* | 201 | 13 | 12 | 8 |

| | Sadece beşincide | İkinci ve beşincide | Dördüncü ve beşincide | İkinci, dördüncü ve beşincide |
|---|---|---|---|---|
| Ortalama | 1,7000 | 1,9000 | 1,8986 | 1,4225 |
| Std. Sap. | 0,8788 | 1,0318 | 0,3045 | 0,5884 |
| *n* | 14 | 9 | 7 | 8 |

© Cengage Learning 2013

Varyans analizi yapınız ve uygun karara ulaşınız.

11.18 Alıştırma 11.17'deki analizi, ADD hiç görülmedi grubunu çıkararak tekrar yapınız. Analizin bu şekilde yapılması verinin yorumlanmasını nasıl daha açık hale getirmektedir?

11.19 Tablo 11.6'daki veri için karekök dönüşümü uygulayınız.

11.20 Alıştırma 11.19'da elde ettiğiniz dönüşümü yapılmış veri için varyans analizi yapınız.

11.21 Alıştırma 11.17'deki veri için $\eta^2$ ve $\omega^2$ yi hesaplayınız.

11.22 Darley ve Latane (1968) başı dertte olan bir kişi için deneklerin yardım çağırma hızını kaydettiler. Denekler, kişiyi duyanın sadece kendileri olduğunu düşünenler (Group 1, *n* = 13), başka birinin de o kişiyi duyduğunu düşünenler (Group 2, n = 26) veya dört başka insanın da kişiyi duyduğunu düşünenler (Group 3, *n* = 13) olarak gruplandırıldılar. Bağımlı değişken kişinin yardım çağırma hızı (=1/zaman x 100) olarak alındı. Hız puanlarının üç grup için ortalamaları sırasıyla, 0,87, 0,72 ve 0,51 olarak bulundu. $KO_{hata}$ 0,053 olarak hesaplandı. Varyans analiz tablosunun özetini tekrar oluşturunuz. Ne karara varabilirsiniz?

11.23 Alıştırma 11.22'de veri orijinal birimlerinden saniyelere dönüştürülmüştür. Bu, dağılımın şeklinde nasıl bir değişime yol açabilir?

11.24 Tablo 11.2'deki dönüşüm varyans eşitleme açısından faydalı olabilir mi? Ne tür bir dönüşüm uygulaması önerirsiniz?

11.25 İsim ve fiil hatırlamayı karşılaştırdığımız bir çalışma yaptığımızı varsayalım. Her deneğe 25 isim ve 25 fiil sunuyoruz ve listeden hatırlamalarını istiyoruz. Hem konuşma parçaları arasındaki farka hem de isim kategorisindeki kelimeler arasındaki farka bakıyoruz. Hangi değişken sabit değişken ve hangi değişken muhtemelen tesadüfî değişken olur?

11.26 Ana bağımsız değişkenin tesadüfî bir değişken olduğu bir çalışmaya örnek veriniz.

11.27 Davey, Startup, Zara, MacDonald ve Field (2003) kişiye bağlı olan kompulsif kontrol derecesinin ruh hali üzerindeki rolü ile ilgilenmişlerdir. (Kompulsif kontrol, psikopatolojilerin bir dizisinin içindedir.) 10 katılımcının yer aldığı üç grubun her birine pozitif, negatif ve doğal modda müzik dinletilmiştir. Sonra katılımcılara "üç haftalığına evden ayrılırken güvenlik için kontrol ettikleri evin civarındaki şeyleri listeleyin" talimatı verilmiştir. Bağımlı değişken listelenen şeylerdir. Gerçek veri kümesi aşağıdaki gibidir:

| Uyarılmış Amaç | | |
|---|---|---|
| Negatif | Pozitif | Hiçbiri |
| 7 | 9 | 8 |
| 5 | 12 | 5 |
| 16 | 7 | 11 |
| 13 | 3 | 9 |
| 13 | 10 | 11 |
| 24 | 4 | 10 |
| 20 | 5 | 11 |
| 10 | 4 | 10 |
| 11 | 7 | 7 |
| 7 | 9 | 5 |

© Cengage Learning 2013

a. Uygun varyans analizini yapınız ve kendi sonuçlarınızı çiziniz.
b. Bir sonraki bölümde çoklu karşılaştırmalar yaparken, teorik nedenlerle karşılaştırmalarda hangi sütun ortalamaları ile ilgilenirsiniz?

11.28 Alıştırma 7.46'da ailelerini ölüm ile kaybetmiş, boşanmış bir aileden gelen veya iki aile ile büyüyen öğrencilerle ilgili veri kümemiz vardı. Bu üç grubu karşılaştıran üç ayrı $t$ testi yaptınız.
a. GSIT bağımlı değişken olmak üzere bu veri kümesi üzerinde varyans analizi yaparak analizi tekrar yapınız.
b. Cevaplarınız, Bölüm 7'deki cevaplarınızdan nasıl farklılaşıyor?

Alıştırma 11.29-11.31 arasında aşağıda verilen materyali kullanınız.

Introini-Collison ve McGaugh (1986) vücutta normal olarak üretilen hormonların bellek üzerinde bir rol oynayabileceği hipotezini test etmişlerdir. Özellikle, eğitim sonrası adrenalin enjeksiyonunun daha önceden öğrenilen ayrımları akılda tutma üzerindeki etkilerine bakmışlardır. Deneyi basitleştirmek için ilk olarak, fareyi elektrik şokundan Y labirentinin sol kolunu seçerek kaçacak şekilde eğitmişlerdir. Eğitimin hemen sonrasında farelere 0,0; 0,3 veya 1,0 mg/kg adrenalin enjekte etmişlerdir. (İlk gruba fizyolojik serum enjekte edilmiştir.) Düşük doz adrenalinin akılda tutmayı kolaylaştırdığını, yüksek dozun ise engellediğini keşfetmişlerdir.

Orijinal eğitimden 1 gün, 1 hafta veya 1 ay sonra, her fare Y labirentine tekrar konulmuştur, fakat bu kez şoktan kaçmak için labirentin sağ kolunu çalıştırmak gerekmektedir. Muhtemelen orijinal eğitimin güçlü belleği, yeni görevin öğrenilmesine engel olur ve denek daha fazla hata yapar.

Bu deneyde iki veri kümesi vardır ve Bilgisayar Veri Kümeleri ekinde verilmiştir. Orijinal çalışma 1 gün sonra test edilen üç farklı dozajdaki 18 hayvanı ve 1 hafta ve 1 ay sonra her grupta test edilen 12 hayvanı içerir. Hipotetik veri, Epinuneq.dat'ta bulunan orijinal sonuçlara yakın şekilde elde edilir, buna rağmen amacımız için verimizde, 1 aylıkken test edilmiş 1,0 mg/kg dozda 7 katılımcı yer almaktadır. İkinci veri kümesi her 9 hücrede 12 gözlem bulunacak şekilde oluşturulmuştur ve Epineq.dat olarak adlandırılmıştır. Her iki durumda da veri kümelerini, sonuçları tam sayı verecek ve gerçek veri ile yakından ilişkili olacak şekilde oluşturmaya ihtiyacımız vardır. Ancak $H_0$ ile ilgili sonuçlar aynıdır.

11.29 Bir aralıktan, onu izleyen aralık arasında önemli bir fark olmaması mâkul bir varsayımdır. Aralık değişkenini yok sayarak veri kümelerini birleştiriniz ve Dozaj üzerinde varyans analizi yapınız. Epinuneq.dat'taki veriyi kullanınız. (0,0 ve 0,3 dozları için 42 ve 1,0 mg/kg için de 37 gözleminiz olacaktır.)

11.30 Epinuneq.dat'taki veriyi kullanarak üç ayrı tek-yönlü varyans analizi yapınız, her akılda tutma aralığında. Her durumda, üç dozaj ortalaması eşittir sıfır hipotezini test ediniz. İstatiksel paketiniz, her analiz için üç dozaj grubunun ortalamalarını, standart sapmalarını çıkarıyor mu? Şimdi üç Aralık ortalamaları eşit hipotezini test edecek şekilde her analizi yapınız. Bu durumda dozajı göz ardı edeceksiniz.

11.31 Alıştırma 11.29'u tekrar yapınız, bu kez Epineq.dat. veri kümesini kullanınız. (Sonuçlar değişecektir çünkü veri kümeleri farklıdır.) Üç hata teriminin ($KO_{\text{hata}}$) ortalamalarını hesaplayınız ve deneydeki dokuz grup için de varyans ortalamalarının eşit olduğunu gösteriniz. Bu değerleri Bölüm 13'te kullanmak üzere saklayınız.

11.32 Strayer, Drews ve Couch (2006) alkol için yasal limitlerde olan bir kontrol grubu ve telefonla konuşanlardan oluşan bir grubu otomobil kullanma davranışları açısından karşılaştıran bir çalışma yapmışlardır. Bu bölüme uygun olması açısından bu çalışmayı biraz değiştirdim, ama sonuçlar onların sonuçlarına uymaktadır. Bu üç grup aşağıda verilmiştir. (Veri kümelerine bu kitap için hazırlanan Web sitesinde Ex 11-32.dat. dosyasından ulaşılabilir.)

Kontrol

808 757 773 937 726 788 806 792 751 765 853 655 626 721 630 722 683 709
718 812 703 791 586 864 737 701 799 844 639 705 822 935 842 827 784 838
795 823 791 819

Alkol

631 656 621 706 937 538 947 855 661 887 750 945 663 880 873 830 751 855
815 593 666 730 1021 906 821 956 606 660 802 961 629 603 826 531 828 959
743 745 922 829

Cep Telefonu

909 712 805 852 859 781 841 822 740 910 900 912 863 785 863 809 927 847
918 810 788 929 798 863 981 842 1021 827 876 736 640 851 787 703 942 758
843 781 969 872

Bu veri kümesine göre, cep telefonunun temel koşullara göre daha uzun süre reaksiyona neden olduğu konusunda bir kanıt mıdır? Cep telefonu durumu alkol-engeli durumundan nasıl farklılaşmaktadır?

11.33 Gouzoulis-Mayfrank ve arkadaşları (2000) bir grup Ecstacy, bir grup Cannibis kullanıcısı ve bir de kullanıcı olmayan kontrol grubu alarak ilk iki grubun arasında ve kontrol grubu ile bunların aralarında verilen görev yönünden performanslarında farklılık olup olmadığını karşılaştırmışlarıdır. Her grupta 28 katılımcı vardır. Performanslar birkaç farklı görevde değerlendirilmiştir, ama biz soyut düşünme üzerine odaklanacağız. Aşağıdaki veri kümeleri orijinal çalışmadaki gibi aynı ortalama ve varyanslar ile oluşturulmuştur, yüksek puanlar daha iyi performansı temsil etmektedir. Veri kümelerine Web sitesindeki Ex.11-33.dat. dosyasından ulaşabilirsiniz.

Ecstacy: 25 25 23 32 21 28 34 26 23 22 26 21 29 28
23 24 29 23 30 18 25 25 25 25 32 23 29 32

Kontrol: 29 31 31 25 33 21 18 40 35 32 29 31 25 32
33 34 28 28 25 22 27 34 38 31 30 31 26 30

Camibis: 28 28 28 29 37 34 27 25 34 28 25 27 30 29 31 32
30 29 32 25 32 31 32 22 31 25 28 34 29 32

a. Üç grubun ortalamalarını karşılaştırmak için varyans analizi yapınız.
b. İkili etki büyüklüklerini incelemek için Cohen'ın *d*'sini hesaplayınız - başka bir deyişle, her grubu her diğer grupla karşılaştırmak için *d*'yi hesaplayınız.
c. Bu çalışmadan çıkarılacak mantıklı sonuç nedir?

## Tartışma Soruları

11.34 Deney yapan bazı kişiler veriyi dönüştürünce suçluluk hissederler. Dönüşümlerin genellikle mükemmel şekilde kabul edilebilir olduğu ile ilgili mantıklı bir gerekçe oluşturunuz.

11.35 Conti ve Musty (1984) çalışmasında THC'nin aktivite üzerindeki etkisinde, ortalamalar doz ile birlikte doğrusal olarak artmıyordu. Eğer bir etkisi varsa, herhangi bir etki büyüklüğünde ölçülmesi gerekir mi?

11.36 Bir doğrultuda aralık boyunca ortalamaları eşit bulunan dört grubunuz veya üç grubunuzun ortalaması yaklaşık olarak eşit ve dördüncünün ortalaması farklı veya yaklaşık olarak eşit ortalamaya sahip ama diğer ikisi yaklaşık olarak eşit ortalamaya sahip grubunuzdan farklı iki grubunuz veya başka bir yapıda grubunuz var diyelim. Kendi kendinize oluşturduğunuz basit veriyi kullanınız (grup içi varyansları içersin), ortalamaların yapısının bir fonksiyonu olarak *F* istatistiği nasıl değişir?

11.37 Linda Teri ve çalışma arkadaşları (Teri, 1997) depresyondan dolayı acı çeken Alzheimer hastlarının tedavisinde ilâç dışı yöntemler denemişlerdir. İlki hoş olayları vurgulayan (BT-PE) ve diğeri ise bakıcı problem çözmeyi vurgulayan (BT-PS) iki davranışsal tedavileri, ayrıca tipik bakım grubu (TCC) ve bekleme listesi-kontrol grubu (WLC) vardır. Bağımlı değişken, tedavinin başlangıcından sonucuna kadar, Hamilton Depresyon Puanlama Ölçeği ile ölçülen depresyondaki değişikliklerdir. (Yüksek puanlar daha fazla iyileşmeyi gösterir.) Veri kümelerine www.uvm.edu/~dhowell/methods8/DataFiles/Teri.dat. Web sitesinden ulaşabilirsiniz. Bu veri kümeleri orjinali gibi öncesi ve sonrası test puanlarının arasındaki korelasyonun tahmini ile yaklaşık olarak aynı ortalama ve varyanslarla oluşturulmuştur. Bölüm 14'te bu veri kümelerine geri döneceğiz. Bu veri kümeleri için nasıl sonuçlar çıkarırdınız?



Bölüm 12

# İşlem Ortalamaları Arasında Çoklu Karşılaştırmalar

## Hedefler

Varyans analizinin genişletilmesi olarak bir ortalamalar kümesi içinde karşılaştırma yapmanın yollarının incelenmesi.

## Konular


12.1 Hata Oranları
12.2 Morfin Toleransı Üzerine Yürütülen Basit Bir Deneydeki Çoklu Karşılaştırmalar
12.3 Önsel Karşılaştırmalar
12.4 Karşılaştırmalar İçin Güven Aralıkları ve Etki Büyüklükleri
12.5 Sonuçların Rapor Edilmesi
12.6 Planlı-Çoklu Doğrusal Karşılaştırmalar
12.7 Tukey Testi
12.8 Hangi Test?
12.9 Bilgisayar Çözümlemeleri
12.10 Eğilim Analizi

BİR VARYANS ANALİZİNDE MANİDAR F DEĞERİ, bütün evren ortalamalarının eşit olmadığının basit bir işaretidir. Bu bize hangi ortalamaların diğer ortalamalardan farklı olduğunu söylemez. Sonuç olarak, varyans analizi genellikle cevapladığından daha fazla soru doğurur. Bu durumda özel hipotezleri test etme veya manidar olarak farklı olanları ayırmak için tekil ortalamaların veya ortalama kümeleri arasındaki farklılıkların incelenmesi problemiyle karşılaşırız. İlk üç ortalama birbirine $\mu_1 = \mu_2 = \mu_3$ daha sonraki iki ortalama birbirine $\mu_4 = \mu_5$ eşit olmakla birlikte, ilk üç ortalama son iki ortalamadan veya bunların tamamı $\mu_6$'dan farklıdır şeklinde ifadelerde bulunmak isteriz.

Ortalamalar arasındaki farkların test edilebilmesi için birçok teknik bulunmakla beraber bu liste her yıl giderek büyümektedir. Buradaki içerik, en çok kullanılan ve en yararlı olan yöntemlerle sınırlandırılacaktır. Bu konudaki bir tartışma Hochberg ve Tamhane'de (1987) ve Toothaker'da (1991) bulunabilir. Keselman, Holland ve Cribbie (2005) yeni yöntemlerle ilgili muhteşem bir derleme sunmuşlardır. Games'in (1978a, 1978b) çalışması yardımcı olabileceği gibi Games and Howell'ın (1976) çalışması da eşit olmayan örneklem büyüklüklerinde yapılan işlemler konusunda faydalı olabilir.

Bu, bu bölümün çeşitli basımlarda nasıl değiştiğinin anlaşılması konusunda okuyucuya yardımcı olabilir. Değişiklikler genel olarak insanların deneysel sonuçlara bakışını yansıtmakta ve en kullanışlı yöntemlere odaklanmaktadır. Bu bölümün orijinalinde en yaygın olarak kullanılan birkaç yöntem kapsanmıştı. Daha sonra zaman ilerledikçe yöntemleri eklemeye devam ettim ve ortalamalar arasında tekil karşılaştırma yollarına ilişkin sayının artışına odaklandım. Ancak bu basımda günümüzde neredeyse herkesin kullandığı birkaç yöntemin kapsanmasına geri dönüyorum. Şu gerçeği vurgulamak isterim ki; dikkatimizi gerçekten umursamamız gereken farklılıklara odaklamamalıyız, bulgular bölümümüzü test edebileceğimiz ancak ilgili olmayan diğer farklılıklarla doldurmamalıyız. Bu felsefe etki büyüklüklerinin hesaplanması ve uygun hata teriminin seçimi ile devam eder. Eğer birkaç özel karşılaştırmayla ilgileniyorsanız bu durumda Tukey (desenlendiği amaç için oldukça iyi bir testtir) gibi standart çoklu karşılaştırma testlerinden birini alıp düşünülebilen bütün ikili sıfır hipotezlerinin testi oldukça kötü bir fikirdir. Gücü boşa harcar, sık sık uygun olmayan hata terimlerinin kullanımına yol açar, zayıf etki büyüklüğü ölçüleri verir ve genel olarak açık ve basit olan bulguları karıştırır. Bir şeyi yapabiliyor olmanız, o eylemi yapmanız için her zaman yeterli sebep değildir. Çok sık kullanılan bir yaklaşım olduğu için Tukey testi kapsama alınacaktır fakat ana vurguyu başka bir yere koymak, Tukey'nin isatistiğe ilişkin genel yaklaşımıyla daha çok örtüşmektedir.

## 12.1 Hata Oranları

Çoklu karşılaştırma yöntemleriyle ilgili tartışmaların en önemli konusunu I. Tip hata olasılığı sorusu oluşturur. Alternatif yöntemlerin bulguları arasındaki en büyük farklılık bu hataları nasıl kontrol edecekleri sorusundan gelir. Problem kısmen teknik olsa da gerçekte daha çok, hatanın nasıl tanımlandığı ve maksimum hata oranının ne kadar olmasının beklendiği gibi subjektif sorularla ilgilidir.

Burada hata oranının veya I. Tip hata olasılığının belirtilmesinin iki temel yolu ayırt edilecektir.[1] (Daha sonra son yıllarda yoğun ilgi toplayan, Yanlış Buluş Oranı adı verilen alternatif hata oranı görüşleri hakkında tartışacağız.) Böyle yaparak Tukey'nin 1953'teki yayımlanmamış çok önemli bir makalesinden bu yana aşağı yukarı standart hale gelmiş bir terminolojiyi kullanacağız (bkz. Ryan, 1959; O'Neil ve Wetherill, 1971).

---

[1] Bundan başka her deneyin hata oranı (rate per experiment-PE) olarak adlandırılan ve bir ikilli karşılaştırmalar kümesindeki beklenen I. Tip hata olasılıkları *sayısı* anlamına gelen bir hata oranı daha vardır. Her deneyin hata oranı bir olasılık değeri değildir ve bunu kontrol etmeye genellikle doğrudan niyetlenmeyiz. $PE = c\alpha$ olarak kolayca hesaplanabilir. Burada $c$ karşılaştırma sayısını ve $\alpha$ her karşılaştırmanın hata oranını temsil eder.

## Her Karşılaştırmanın Hata Oranı

**her karşılaştırmanın hata oranı**

Daha öncesinde **her karşılaştırmanın hata oranı** (*PC*) terimini kullanmıştık ve bunun bir miktar dikkat edilmesi gereken bir konu olduğunu belirtmiştik. Bu, herhangi bir ikili karşılaştırmadaki I. Tip hata olasılığıdır. Örneğin, 2 grubu *t* testi kullanarak karşılaştırırsak ve ardından sıfır hipotezini, *t* değerimiz $t_{,05}$'i geçtiğinden reddedersek bu durumda her karşılaştırma için ,05 hata oranı ile çalışıyoruz demektir.

## Aileboyu (Familywise) Hata Oranı (*AHO/FW*)

Grup ortalamalarımız arasındaki karşılaştırmaları tamamladığımızda, bir sonuçlar kümesine (genellikle aile olarak adlandırılır) ulaşırız. Örneğin, aile aşağıdaki ifadeleri kapsayabilir.

$\mu_1 < \mu_2$
$\mu_3 < \mu_4$
$\mu_1 < (\mu_3 + \mu_4)/2$

**aileboyu hata olasılığı (AHO)**

Bu sonuçlar ailesinin en az bir I. Tip hata barındırma olasılığı **aileboyu hata oranı *(AHO)*** olarak adlandırılır.[2] İnceleyeceğimiz süreçlerin birçoğu AHO hatasını kontrol etmeye yönelmiştir. Bunların AHO'yu kontrol etme amacı gütmeyenleri bile, AHO'nun muhtemel düzeyine göre değerlendirilmektedir.

Bir deneyde sadece bir karşılaştırma yapılıyorsa, her iki hata oranı (bir karşılaştırmadaki hata oranı ve aileboyu hata oranı) da aynı olacaktır. Bununla beraber yapılacak olan karşılaştırmaların sayısı arttıkça iki oran farklılaşacaktır. Eğer $\alpha'$ herhangi bir karşılaştırmanın hata oranı olarak kabul edilir ve *c* de yapılacak karşılaştırmaların sayısını belirtirse bu durumda:

| | |
|---|---|
| Her karşılaştırmanın hata oranı: | $\alpha = \alpha'$ |
| Ailesel hata oranı: | $\alpha = 1 - (1 - \alpha')^c$ |
| | (Eğer karşılaştırmalar bağımsız ise) |

Eğer karşılaştırmalar bağımsız değilse, her karşılaştırmaya ilişkin hata oranı değişmemekle beraber aileboyu hata oranı bundan etkilenir. Çoğu durumda $1-(1 - \alpha')^c$ AHO için makul bir yaklaşımı temsil eder. AHO üzerindeki $HKHO \leq \mathbf{AHO} \leq c\alpha$; pek çok mâkul durumda AHO, $c\alpha$'nın civarındadır. Bu durum, Bonferroni testi düşünüldüğünde önemli hale gelir.

## Sıfır Hipotezi ve Hata Oranları

Buraya kadar olan kısımda *tam* veya *geniş kapsamlı sıfır hipotezi* ($\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$) için konuştuk. Bu varyans analizinde genel olarak test edilen sıfır hipotezidir. İnsanlar tamamında olmasa da çoğu deneyde, tam bir sıfır hipoteziyle ilgilenmek yerine ($\mu_1 = \mu_2 = \mu_3$, $\mu_4 = \mu_5$, $\mu_6 = \mu_7$) gibi daha sınırlandırılmış çeşitli alt test kümeleri arasındaki farklara yönelik yokluk hipotezleriyle ilgilenirler. Eğer buna benzer bir durum varsa, süreç daha karmaşık bir hâl alır ve evren ortalamaları arasındaki yapıyı bilmeden AHO'nun belirlenebilme olasılığı bazen mümkün olmayabilir. Bu durum, hata oranları tartışılırken hesaba katılmalıdır.

---

2 Bu hata oranı özellikle eski kaynaklarda "deneysel" hata oranı olarak anılır. Bununla beraber Tukey'nin "ailesel" terimi daha çok kullanılagelmiştir. Daha kompleks varyans analizlerinde deney kavramı; sıklıkla farklı karşılaştırma ailelerini içerir nitelikte düşünülmektedir.

## Bir Önsel Karşılaştırmaya Karşı Planlı-Çoklu Karşılaştırmalar

**önsel karşılaştırmalar**

**çoklu karşılaştırmalar**

Bu kitabın önceki basımlarında geleneğe dikkatlice bağlı kalarak, veri toplanmadan önce seçilen anlamındaki **önsel karşılaştırmalar** ve veri toplandıktan sonra deneyci tarafından planlanan anlamındaki planlı **çoklu** (**post hoc) karşılaştırmalar** kavramlarını birbirinden ayırdım. Her iki karşılaştırma da ortalamaların birbirine yakınlığı veya uzaklığıyla ilgilenir. Bu geleneksel bir ayrım olmakla birlikte, karşılaştırmaları yapanların bazılarına oldukça önemsiz görünebilir. Pratikte gerçek farklılık teorik veya pratikteki doğası gereği kasıtlı olarak yapılan analizlerle, bütün olası ikili ortalama farklarına yönelik karşılaştırmalar (sadece ortalama farklarına kilitlenmeden) arasındaki ayrışma olarak kendini gösterir. Biraz bulanıklık yaratsa da, önsel ve çoklu olarak ayrım yapmaya devam edeceğim, çünkü bu kavramlar materyali iyi organize etmekte ve sıkça anılmaktadır.

Örnek olarak daha önce gördüğünüz, önceki bölüm Alıştırma 11.8'de yer alan Foa, Rothbaum, Riggs ve Murdock (1991) tarafından yapılmış çalışmaya dönebiliriz. Foa ve arkadaşları (1991) tecavüz kurbanlarına ilişkin dört farklı terapiyi değerlendirmiştir. Strese Karşı Aşılama Terapisi (SKAT) *(Stress Inoculation Therapy (SIT))* grubu stresle baş edebilmek için eğitim almıştır. Uzatılmış Pozlama (UP) grubu ise hafızalarındaki olaylar üzerine tekrarlı bir şekilde gitmişlerdir. Destekleyici Danışma (DD) grubuna genel bir problem çözme tekniği (General Problem Solving) öğretilmiştir. Son olarak Bekleme Listesi (BL) (Waiting List) (WL) grubu ise kontrol grubu olarak herhangi bir terapi almamıştır.

Bu çalışmayı yürüttüğünüzü düşünün. Gerçek amacınız Strese Karşı Aşılama Terapisi (SKAT) ile Uzatılmış Pozlama (UP) gruplarını karşılaştırmak, buna ek olarak bu iki terapinin kontrol durumlarına (SC ve BL) göre daha etkili olup olmadığını kontrol etmek istiyorsunuz. Fakat daha sonra sonuçlara baktığınızda oldukça etkisiz olmasını beklediğiniz SC'nin, Bekleme Listesi kontrol grubuna göre çok daha iyi çıktığını düşünün. Bu sizin daha önce yapmayı hiç düşünmediğiniz ancak yapmanız gereken bir karşılaştırmadır. Bu durumda ne yaparsınız? Bu artık basit bir sorudur? İlk iki karşılaştırma açık bu araştırmayı yapma amacınızdan kaynaklı önsel niteliktedir. Ancak üçüncüsü, planlı-çoklu karşılaştırma olarak nitelendirilebilir, çünkü sonuçları görünceye kadar böyle bir şeye niyetlenmemiştiniz. Geleneksel cevap, bütün ikili karşılaştırmaları yapmak için çoklu test kullanmanız yönündedir. Aksi halde, çünkü kontrol grupları arasındaki farklılıklardan dolayı I. Tip hata yapma riskine girersiniz. Alternatif bir yaklaşım şöyle olabilir: Teoriye dayanmasa da son karşılaştırmaya önsel olarak davranın, manidar çıkması durumunda özellikle olasılık değeri kritik değerin yakınlarındaysa biraz daha dikkatlice inceleyin. Her iki yaklaşımın neler getireceğini düşüneceğiz.

Önsel bir test hakkında konuşurken genel olarak kastettiğimiz, nispeten küçük bir karşılaştırma kümesi olduğu, aklımızda tutmamız gereken önemli bir konudur. Eğer ortalamalar arasındaki farka yönelik *bütün* ikili karşılaştırmaları yapıyorsak, analizin daha önceden düşünülüp düşünülmemesi bir fark yaratmaz. (Gerçekten bütün mümkün karşılaştırmaları yapmak isteyip istemeyeceğinizi merak ederim.)

## Genel *F*'nin Manidarlığı

Grup ortalamaları arasında çoklu karşılaştırmalar yapmadan önce manidar bir genel *F* değeri konusunda ısrarcı olunup olunmaması sorusuna yönelik çeşitli tartışmalar vardır. Geçmişteki genel tavsiye, manidar bir grup etkisi yokken, ikili karşılaştırmaların yapılmasının uygun olmayacağı şeklindeydi. Gerçekte Bölüm 12.4'te yer alan Fisher'ın en küçük manidar farklılıklar testindeki hata oranlarının rasyoneli için genel bir manidarlığa ihtiyaç vardır.

Ancak genel tavsiyenin yanlış olduğu durumlar vardır. Çoğu çoklu karşılaştırma sürecimizin arkasında yer alan mantık, özel karşılaştırmalar yapmadan önce genel bir manidarlığa gereksinim duymaz. İlk olarak, genel test ile çoklu karşılaştırma testi tarafından test edilen hipotezler oldukça farklı güç düzeyleriyle birbirinden oldukça farklıdır. Örneğin, genel *F*, gruplar arasındaki farklılıkları grupların serbestlik derecesine göre dağıtır. Bu çeşitli grup ortalamalarının birbirine eşit olduğu ancak başka bir ortalamadan farklı olduğu durumlarda, genel *F* değeri üzerinde seyreltici bir etki yaratır. İkinci olarak, genel bir manidarlık istenmesi *AHO* değerini değiştirecektir ve çoklu karşılaştırma testlerini tutucu hale getirecektir. Bu testler genel *F* değeri göz önüne alınmadan desenlenmiş ve manidarlık düzeyleri belirlenmiştir.

Wilcox (1987a), bu durumu düşünerek "genel bir *F* testi uygulamak için küçük bir neden vardır" (s. 36) şeklinde tavsiyede bulunmuştur. Wilcox, *F* değeri hesaplamadan doğrudan çoklu karşılaştırmalara geçmektedir. Diğerleri de büyük ölçüde buna benzer sözler söylemişlerdir. Bu durum geçmiş için biraz uç bir durum olsa da, bu şekilde vurgu yapılmıştır. Bununla beraber, günümüzde bu durum, 20 yıl önce göründüğü kadar uç bir durum olarak görülmemektedir. Hatırlarsanız tipik çoklu karşılaştırmalar manidar bir genel *F* değerine ihtiyaç duymamaktaydı ve gruplar arasındaki fark bu *F* değerine ihtiyaç olmadan incelenebilmekteydi. Öyleyse neden hâlâ *F* değerine ihtiyaç hissediyoruz? Düşünebildiğim tek neden "gelenek" ki bu da oldukça güçlü bir kaynaktır. Her durumda $KO_{\text{hata}}$ hesaplamaya kadar gideceğinden fazladan bir adım olarak geniş kapsamlı *F* hesaplanacaktır.

## 12.2 Morfin Toleransı Üzerine Basit Bir Deneydeki Çoklu Karşılaştırmalar

Çeşitli yöntemler tartışılırken, bütün yaklaşımların uygulanabileceği bir veri kümesi yardımcı olabilir. Siegel'ın (1975) morfin toleransı üzerine yaptığı bir deneye benzer bir çalışmayı ele alabiliriz. Hayalî veri kullanılmasına ve ortam koşulları özgürce tanımlanmasına rağmen ortalamalar (ve ortalamalar arasındaki farkların manidarlığı) Siegel'ın makalesinde olduğu gibidir. Bu çalışmayı daha ayrıntılı bir şekilde açıklamak gerekli olabilir. Hipotezleri ve uygulama etiketlerini anlamak için zaman ayırmanız size avantaj sağlayacaktır.

Morfin ağrı yatıştırmak için sıklıkla kullanılan bir ilâçtır. Tekrarlı morfin uygulamaları, morfin toleransına neden olur ki morfinin zaman içerisindeki ağrı hafifletme etkisi azaldıkça azalır. (Bunu baharatlı yemek tüketiyorsanız siz de deneyimlemişsinizdir. Başladığınız acı seviyesine göre her defasında daha fazla acı yiyebildiğinize şahit olursunuz.) Morfin toleransı için yaygın olarak düzenlenen bir deneysel etkinlik, bir farenin rahat edemeyeceği kadar sıcak bir yüzeye yerleştirilmesini içerir. Sıcaklık çok rahatsız edici bir seviyeye ulaştığında, fare pençelerini yalar. Sadece bir morfin enjeksiyonu alan fare, acı hassasiyetinin azaldığının göstergesi olacak şekilde, tipik olarak daha uzun pençe–yalama tepki zamanı gösterir. Tekrarlanan morfin enjeksiyonlarıyla birlikte morfin toleransının gelişme göstergesi olarak pençe–yalama tepki zamanı kısalır (artan hassasiyet veya azalan duyarsızlaşma göstergesi olarak).

Siegel morfinden başka ilâçları içeren bir dizi durumda *koşullanmış* (öğrenilmiş) ilâç tepkilerinin, koşullanmamış (doğal) ilaç tepkilerinin tersi yönde hareket ettiğine dikkat etmiştir. Örneğin, atropin enjekte edilen bir hayvanın genellikle tükürük salgılamasında belirgin bir düşüş olur. Tekrarlı antropin enjeksiyonlarından sonra (aynı fiziksel koşullarda) fizyolojik tuz solüsyonu (bir etkisi olmayan) aniden enjekte edilirse, hayvanın tükrük üretiminde *artış* görülür. Bu hayvanın, atropinin beklenen etkisini telâfi etmesi gibidir. Bu tip

çalışmalarda görülmüştür ki denemeler boyunca öğrenilmiş tamamlayıcı mekanizma gelişir ve ilâcın etkisini dengeler. (Bu etkiyi baharat kullanarak yediğiniz bir yemekte baharat kullanmayı bıraktığınızdaki durum olarak deneyebilirsiniz. Üzüm kurusu, yediğinizde çok yumuşak tatta olsa da, kahvaltıda yediğiniz zaman tadı yumuşak olmayabilir - Üzüm kurusuna baharat koymayacağınızı umarım.)

Siegel böyle bir sürecin morfin toleransını açıklayabileceğini teorik hale getirmiştir. Bunu da şu şekilde nedenlemiştir: Eğer hayvanlara morfin enjekte ettikten sonra sıcak bir yüzeye koymaya devam ederseniz morfin toleransı gelişecektir. Eğer hayvana tekrar aynı enjeksiyonu yapıp benzer uygulamayı devam ettirirseniz, toleransın tamamen gelişmesinden kaynaklı olarak hayvan daha önce hiç morfin almamış bir hayvan kadar acıya duyarlı olacaktır. Normal morfin enjeksiyonundaki düzenekte olduğu gibi, hayvana morfin yerine fizyolojik tuz çözeltisi *enjekte* edilirse, *tekrarlı morfin uygulamalarındaki* koşullanmış aşırı duyarlılık morfinin varlığıyla dengelenmez ve hayvan çok kısa pençe–yalama tepki zamanı gösterir ve hassasiyet artar. Siegel bir düzenekte hayvana tekrarlı morfin uygulamasında bulunulduktan sonra yeni bir düzenekte morfin test edilirse, yeni düzeneğin morfini dengelemek için koşullanmış tamamlayıcı aşırı duyarlılığı ortaya çıkarmayacağını belirtmiştir. Sonuç olarak hayvan ilk defa morfin enjekte edilmiş bir hayvan gibi tepkide bulunur. Eroin bir morfin türevidir. Tolerans geliştirmekten kaynaklı olarak yüksek dozlarda eroin alan bir eroin bağımlısını hayal edin. Düzeneğin değiştirilmesi sonucu, bu bağımlının bu yüksek doza olan tepkisi birden yüksek toleranslı biri yerine (hoşgörü yerine) ilk kullanan birinin tepkisi gibi olursa sonuç sıklıkla ölümcül olacaktır. Burada ciddi bir konu hakkında konuşuyoruz, uyuşturucu aşırı dozları genellikle yeni düzeneklerde oluşur.

Bu kitabın okuyucularının çoğunun doğmadığı, 30 yıl önce yapılmış bu deneyin ilgi çekici olabilmesi için çok eski olduğunu düşünebilirsiniz. Google'da yapılacak basit bir tarama ile Siegel'ın çalışmasından türetilmiş birçok güncel çalışma ile karşılaşılabilir. Özellikle ilgi çekici olanlardan birisi Mann-Jones, Ettinger, Baisden ve Baisden tarafından yapılan ve Dekstromethorfan adlı ilâcın morfin toleransını etkisizleştirmesini içeren çalışmadır. Bu Dekstromethorfan adlı ilâcın öksürük şurubunun önemli bir bileşeni olduğunu öğrendiğimizde ilgi çekici hale gelir. Bu, eroin bağımlılarının öksürük şurubu almayı, yeni ortamlarda eroin almaktan daha çok istememelerini tavsiye eder. Bu çalışmaya www.eou.edu/psych/re/morphinetolerance.doc adresinden ulaşılabilir.

Siegel deneyinin bizim olan versiyonu sadece özetlenen öngörülere dayanmaktadır. Deney beş farklı fare grubunu içerir. Her grup dört deneme alır ama veri sadece kritik olan dördüncü (test) denemesinden elde edilir. Gruplar ilk üç denemedeki işlemleri ve son denemedeki işlemi belirtecek şekilde desenlenmiştir. Grup M-M ilk üç denemede morfin almıştır, dördüncü denemede de aynı test durumu içinde bu tekrarlanmıştır. Bu standart morfin toleransı grubudur, morfin toleransı çok çabuk geliştiğinden, dördüncü denemede normal veya normale yakın düzeyde acı hassasiyeti görmeyi bekleriz. Grup M-S ilk üç denemede morfin, ancak dördüncü denemede tuzlu solüsyon almıştır (aynı test düzeneği içerisinde). Bu hayvanların ağrı uyarıcısına karşı aşırı duyarlı olmaları beklenir; çünkü, morfinin tamamlayıcı etkileri olmadığından koşullanmış aşırı duyarlılıkları dengelenmeyecektir. Grup $M_{(kafes)}$-M (kısaltılmış hali $M_c$-M) ilk üç denemede morfini kendi kafeslerinde, dördüncüsünü onlar için yeni olan standart test düzeneğinde almışlardır. Bu grup; morfin enjeksiyonu ile ilişkilendirilen durumlar, test durumunda olmadığı için Siegel'ın modeline göre morfin toleransı göstermemelidir. 4. grup olan S-M ilk üç denemede tuzlu solüsyon almışlar, dördüncü denemede ise morfin almışlardır. Bu hayvanların acıya en az duyarlılığı göstermeleri beklenir; çünkü morfin toleransının gelişmesi için bir seçenek yoktur. Son olarak grup S-S dört denemede de tuzlu solüsyon almıştır.

**Tablo 12.1** Veri ve morfin toleransı üzerine analiz

**(a) Veri**

| | M-S | M-M | S-S | S-M | $M_c$-M |
|---|---|---|---|---|---|
| | 3 | 2 | 14 | 29 | 24 |
| | 5 | 12 | 6 | 20 | 26 |
| | 1 | 13 | 12 | 36 | 40 |
| | 8 | 6 | 4 | 21 | 32 |
| | 1 | 10 | 19 | 25 | 20 |
| | 1 | 7 | 3 | 18 | 33 |
| | 4 | 11 | 9 | 26 | 27 |
| | 9 | 19 | 21 | 17 | 30 |
| Ortalama | 4,00 | 10,00 | 11,00 | 24,00 | 29,00 |
| Std. Sapma | 3,16 | 5,13 | 6,72 | 6,37 | 6,16 |

**(b) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| İşlem | 4 | 3497,60 | 874,40 | 27,33* |
| Hata | 35 | 1120,00 | 32,00 | |
| Toplam | 39 | 4617,60 | | |

*p <. ,05

Eğer Siegel'ın modeli doğru ise, Grup S-M en uzun tepki zamanına (en düşük duyarlılık göstergesi olarak) sahip olurken, Grup M–S en kısa tepki zamanına (en yüksek duyarlılık) sahip olmalıdır. Grup $M_c$–M ve Grup S-M benzer olmalıdır çünkü, Grup $M_c$–M'nin ilk üç denemeyi alma koşulları, test denemesi olan sonuncusundan farklıdır. Grup M-M ve Grup S-S orta düzeyde olmalıdır. Grup M-M'nin S-S'ye eşit olması morfin toleransının gelişmesi oranına bağlıdır. Beklenen sonuçlar aşağıdaki gibidir:

$$S - M = M_c - M > M - M \; ? \; S - S > M - S$$

"?" herhangi bir öngörü olmadığı anlamına gelmektedir. Bağımlı değişken pençe-yalama tepki zamanı (saniye cinsinden) ele alınmıştır.

Siegel'ın sonuçlarına yakın olan bu deneyin sonuçları, Tablo 12.1a'da ve tüm varyans analizi sonuçları Tablo 12.1b'de verilmiştir. Grup-içi varyanslar aşağı yukarı eşit (varyansların heterojenliği testi manidar değil), uç değer yoktur. Bu verinin tamamına ait varyans analizi açık olarak manidardır; bu beş işlem grubu arasında fark olduğunu belirtir.

## Etki Büyüklüğü

Bu veri kümeleri için $\eta^2$'yi $KT_{\text{uygulama}}/KT_{\text{toplam}}$ = 3497,60/4617,60 = ,76 biçiminde hesaplayabiliriz. Buna göre uygulamanın tamamına ait değişkenliğin %76'sının işlemlerden (uygulamalardan) geldiğini söyleyebiliriz. Neredeyse $\omega^2$ yansız kestiricisi kadardır, bu kestiricinin değeri

$$\omega^2 = \frac{KT_{\text{işlem}} - (k-1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{397,60 - 4(32)}{4617,60 + 32} = \frac{3369,6}{4649,6} = 0,72$$

Her iki kestirim de göstermektedir ki gruplara yapılan uygulamalar çalışmadaki değişkenliğin önemli bir kısmını oluşturmaktadır.

## 12.3 Önsel Karşılaştırmalar

Tartışmaya önsel karşılaştırmalar ve *t* testi ile başlamamızın iki nedeni vardır. Birinci sırada, ortalama çiftleri arasındaki standart *t* testi mükemmel ve akla uygun bir karşılaştırma yöntemidir. İkincisi, *t* için temel formül ve bu formül üzerinde yapılan küçük düzeltmeler çok sayıda yöntem için (önsel ve çoklu) uygulanabilir niteliktedir. Bunun gözden geçirilmesi yararlı olacaktır.

**karşılaştırmalar**

Gördüğümüz gibi, önsel karşılaştırmalar (**karşılaştırmalar** da denir) veri toplanmadan önce planlanır. Çeşitli önsel karşılaştırma türleri vardır; bunlar üzerinde tartışacağız.

### Çoklu *t* Testleri

Daha önceden planlanan grup çiftlerini karşılaştırmanın en kolay yollarından biri *t* testleridir. Eğer varyansların homojenliği sayıltısı sağlanabilirse, *t* testi yapılırken genellikle yaptığımız varyans analizindeki $KO_{\text{hata}}$ yerine varyansın kendisi veya birleşik varyans kestirimini kullanıp *t* değerini $sd_{\text{hata}}$ serbestlik derecesi üzerinden değerlendirmektir. Varyanslar homojen değilse, ancak örneklem büyüklükleri eşitse bu durumda $KO_{\text{hata}}$'yı kullanmak yerine örneklemlerin varyansını kullanıp *t* değerini 2(n-1) serbestlik derecesi üzerinden değerlendiririz. Son olarak eğer varyanslar heterojense ve örneklem büyüklükleri birbirine eşit değilse varyanslar kullanılır ancak serbestlik derecesi Welch–Satterthwaite yaklaşımı ile düzeltilir (bakınız Bölüm 7). (Bölüm 7'de Hayes ve Cai (2007), bağımsız ortalamaların karşılaştırılmasında *t* testi kullanılırken varyansların birleştirilmesi temel fikrine karşı olacak şekilde tartışmışlardır. Ancak çeşitli gruplarla varyans analizi yaparken, sadece sorudaki gruplardan varyans kestirimi elde edilmesi hatırı sayılır ölçüde güç kaybetmemize yol açar. Bu konuda anlaşmazlıklar olsa da, *t* değerimi hesaplarken eğer örneklem varyansları homojen gibi görünüyorsa, $KO_{\text{hata}}$ ve serbestlik derecesini kullanabilirim, fakat bu düşünceye itiraz edilebilir.)

Çoklu *t* testlerinin ayrım yapmadan kullanılması tipik olarak çoklu karşılaştırmalar için kötü bir örnek olmayı beraberinde getirir. Çeşitli açılardan bu haksız bir eleştiridir. İlgi çekici olabileceğinden dolayı, bütün ortalama çiftlerini karşılaştırmak için bir veri kümesine atlamak oldukça kötü bir şeydir. Aileboyu hata oranı oldukça yüksek olacaktır. Ancak yapacak bir ya da iki karşılaştırmanız varsa ve bu karşılaştırmaları gerçekten daha önceden planladıysanız bunun için *t* testi tavsiye edilebilir. ("Bunları karşılaştırmayı düşünseydim planlardım." diyemezsiniz.) Sadece iki karşılaştırma için, örnek vermek gerekirse, eğer her karşılaştırma $\alpha$ = ,05 düzeyi için yapıldıysa maksimum *AH*O 0,10 olabilir ve her karşılaştırma $\alpha$ = ,01 düzeyi için yapıldıysa maksimum *AH*O 0,02 olabilir. Analizde bireysel farklılıkların oynayabileceği rol için Howell (2008c) incelenebilir.

Daha önce tanımlanan morfin toleransı çalışmasında çoklu *t* testlerini basit bir şekilde kullanamayız, çünkü çok sayıda önemli karşılaştırma düşünülmüştür. (Kendimizi görece daha az sayıda karşılaştırma için sınırlamadığımız sürece, gerçekte bütün ikili karşılaştırmaları yapacak bir çoklu karşılaştırma (planlı çoklu karşılaştırma) yöntemi kullanırız.) Bir örnek vermek gerekirse, teori çerçevesinde önceden açık bir şekilde öngördüğümüz *t* testi ile kolayca test edilebilecek iki temel karşılaştırmayı düşünün. Teori demektedir ki; ilk üç uygulamada morfin almış ancak son uygulamada tuzlu solüsyon almış tavşan, hep tuzlu çözelti alarak test edilmiş bir hayvana göre daha büyük acı duyarlılığı gösterir. Bu, Grup M-S ile Grup S-S'nin karşılaştırılmasını içerir. Daha ileri gitmek gerekirse, teori, Grup $M_c$–M'nin M–M'den daha az acı duyarlılığı göstermesi gerektiğini söylemektedir; çünkü $M_c$ – M daha

önce morfin aldığı bir ortamdan farklı bir ortamda test edilmiştir. Örneklem varyansları benzer ve örneklem büyüklükleri eşit olduğundan, birleşik varyansın kestirimi olarak $KO_{\text{hata}}$ kullanacağız ve sonucu $sd_{\text{artık}}$ üzerinden değerlendireceğiz.

Genel *t* genel formülümüzdeki varyanslar $KO_{\text{artık}}$ ile yer değiştirirse

$$t = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{KO_{\text{artık}}}{n} + \frac{KO_{\text{artık}}}{n}}} = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{2KO_{\text{artık}}}{n}}}$$

yazılabilir. Örneğimizdeki veriyi kullanarak Grup M-S ile Grup S-S karşılaştırıldığında aşağıdakiler elde edilir:

$$\overline{X}_{M-S} = 4{,}00 \quad \overline{X}_{S-S} = 11{,}00 \quad KO_{\text{artık}} = 32{,}00$$

$$t = \frac{\overline{X}_{\text{MS}} - \overline{X}_{\text{SS}}}{\sqrt{\frac{2KO_{hata}}{n}}} = \frac{4{,}00 - 11{,}00}{\sqrt{\frac{2(32{,}00)}{8}}} = \frac{-7}{\sqrt{8}} = -2{,}47$$

Ve grup $M_c$-M ile grup M-M karşılaştırılması için

$$\overline{X}_{Mc-M} = 29{,}00 \quad \overline{X}_{M-M} = 10{,}00 \quad KO_{\text{artık}} = 32{,}00$$

$$t = \frac{\overline{X}_{Mc} - \overline{X}_{M-M}}{\sqrt{\frac{2KO_{\text{artık}}}{n}}} = \frac{29{,}00 - 10{,}00}{\sqrt{\frac{2(32{,}00)}{8}}} = \frac{19}{\sqrt{8}} = 6{,}72$$

bulunur. Elde edilen her iki *t* değeri $t_{,025}(35) = 2{,}03$ kullanılarak değerlendirildikten sonra ilgili sıfır hipotezinin reddedileceği söylenebilir. Tuzlu çözelti ile test edilmiş iki hayvan grubundan, daha önce aynı durumda morfin almış grubun daha yüksek acı duyarlılığı gösterdiği sonucunu çıkarabiliriz. Buna ek olarak morfinin verildiği düzeneğin değiştirilmesi, koşullanmış morfin toleransı etkisini yok etmediği takdirde, manidar derecede azaltmıştır. Her iki yokluk hipotezini $\alpha = ,05$ düzeyinde test ettiğimiz için, hiç öyle görünmemekle beraber eğer *her iki yokluk hipotezi de doğru ise AHO*, 10'a ulaşacaktır. Jones ve Tukey'de (2000) olduğu gibi, yokluk hipotezlerinin doğru olma olasılığının veya manidar farklılığı yanlış yönde bulmamızın son derece düşük olduğu bu durumda, hata olasılığı neredeyse ,05 olacaktır. Ortalama çiftleri arasındaki karşılaştırmalar hakkında konuşurken avantajlarını ve dezavantajlarını akılda tutmamız önemlidir.

Biraz önce kullandığımız temel *t* testi bundan sonrakilerin neredeyse hepsinin temelini oluşturur. Formülü oraya ya da buraya çekiştirdiğimde, çeşitli tablolar ve karar kuralları kullandığımda, hatta formülünü değiştirip adına *q* dediğimde, temel *t* testimiz var olacaktır.

## Doğrusal Karşılaştırmalar

*t* testinin kullanımı, çoklu doğrusal karşılaştırmalar[3] olarak bilinen çok daha genel bir tekniğin özel bir halidir. *t* testi bir grup ile diğer grubu karşılaştırmamıza izin verir ancak doğrusal karşılaştırmalar bir grup veya *gruplar kümesini* başka bir grup veya gruplar kümesi ile karşılaştırmamıza imkân sağlar. Çoklu doğrusal karşılaştırmalara ilişkin hesaplama tekniklerini çoklu testlerle kullanabildiğimiz gibi önsel testler ile de kullanabiliriz, ancak burada en çok kullanıldığı şekli yani önsel testler altında tartışacağız. Karşılaştırmalar için olan testlerin standart *t* testinin uzantısı olduğunu unutmayınız.

3 "Contrast" ve "comparison" kelimelerinden birinin yerine diğeri kullanılabilmektedir.

**doğrusal (lineer) kombinasyon**

Doğrusal farklılıkları tanımlamak için, öncelikle doğrusal kombinasyonun tanımlanması gerekir. Ortalamaların **doğrusal (lineer) kombinasyon**u aşağıdaki formdadır;

$$L = a_1\overline{X}1 + a_2\overline{X}2 + ... + a_k\overline{X}k = \sum a_j\overline{X}j$$

Bu eşitlik basitçe bir doğrusal kombinasyonun işlemlere (uygulamalara) ait ağırlıklandırılmış toplamıdır. Eğer, örneğin, tüm $a_j$'lerin toplamı 1'e eşit ise $L$ ortalamaların toplamı, diğer taraftan, bütün $a_j$'ler 1/$k$'ya eşitse $L$ ortalamaların ortalaması olur.

**doğrusal karşılaştırma**

$\sum a_j = 0$ sınırlaması yapılırsa bir doğrusal kombinasyon, **doğrusal karşılaştırma** adını alır. Geleneksel olarak bunun bir doğrusal karşılaştırma olduğunu belirtmek için "$L$" yerine psi ($\psi$) harfi kullanılır. Uygun $a_j$ seçimleri ile bir doğrusal karşılaştırma çok faydalı olabilir. Örneğin $t$ testi ile aynı sonucu verecek şekilde bir ortalamayı diğer bir ortalama ile karşılaştırmak için veya bir koşulun ortalamasının çeşitli koşulların ortalamalarının kombinasyonuyla gösterimi için kullanılabilir. Örnek, üç ortalama ($\overline{X}_1, \overline{X}_2$ *ve* $\overline{X}_3$) düşünün. $\sum a_j = 0$ olacak şekilde $a_1 = 1$, $a_2 = -1$ ve $a_3 = 0$ olsun,

$$\psi = (1)\overline{X}_1 + (-1)\overline{X}_2 + (0)\overline{X}_3 = \overline{X}_1 - \overline{X}_3$$

ifadesi bir doğrusal karşılaştırmadır. Burada ($\psi$), birinci ve ikinci gruba ait ortalamaların farkıdır. Diğer taraftan, $a_1 = 1/2$, $a_2 = 1/0$ ve $a_3 = -1$ alınırsa elde edilen

$$\psi = (1/2)\overline{X}_1 + (1/2)\overline{X}_3 + (-1)\overline{X}_3 = \frac{\overline{X}_1 + \overline{X}_2}{2} - \overline{X}_3$$

ifadesi de bir doğrusal karşılaştırmadır. Bu doğrusal karşılaştırma ($\psi$), birinci ve ikinci gruplara ait ortalamaların ortalamasının üçüncü grup ortalamasından farkı olarak tanımlanmıştır.

## Doğrusal Karşılaştırmalar İçin Kareler Toplamı

Doğrusal karşılaştırmaların avantajlarından biri de, kareler toplamına kolaylıkla dönüştürülebilmesi ve bir dizi işlemin (uygulamanın) ortalamaları arasındaki farkın kareleri toplamını temsil edebilmesidir. Eğer

$$\psi = a_1\overline{X}1 + a_2\overline{X}_2 + ... + a_k\overline{X}_k = \sum a_j\overline{X}_j$$

yazarsak

$$KT_{\text{doğ.fark}} = \frac{n\psi^2}{\sum a_j^2} = \frac{n(\sum a_j\overline{X}_j)^2}{\sum a_j^2}$$

eşitliği elde edilir. Bu ifade, n her uygulamadaki puan sayısını göstermek üzere $KT_{\text{işlem}}$'in 1 serbestlik derecesinde bir bileşenidir.[4]

$n = 10$ $\quad \overline{X}_1 = 1{,}5$ $\quad \overline{X}_2 = 2, 0$ $\quad \overline{X}_3 = 3{,}0$

olan işlemler olsun, varyans analizi için,

---

4 Eşit olmayan örneklem büyüklükleri için $KT_{\text{islem}} = \dfrac{\psi^2}{\sum(a_j^2/n_j)}$

$$KT_{\text{işlem}} = n\sum(\overline{X}_j - \overline{X}_{..})^2 = 10\left[(1,5-2,167)^2 + (2-2,167)^2 + (3-2,167)^2\right]$$
$$= 10\left[0,4449 + 0,0278 + 0,6939\right] = 11,667$$

elde edilir. 1 ve 2 numaralı işlemlerin ortalamasını 3 numaralı işlemle karşılaştırmak isteyelim.

$$\psi = \sum a_j \overline{X}_j = (1/2)(1,5) + (1/2)(2,0) + (-1)(3,0) = -1,25$$
$$KT_{karşılaştırma} = \frac{n\psi^2}{\sum a_j^2} = \frac{10(-1,25)^2}{1,5} = \frac{15,625}{1,5} = 10,417$$

Bu kareler toplamı tüm kareler toplamının, $\text{İşlem}_{\text{işlem}}$, 1 serbestlik derecesinde bir bileşenidir. 1 serbestlik derecemiz var, çünkü gerçekte iki niceliği (ilk iki ve üçüncünün ortalamaları) karşılaştırıyoruz.

Şimdi İşlem 1 ile İşlem 2 ek olarak üçüncü bir doğrusal farklılık elde ettiğimizi düşünelim. $a_1 = 1$, $a_2 = -1$ ve $a_3 = 0$ olsun.

$$\psi = \sum a_j \overline{X}_j = (1)(1,5) + (-1)(2) + (0)(3,0) = -0,5$$
$$KT_{karşılaştırma} = \frac{n\psi^2}{\sum a_j^2} = \frac{10(-0,5)^2}{2} = \frac{2,5}{2} = 1,25$$

Bu *KT*'da $\text{KT}_{\text{işlem}}$'in 1 serbestlik derecesinde bir bileşenidir. Buna ek olarak kısmi farklılıklar seçilerek çalışıldığından

$$KT_{\text{işlem}} = KT_{\text{karşılaştırma1}} + KT_{\text{karşılaştırma2}}$$

$$11,667 = 10,417 + 1,25$$

olacaktır. Her iki farklılık da tüm kareler toplamına hizmet eder ve bütün serbestlik dereceleri işlemlere yüklenebilir. Buna $KT_{\text{işlem}}$'in *tam olarak* **parçalanma**sı deriz.:

**parçalanma**

## Katsayıların Seçilmesi

Bir önceki örnekte katsayıları neden öyle seçtiğimiz oldukça açıktı. Sorudaki doğrusal farklılıkları mantıksal olarak gerçekleştirebilmek için işlem ortalamaları ağırlıklandırıldı. Beş eşit büyüklükte grubumuz olduğunu düşünün ve ilk üçü ile son ikisini karşılaştırmak isteyelim. $\sum a_j = 0$ olacak şekilde bir $a_j$ kümesine ihtiyacımız vardır. En basit kural, iki uygulama kümesi yaratmak ve her kümeye o kümenin eleman sayısının tersi sayıyı ağırlık olarak atamaktır. Daha sonra kümelerden birindeki katsayılara keyfî olarak eksi verilebilir. Örneğin, aşağıdaki ortalamaları ele alalım:

$$\overline{X}_1 \quad \overline{X}_2 \quad \overline{X}_3 \quad \overline{X}_4 \quad \overline{X}_5$$

$\overline{X}_1, \overline{X}_2, \overline{X}_3$ birleşimi ile $\overline{X}_4$ ve $\overline{X}_5$ birleşimini karşılaştırmak istiyoruz. İlk küme 3 ortalama içermektedir, böylece $\overline{X}_1, \overline{X}_2, \overline{X}_3$ için $\alpha_j = 1/3$ olacaktır. İkinci küme 2 ortalama içerdiğinden, $\overline{X}_4$ ve $\overline{X}_5$ için $\alpha_j = 1/2$ alınacaktır ve 1/2'ler negatif olacaktır. Öyleyse elimizdekiler

| **Ortalamalar:** | $\overline{X}_1$ | $\overline{X}_2$ | $\overline{X}_3$ | $\overline{X}_4$ | $\overline{X}_5$ | |
|---|---|---|---|---|---|---|
| $\mathbf{a_j}$**:** | 1/3 | 1/3 | 1/3 | –1/2 | –1/2 | $\sum a_j = 0$ |

Böylece $\sum a_j \overline{X}_j$ terimi, görebileceğiniz gibi, ilk üç terimin ortalamasından son iki terim ortalamasının farkına indirgenmiş olur ve $1/3(\overline{X}_1 + \overline{X}_2 + \overline{X}_3) - 1/2(\overline{X}_4 + \overline{X}_5)$ ifadesiyle yazılabilir.

(Eğer Siegel'ın morfin üzerine olan deneyine geri döner, ilk üç grubu beraber ve son iki grubu beraber ele alırsanız ve birleştirilmiş uygulamaların ortalamalarına bakarsanız, bu sistemin neden mantıklı geldiğini anlarsınız.)[5]

Tam sayıları kullanarak aynı sonucu alabileceğiniz başka yollar da vardır. Kesirleri karışık bulduğumdan dolayı alternatif yöntemleri kullandım. Ancak burada yaptığım gibi, *mutlak değerleri* toplamı 2'ye eşit olan kesirleri kullanmanın, etki büyüklüğünü hesaplarken avantajları vardır. Katsayılar kümesinin toplamının 2'ye eşit olduğu küme **standart küme** olarak da adlandırılır.

**standart küme**

## Manidarlık Testi

Doğrusal karşılaştırmaların kolaylıkla 1 serbestlik derecesinde kareler toplamına dönüştürülebildiğini gördük. Bu kareler toplamına diğer kareler toplamına davranıldığı gibi davranılabilir. Bunlar aynı zamanda 1 serbestlik derecesine sahip olduklarından (hep iki nicelik karşılaştırılmaktadır) kareler ortalaması da olabilirler ve $KO_{\text{hata}}$ ile bölünerek $F$ üretilebilir. Çünkü *bütün* farklılıklar 1 serbestlik derecesine sahiptir.

$$F = \frac{KO_{\text{karşılaştırma}}}{KO_{\text{hata}}} = \frac{n\psi^2/\sum a_j^2}{KO_{\text{hata}}} = \frac{n\psi^2}{\sum a_j^2/KO_{\text{hata}}}$$

Bu $F$ bir ve $sd_{\text{hata}}$ serbestlik derecesine sahip olacaktır. Eğer kendinizi $t$ ile daha rahat hissedecekseniz, $F$'nin karekökünü alıp $sd_{\text{hata}}$ serbestlik derecesinde $t$'ye sahip olabilirsiniz.

Örneğimiz için morfinin en çok etkileyeceği daha önce morfin almayan (S–M) ve morfini farklı bir biçimde alan ($M_c$-M) grupları ile diğer 3 grubun (M–M, S–S ve M–S) ortalamalarını önsel olarak karşılaştırmak istediğimizi farz edelim. Buna ek olarak $t$ testi için yapılan önceki tartışmadaki nedenlerden dolayı grup $M_c$-M ile grup M–M ve grup M-S ile S–S'yi karşılaştırmayı planlayalım. Son olarak Grup M-M ile Grup S-S karşılaştırması istenmiştir ki hep morfin alıp morfin toleransı gelişen hayvanların dördüncü dönemden sonra hep tuzlu çözelti alan hayvanlardan farklılaşıp farklılaşmadığı görülsün. (Gördüğümüz gibi bu dört farklılık birbirinden bağımsız değildir ancak önemli soruları cevaplandırır.) Analizler Tablo 12.2'de gösterilmiştir.

Bütün F değerleri $F_{.05}(1,35) = 4{,}12$'ye karşı değerlendirilecektir. Beklendiği gibi, ilk üç karşılaştırma manidardır. Dört denemede morfin toleransının tamamen gelişmiş olduğunun göstergesi olarak Grup M–M ile S–S'i karşılaştıran dördüncü karşılaştırma manidar değildir. (Burada dikkatli olunuz, mümkün olmadığını bildiğimiz gibi sıfır hipotezini kanıtlayabilecekmişim gibi hareket ediyorum.) Dikkat ediniz, 2 ve 3'ün karşılaştırılması $t$ testi ile test ettiğimiz aynı hipotezi test etmektedirler. Bu zıtlıklar için $F$'lerin karekökünü alırsanız bunlar 6,72 ve 2,47'ye eşit olacaktır ve bu değerler daha önceden $t$ kullanarak elde ettiğimiz değerlerdir. Bu basitçe göstermektedir ki $t$ testleri doğrusal karşılaştırmaların özel halidir.

---

5 Eğer gruplarımızda farklı sayıda denekler varsa, katsayılarımızı daha farklı yollardan elde etmemiz gerekebilir. Eğer örneklem büyüklükleri, birkaç deneğin kaybolması gibi şart olmayan farklılıklar gösteriyorsa yukarıdaki yol uygun olacaktır. Bu herhangi bir grup ortalamasını diğerinden birkaç birey daha fazla diye ağırlıklandırmaz. Ancak örneklem büyüklükleri tesadüfî değil de sistematik olarak farklılaşıyorsa ve daha geniş örneklemli gruba daha fazla ağırlık vermek istiyorsak bu durumda daha farklı bir şeyler yapmamız gerekir. Farklı örneklem büyüklüğünün önemli rol oynayabileceğini hayal edebildiğim birkaç yer var ve bu yaklaşımı kitabın bu baskısından çıkardım. Bunu daha önceki basımlarda ve Web'de daha önceden belirtilen yerlerde bulabilirsiniz.

**Tablo 12.2** Morfin verisine ilişkin bir önsel karşılaştırma kümesi

| Gruplar: | M-S | M-M | S-S | S-M | Mc-M | | |
|---|---|---|---|---|---|---|---|
| Ortalamalar: | 4,00 | 10,00 | 11,00 | 24,00 | 29,00 | | |
| | | | **Katsayılar** | | | $\Sigma a_j^2$ | $\psi = \sum a_j \overline{X}_j$ |
| $a_j$ | –1/3 | –1/3 | –1/3 | 1/2 | 1/2 | 0,833 | 18,167 |
| $b_j$ | 0 | –1 | 0 | 0 | 1 | 2 | 19 |
| $c_j$ | –1 | 0 | 1 | 0 | 0 | 2 | 7 |
| $d_j$ | 0 | 1 | –1 | 0 | 0 | 2 | –1 |

$$KT_{\text{karşılaştırma}} = \frac{n\left(\sum a_j \overline{X}_j\right)^2}{\sum a_j^2} = \frac{0(18,17)^2}{0,8333} = \frac{2641,19}{0,8333} = 3169,42$$

$$F = \frac{KO_{\text{karşılaştırma}}}{KO_{\text{hata}}} = \frac{3169,42}{32,00} = 99,04$$

$$KT_{\text{karşılaştırma}_2} = \frac{n\left(\sum b_j \overline{X}_j\right)^2}{\sum b_j^2} = \frac{8(19)^2}{2} = \frac{288}{2} = 1444,00$$

$$F = \frac{KO_{\text{karşılaştırma}_2}}{KO_{\text{hata}}} = \frac{1444,00}{32,00} = 45,125$$

$$KT_{\text{karşılaştırma}_3} = \frac{n\left(\sum c_j \overline{X}_j\right)^2}{\sum c_j^2} = \frac{8(7)^2}{2} = \frac{392}{2} = 196,00$$

$$F = \frac{KO_{\text{karşılaştırma}_3}}{KO_{\text{hata}}} = \frac{196,00}{32,00} = 6,125$$

$$KT_{\text{karşılaştırma}_4} = \frac{n\left(\sum d_j \overline{X}_j\right)^2}{\sum d_j^2} = \frac{8(-1)^2}{2} = \frac{8}{2} = 4,00$$

$$F = \frac{KO_{\text{karşılaştırma}_4}}{KO_{\text{hata}}} = \frac{4,00}{32,00} = 0,125$$

© Cengage Learning 2013

Eğer yokluk hipotezi doğruysa -büyük olasılıkla değil- dört doğrusal karşılaştırmanın ,20'ye yaklaşan bir *AHO* değeri vardır.[6] Bu hata olasılığının rahatsız edici miktarda yüksek olmasına rağmen bazı deneyciler özellikle önsel olduğu için kabul edebilmektedirler. Hata oranını düşürmenin bir yolu, her karşılaştırmayı $\alpha$ = ,01 gibi daha sıkı bir $\alpha$ düzeyiyle yapmaktır. Başka bir alternatif yöntem, Bonferroni tekniğini kullanmak olabilir. Bu yaklaşım neredeyse birincisiyle aynı miktarı üretmekle birlikte daha titiz bir tekniktir. Dik doğrusal farklılıklar adını alan özel bir doğrusal farklılık hakkında kısaca yapacağımız bir tartışmadan sonra bu tekniği düşünebiliriz. *AHO*'yu kontrol etmenin üçüncü bir yolu, daha az

6 Bu konuyu biraz daha açıklamalıyım. Morfin toleransı bildiğimiz gibi iyi tanımlı bir olgudur. Örneğin, ilk defa morfin alan farelerle dört defa morfin alan fareler aynı şekilde davranabilir mi? Zannetmem. Dört yokluk hipotezinin yanlışlıkla reddedilme durumu yoktur ve olasılığın ,20 gibi büyük bir sayı olması mantıklı değildir. İkiden fazla yokluk hipotezi doğru olsaydı şaşırırdım ve böyle bir durumda gerçek *AHO* hata oranı ,10'un üzerinde olmamalıdır.

sayıda doğrusal karşılaştırma yapmaktır. Örneğin, grup M-M ile grup S–S karşılaştırması büyük olasılıkla çok da önemli değildir. Dördüncü denemede, altıncı ya da yedinci denemede toleransın tamamen gelişmesi teorik olarak çok ilgi çekecek bir husus değildir. Karşılaştırmaları eleyerek *AH*O değerini, 15'e düşürebiliriz. Doğrusal farklılıkları çalıştırmak için seçtiğiniz yol fındık fıstık yeme şeklinizden veya hep dağa çıkma şeklinizden farklı olmalıdır. Sadece orda hep var olmaları yeterli olmamalıdır. Genel olarak eğer bir doğrusal farklılık önemli değilse karşılaştırmayı yapmayın.

## Dik Doğrusal Karşılaştırmalar

Tanımlandığı gibi, doğrusal karşılaştırmalar bize işlem farklılıklarına ilişkin bir dizi hipotezi test etme imkânı sağlar. Bazen bu farklılıklar birbirinden bağımsız olurken bazen de olmazlar. Örneğin; $\bar{X}_1$'nın $\bar{X}_2$ ve $\bar{X}_3$ ortalamalarından büyük olduğunu bilmek $\bar{X}_3$'ün $\bar{X}_4$'ten büyük olduğu hakkında bir şey söylemediği gibi $\bar{X}_4$'ün $\bar{X}_5$'ten büyük olabilme durumu için de herhangi birşey söylemez. Bu doğrusal karşılaştırmalar bağımsızdır. Bununla beraber $\bar{X}_1$'in $\bar{X}_2$ ve $\bar{X}_3$'ten büyük olduğunu bilmek, 50:50 şanstan daha iyi olacak şekilde $\bar{X}_1$'in $\bar{X}_2$'den büyük olacağını söyler. Bir doğrusal karşılaştırma kümesinin elemanlarının birbirinden bağımsız olması **dik karşılaştırmalar** olarak adlandırılır ve bütün dik karşılaştırmaların kareleri toplamı $KT_{\text{işlem}}$'i verir. (Eğer doğrusal karşılaştırmalar dik değilse, örtüşen bilgi içerirler ve bu toplanabilirlik özelliği sağlanamaz.) Hesaplama açısından bakıldığında, dik karşılaştırmaları, diğer karşılaştırmalardan ayırıp seçmemizi sağlayan dik karşılaştırma katsayıları arasındaki ilişkilerdir. Bunun dışındaki hesaplamalar aynıdır.

**dik karşılaştırmalar**

## Dik Katsayılar

Örneklem büyüklükleri eşit olmak üzere, doğrusal karşılaştırmaların dik olması için katsayıların, $a_j$ ve $b_j$ farklı karşılaştırmalar için katsayı kümeleri olmak üzere, aşağıdaki üç özelliği karşılaması gerekir.

1. $\sum a_j = 0$

2. $\sum a_j b_j = 0$

Ayrıca $KO_{\text{karşılaştırma}}$'yı $KO_{\text{işlem}}$ ile toplayabilmek için üçüncü bir kritere ihtiyacımız vardır.

3. Karşılaştırmalar sayısı = işlemler için *sd*.

Birinci kısıtlamayı tartışmıştık ve doğrusal karşılaştırmaların kareleri toplamı ile sonuçlanmıştı. İkinci sınırlama doğrusal karşılaştırmaların birbirinden bağımsız olması (dik olması) ve örtüşmeyen bileşenlerin toplanmasıdır. Üçüncü sınırlama, parçalardan bütünü oluşturmak için bütün parçalara ihtiyaç olduğudur.

İlk bakışta, $\sum a_j b_j = 0$ şartını sağlayacak katsayı kümelerini bulmak uğraştırıcı deneme yanılma süreçleri veya eş zamanlı denklem sistemlerinin çözümlerini gerektiriyor görünebilir. Aslında, bütün olası kümeleri bulamasa da bütün dik katsayı kümelerini bulabilecek basit bir kural vardır. Bu kural, ağaç diyagramı kullanarak $KT_{\text{işlem}}$'i parçalara ayırarak görselleştirmektir. Beş özellik için hesaplanacak genel *F*, eş zamanlı olarak beş özelliğin ortalamasıyla baş eder. Bu, ağacın gövdesini oluşturur. Eğer daha sonra 1 ve 2'nin kombinasyonunu 3, 4 ve 5'in kombinasyonu ile karşılaştıracaksak, biri İşlem 1 ve İşlem 2'yi, diğeri İşlem 3, İşlem 4 ve İşlem 5'i gösteren ağacımızın iki dalını oluştururuz. Daha önce de tartıştığımız gibi, işlem ortalamaları için olan $a_j$ değerleri uygulama sayısının tersi ve kümelerden birine ait katsayı negatif olacağından beş uygulama için katsayılarımız işlemler için sırasıyla (1/2, 1/2, –1/3, –1/3, –1/3) olur.

Bu durumda ağacımızın iki dalını oluşturmuş oluyoruz. Farklı dallardaki ortalamayı karşılaştıramayıp sadece aynı dalda yer alan ortalamaları karşılaştırabilmekteyiz. Bu durumda İşlem 3 ile İşlem 4 ve İşlem 5'in birleşiminin karşılaştırılması akla yakın bir karşılaştırma örneğidir. Bu durumda katsayılar (0, 0, 1, -1/2, -1/2) şeklinde olacaktır. İşlem 1 ve 2, 0 katsayısına sahiptir, çünkü istenen karşılaştırmanın bir parçası değildirler. İşlem 3'ün

katsayısı 1'dir, çünkü 1 işlemi içermektedir. İşlem 4 ve 5, kümede iki uygulama olduğu için her ikisi de -1/2 katsayısını almıştır. Eksi işareti keyfî olarak karşılaştırma taraflarından birine verilebilir.

Bir önceki yöntemi, yorgun düşene kadar, mümkün olan bütün karşılaştırma kümelerine uygulayabiliriz. Bu durum işlemlerin serberbestlik derecesi kadar karşılaştırma yaptığımız zaman mümkün olur. Bu yöntemin bir sonucu olarak, Şekil 12.1'de gösterilen karşılaştırmalara ve katsayılara dönelim. Bu katsayıların dik olduğunu göstermek için, sadece bütün *çiftlerin* katsayılar çarpımlarının toplamının sıfıra eşit olduğunu göstermemiz gerekir. Örneğin:

$$\sum a_j b_j = (\tfrac{1}{2})(1) + (\tfrac{1}{2})(-1) + (-\tfrac{1}{3})(0) + (-\tfrac{1}{3})(0) + (-\tfrac{1}{3})(0) = 0$$

ve

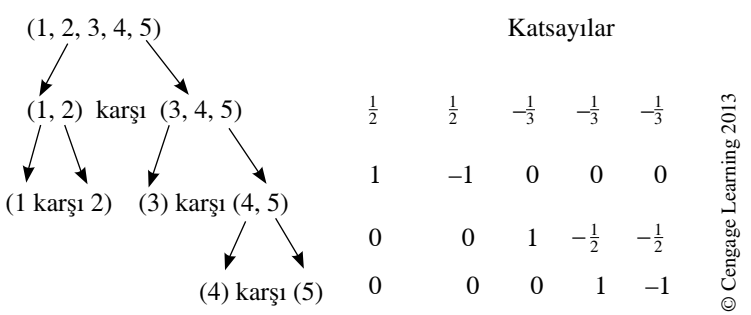
$$\sum a_j c_j = (\tfrac{1}{2})(0) + (\tfrac{1}{2})(0) + (-\tfrac{1}{3})(2) + (-\tfrac{1}{3})(-1) + (-\tfrac{1}{3})(-1) = 0$$

Böylece birinci ile ikinci ve birinci ile üçüncü dik karşılaştırmaların bağımsız olduğunu görürüz. Benzer hesaplamalar diğer doğrusal karşılaştırmaların da birinin diğerinden bağımsız olduğunu gösterecektir.

Bu katsayılar mümkün olan dik karşılaştırma kümelerinin sadece biridir. Eğer İşlem 1 ile işlem 2, 3, 4 ve 5'in birleşimini karşılaştırmak için başlasaydık sonuçlanacak olan dik karşılaştırma kümesi oldukça farklı olacaktı. Deneyi yapan kişinin hangi dik karşılaştırmaları planladığı önemlidir. Ayrıca akılda tutulmalıdır ki mâkul karşılaştırmaları yapacak şekilde katsayıları ayarlamanız, o karşılaştırmaları gerçekte yapabileceğiniz anlamına gelmez.

$F$ ve $t$'nin dik karşılaştırmalardan elde edilişine yönelik hesaplama ile dik olmayan karşılaştırmaların karşılaştırılması için yapılanın aynıdır. Bundan dolayı, bu aradaki örnek ile çalışırken küçük bir kazancımız vardır. Sizin için tam bir karşılaştırmalar kümesi kurmak ve buna ilişkin aritmetiği yapmak iyi bir pratik olabilir. Cevaplarınızı kareler toplamı toplamının $KT_{\text{işlem}}$'e eşit olup olmadığını göstererek kontrol edebilirsiniz.

İlk istatistik öğretmeye ve yazmaya başladığımda, dik karşılaştırmalar, önsel ve sonradan planlanan testler arasındaki ayrım oldukça büyük bir problemdi. Yazarlar dikliğin önemini kendilerince anlatmakta ve dik olmayan durumlarda karşılaştırma yapınca kendinizi suçlu hissetmenize yol açmaktaydılar. Bu tutum geçen yıllar boyunca değişti. Toplamı $KT_{\text{işlem}}$'e eşit olmasından dolayı dik karşılaştırmalar kümesine sahip olmak güzel olsa da, insanlar dik olmayan karşılaştırmalarla çalışmaya daha istekli oldular. Kesinlikle diğerlerine dik olmadığından dolayı önemli karşılaştırmalardan vazgeçmenizi tavsiye etmem. Gerçekte daha önce yaptığım karşılaştırmalar birbirine dik değildi ve bu konuda da çok endişelenmedim. Önemli sorulara cevap aramaktaydılar (daha önce söylediğim gibi S-S ve M-M değil). Buna benzer olarak ilgimiz olmayan bir karşılaştırmayı sadece dik karşılaştırmalar kümesinin elemanı diye yapmamak gerekir. Hatırımızda tutmamız gerekir ki dik olmayan karşılaştırmalara ait karşılaştırmalar birbirinden bağımsız değildir.




**Şekil 12.1** $KT_{\text{uygulama}}$'nın ağaç diyagramı kullanarak görselleştirilmesi.

## Bonferroni *t* (Dunn testi)

**Dunn testi**

**Bonferroni *t***

**Bonferonni eşitsizliği**

Daha önce doğrusal karşılaştırmaları kullanırken aileboyu hata oranının değerini azaltmanın bir yolunun her karşılaştırma için daha sıkı $\alpha$ değeri kullanmak olduğunu tavsiye etmiştim. Bu tavsiye $\alpha$ = ,05 yerine istatistik tablolarımızda da olan $\alpha$ = ,01 kullanmanız yönündedir. (İstatsitiksel programlar hesaplayıp yazabilse de tablolar genellikle ,05 ile ,01 arasındaki kritik *t* değerini vermezler.) *AHO*'yu kontrol etmenin daha formel bir yolu Dunn (1961) tarafından önerilen ve her karşılaştırmanın hata oranını kontrol eden Dunn testinde bulunur. Bu yöntem genelikle yapılabilecek bütün karşılaştırmalardan birkaçını tercih ettiğiniz durumda uygundur. Bu test uzun zamandır bilinmesine rağmen, Dunn bunu formalize eden ve gerekli tabloları sunan ilk kişidir. Bu nedenle bazen **Dunn testi** olarak anılır. Şimdilerde daha çok **Bonferroni *t*** adını almaktadır. Bonferroni *t*, **Bonferroni eşitşizliği** olarak bilinen yönteme dayanır. Bu yöntem bir veya *birkaç* olayın olabilme olasılığının hiçbir zaman kendi bireysel olasılıkları toplamını geçemeyeceğini söyler.

Bu şu anlama gelir; $\alpha'$ = ,05 düzeyinde I. Tip hata ile üç karşılaştırma yaparsak, *en küçük* I. Tip hata oranı 3 x ,05 = ,15 değerini hiçbir zaman geçmez. Daha formel terimlerle konuşmak gerekirse, *c*'nin karşılaştırma sayısını ve $\alpha'$nün her karşılaştırma için I. Tip hata olasılığını temsil ettiğini düşünürsek, *AHO* $c\alpha'$dan küçük veya ona eşit olacaktır. Buradan $\alpha$ istenilen en yüksek *AHO* düzeyini göstermek üzere, $\alpha' = \alpha/c$ olacak şekilde her karşılaştırma için ayarlanırsa, $AHO \leq c\alpha' = c(\alpha/c) = \alpha$ olur. Dunn (1961) bu testi her karşılaştırma $\alpha' = \alpha/c$ için, $AHO \leq a$ korunacak şekilde desenlemiştir. Bu, standat *t* testi yöntemi kullanılarak ancak sonuçların uyarlanmış *t* tablosuna göre değerlendirilmesiyle elde edilebilir.

Hemen karşılaşacağınız problem, her karşılaştırmayı yapmak istediğiniz $\alpha' = \alpha$/c için Student *t* standart tablolarında karşılık bulamayabilirsiniz. Bu üç karşılaştırmayı $\alpha' = \alpha/c$ = ,05/3 = ,0167 ile yapmak istiyorsanız $\alpha$ = ,0167 için *t* kritik değerlerine ihtiyacınız vardır. Bu değerleri yazılım kullanarak da kolaylıkla hesaplayabilirsiniz.[7] Dunn'ın en büyük katkısı bu tür tabloları kazandırmak olmuştur. Bununla beraber artık bu tablolara ihtiyacımız yoktur, çünkü yazılımlar güncel *p* değeri sağlamaktadırlar. Eğer bunu el ile yapmak istersek hazır olasılık hesaplayıcılarını kullanabiliriz. Onları bu baskıda açıklamayacağım.

İkili karşılaştırmalar için (bir ortalama ile diğer ortalamanın karşılaştırılması) Bonferroni testi aşağıdaki gibi tanımlanır:

$$t' = \frac{\overline{X}i - \overline{X}j}{\sqrt{\frac{KO_{hata}}{n} + \frac{KO_{hata}}{n}}} = \frac{\overline{X}i - \overline{X}j}{\sqrt{\frac{2KO_{hata}}{n}}} \quad veya \quad F' = \frac{\psi^2}{\frac{\sum a_j^2 KO_{hata}}{n}}$$

Hesaplanan $t'$ değeri, $\alpha/c$ manidarlık düzeyindeki ($t_{a/c,s}$) olan $t'$ kritik değeri ile karşılaştırılıp değerlendirme yapılır. Hâlâ standart *t* testini kullandığımıza dikkat ediniz. Standart *t* değeri ile $t'$ arasındaki fark sadece $\alpha$/c değerlendirmesi için kullanılan tablolardır. Eşit olmayan örneklemler, fakat homojen olan varyanslarda ise en soldaki $n_i$'ler ve $n_j$'ler yerine *n* konur. Heterojen varyans durumu için, bu bölümün ileri kısımlarındaki Games ve Howell'ın çözümünü görünüz.

Eşleştirilmiş çiftler halinde olsun veya olmasın ortalamalar arasındaki farkı test etmek için genel bir ifade yazmak gerekirse bunu $t'$ kullanarak doğrusal karşılaştırmalar aracılığıyla aşağıdaki gibi yapabiliriz.

$$\psi = \sum a_j \overline{X}j \quad \text{ve} \quad t' = \frac{\psi}{\sqrt{\frac{\sum a_j^2 KO_{\text{hata}}}{\text{n}}}}$$

7 Ücretsiz olasılık hesaplayıcıları, http://www.danielsoper.com/statcale/.adresinde bulunabilir. Bu kitabın Web sitesinde kritik değerleri hesaplayabilecek olan (tProb, R, FProb.R ve chisqProb.R) hazır R kodları yer almamaktadır.

Yukarıdaki ifade Bonferroni *t* testinin en genel formunu belirtir. $\psi$ herhangi bir doğrusal kombinasyon olmak üzere ($\sum a_j = 0$ olmasını gerektiren doğrusal karşılaştırma olması gerekmez) *c* sayıdaki karşılaştırmanın *AHO*'su yaklaşık $\alpha$ kadardır (Dunn, 1961). En basit açıklamayla Bonferroni *t*, her zamanki *t* testini yapar ancak sonuçları, *AHO*'yu sınırlandırmak için modifiye edilmiş kritik değerler kullanarak değerlendirir.

Bonferroni testi ve varyasyonlarına ilişkin bir uyarım olacak: Bu testler eşleştirilmiş çiftler halinde olsun veya kompleks olsun planlı doğrusal karşılaştırmaları sınırlandırdığınız zaman uygundur. Ancak, SPSS ve SAS Bonferroni testine sadece eşleştirilmiş çiftlerle ilgili çoklu karşılaştırmalar için imkân sağlar ve bu da çoğu zaman uygun değildir. Bu sistem altında, hepsini önemseseniz de önemsemeseniz de, *c*'yi otomatik olarak, bütün *muhtemel* ikili karşılaştırmalara göre ayarlamaktadır. Eğer planlanan doğrusal karşılaştırmalar kümesi için bir düzeltme uygulayacaksanız bu farklılıkları tanımlamanız ve $\alpha/c$ düzeyine göre kendiniz değerlendirme yapmanız ve bunu yapabilmek için Compare Means/One-way ANOVA yolunu kullanmanız gerekir, univariate yolunu değil. SAS'da ise ProcGLM yolu ile ProcGLM ifadesini kullanmalısınız.

**Dunn-Sidak testi**

Bonferroni yönteminin bir varyasyonu Sidak (1967) tarafından önerilmiştir. Bu öneri çarpımsal $p(AHO \leq 1 - (1 - \alpha)^c$ eşitsizliğine dayanır ve $\alpha' = 1 - (1 - \alpha)^{1/c}$ manidarlık düzeyinde *t*'ye göre değerlendirilir. (Bu, **Dunn-Sidak testi** olarak adlandırılır.) İki yöntemin gücüne ilişkin karşılaştırmalar Sidak testinin lehine olmak üzere çok küçük farklılıklar gösterir. Daha yaygın kullanılmasından dolayı Bonferroni testini kullanmaya devam edeceğiz. Birçok program Sidak testine imkân vermektedir.

Daha önce bu bölümde yer alan *AH*O değeri maksimum ,20 olacak şekilde yapılmış olan dört doğrusal karşılaştırmalarını düşünelim. (Bu doğrusal karşılaştırmaların hepsi *F* istatistiği içermekteydi, çünkü her birinin serbestlik derecesi 1'e eşitti ve $t = \sqrt{F}$ eşitliği kullanılarak *t*'den *F* elde edilmişti.) Aynı eşitlikleri *AH*O değeri 4 x ,05 = ,20 olacak şekilde değil de *AHO*'nun maksimum değeri ,05 olacak şekilde yeniden kullanırsak, Bonferroni *t* testini kullanabiliriz. Her seferinde $t'$ çözümlemesi ve $t'$ değerlendirmesi, standart *t* kullanılarak yapılıp $\alpha' = \alpha/4$ alınarak yapılabilir. Önce eşleştirilmiş çiftler testini alalım, hesaplamalar onu izler. Dört doğrusal karşılaştırma için kritik $t'$ değeri, *t* dağılımının 5/4 = ,0125 = %1,25 değerinin çift yönlü hipotez testi değeri olan ,00625'in karşılaştırma sayısına göre düzenlenmiş hali kullanılarak değerlendirilebilir. Eğer *p* değerini veren standart bir yazılım kullanıyorsanız, yapmanız gereken basitçe reddetme düzeyinizi ,0125 olarak değiştirmenizdir. Diğer bir seçenek 6. dipnotta belirtildiği gibi *sd* = 35 için $t_{,00625,35} = 2{,}63$ olacak şekilde kritik değer bulunabilir.[8]

Bu teste ilişkin hesaplamalara aşağıda yer verilmiştir. Doğrusal karşılaştırmaları $t \geq \pm 2{,}63$'e göre bulmak istiyoruz.

**Mc-M'ye karşı M-M:**

$$t' = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{2MS_{\text{hata}}}{\text{n}}}} = \frac{29{,}00 - 10{,}00}{\sqrt{\frac{(2)(32{,}00)}{8}}} = \frac{19}{\sqrt{8}} = 6{,}72$$

**S-S'ye karşı M-S:**

$$t' = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{2MS_{\text{hata}}}{\text{n}}}} = \frac{11{,}00 - 4{,}00}{\sqrt{\frac{(2)(32{,}00)}{8}}} = \frac{7}{\sqrt{8}} = 2{,}47$$

**M-M'ye karşı S-S:**

$$t' = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{2MS_{\text{hata}}}{\text{n}}}} = \frac{10{,}00 - 1{,}00}{\sqrt{\frac{(2)(32{,}00)}{8}}} = \frac{-1}{\sqrt{8}} = -0{,}35$$

8 Daniel Soper'ın mükemmel olasılık hesaplayıcısını kullanmak için, Student *t* dağılımı linkini kullanarak Student *t* hesaplayıcısını kullanınız.

Daha kompleks karşılaştırmalar için önce olduğu gibi $a_j$ = 1/3, 1/3, 1/3, -1/2, -1/2 olacak şekilde ayarlama yapılırsa aşağıdaki elde edilir:

**S-M ve $M_c$-M'ye karşı M-M, S-S ve S-S:**

$$t' = \frac{\sum \alpha_j \overline{X}_j}{\sqrt{\frac{\alpha_j^2 KO_{\text{hata}}}{\text{n}}}} = \frac{(1/2)(24) + \ldots + (-1/3)(4)}{\sqrt{\frac{(0,833)(32,00)}{8}}} = \frac{18,167}{\sqrt{3,3333}} = 9,95$$

Bu durumda, ilk ve son karşılaştırmalar 2,63 değerini geçerek manidar olacak, fakat diğer ikisi olmayacaktır.[9] Aynı ortalamaya sahip evrenlerden örneklenmiş S-S ve M-S gruplarına ilişkin hipotezler daha önce reddedildiği halde daha tutucu olan Bonferroni *t* testine göre reddedilmemektedir. Burada daha önce morfin enjekte edilmesinin aşırı duyarlılığa neden olduğu kararına varamayacağız. İki yöntemin sonuçları arasındaki farklılık doğrudan doğruya daha tutucu olan aile-boyu hata oranından kaynaklanmaktadır. Eğer *AHO*'yu görmezden gelerek her karşılaştırmanın hata oranına odaklanmak istersek, her *t* (veya *F*) değerini $\alpha$ = ,05 düzeyindeki kritik değerle karşılaştırırız. Diğer taraftan, öncelikle *AH*O değerini kontrol altına almak istersek, her *t* veya *F* değerini daha sıkı bir şekilde değerlendirebiliriz. Bu araştırmacıya bağlıdır.

## Çok Adımlı Bonferroni Yöntemleri

Bonferroni çoklu karşılaştırma yöntemlerinin bir dizi türü vardır. Bunlar burada varyans analizi çerçevesinde belirtilse de, aileboyu hata oranını kontrol etmek istediğimiz çoklu hipotez testleri için de her zaman kullanabiliriz. Bu yöntemler Tukey HSD'de olduğu gibi (kısaca tartışılacaktır) bir dizi yokluk hipotezinin *AH*O değerini belirli bir $\alpha$ düzeyinde limitleyebilmemiz açısından avantajlıdır. Eğer ilgi alanımız *doğrusal karşılaştırmaların bir alt kümesi ise*, bu yöntemler Tukey HSD'ye göre daha az sıkıdır. Genel olarak Bonferroni prosedürleri bir ortalama kümesine ilişkin bütün ikili karşılaştırmaları yapmak için kullanılmaz. Bu amaç için yokluk hipotezi reddedildikçe kritik değeri değişen, çok basamaklı prosedürler kullanılır.

Gördüğünüz gibi Bonferroni yöntemleri, bir karşılaştırma ailesindeki doğrusal farklılıklar için olan *AHO*'yu, tekil doğrusal farklılıklar arasında bölme prensibine dayanır. Şöyle ki; eğer *AH*O değerinin ,05 olmasını istiyorsak ve dört doğrusal farklılığımız varsa bu durumda $\alpha$ = ,05/4 = ,0125 olur. Çok adımlı testler de benzer bir prensip izler, fakat en önemli farklılık $\alpha$ değerinin parçalanmasında oluşur. Bunlar temel olarak en büyük farklılığı (veya korelasyon katsayısını) test ederler. Eğer en büyük fark manidar ise test sayısını bir küçülterek ondan sonraki en büyük farka geçerler. Bu ilk manidar olmayan farka kadar devam eder. Bu yaklaşımın altında yatan mantık şu gerçeğe dayanır: Aileboyu hata oranı en az bir I. Tip hata yapma olasılığıdır. Dört tane test edilecek ortalamamızın olduğunu farz edin. Eğer birincisinde hata yaparsanız, zaten birincisini yapmış olursunuz daha sonrakiler zarar vermez. (Elbette bütün hatalar zarar verir ama birden fazlası aileboyu hata oranını etkilemez.) Eğer bu testte hata yapmazsanız bu durumda reddedecek üç yokluk hipoteziniz kalır ve $\alpha' = \alpha/3$ olur.

Bu yöntemlerin ayrıntısına girmek yerine, yerden tasarruf etmek amacıyla bu tartışmayı kitabın Web sayfasına taşıdım. Bunlar testlerinizi aşamalar halinde desenleme durumunda oldukça güçlü prosedürlerdir ve o sayfaları okumanızı tavsiye ederim.

---

9 Gerçek olasılıklar şu şekildedir: ,000; ,018; ,726 ve ,000

## Kırpılmış Ortalamalar

Veri uzun kuyruklu dağılımlardan geldiğinde standart yöntemlere göre daha güçlü olan daha genel bir yaklaşımı da dahil etmek istiyorum. Bu, kırpılmış ortalamalarla ilgili bir konudur. Bu konunun güzel tarafı, bu bölümdeki herhangi bir yönteme adapte edilebilmesidir. Bu işlem uygun kırpılmış ortalama ve standart hataların basitçe yerine konulmasıyla gerçekleştirilebilir.

Uçlarından kırpma yapılacak kadar geniş bir örnekleme sahip olduğunuzu varsayıyorum, çünkü başından ve sonundan kırpacağımızı varsayıyorum. Wilcox %20'lik bir kırpmayı önermiştir. Bu, hatırı sayılır bir örneklem büyüklüğü kaybına yol açsa da güç kazandırmaktadır. Her grupta 40 gözleminizin olduğunu düşünün ve Wilcox'ın önerdiği %20'lik kırpma işlemiyle devam edelim. Bu en düşük ve en yüksek (,20)(40) = 8 gözlemi eleyeceğimiz ve her koşul için 24 gözlemimiz var anlamına gelir. Kırpılmış ortalamalar her grupta yer alan yirmidörder gözlemin ortalaması olacaktır. Varyansı hesaplamak için Winsorize edilmiş (Winsorized) örneklemleri kullanacağız. Şöyle ki en düşük ilk 8 puan ile 9'uncu en düşük puan ve en yüksek 8 puan ile 9'uncu en yüksek puan yerlerine konulacaktır. Bu bize $n_{\text{i}} = 40$ puanlı ancak $i$'yinci örneklemde $h_i = 24$ bağımsız puanlı bir örneklem bırakır. Winsorize edilmiş örneklemden elde edilen Winsorize edilmiş varyansı $S^2_{W_i}$ sembolüyle gösterirsek ortalamanın standart hatasının karesi

$$s^2_{W\bar{X}_i} = \frac{(n_j - 1)s^2_{W_i}}{h_j(h_j - 1)}$$

ifadesiyle yazılır ve çiftlerin farklarına ait dayanıklı $t$ testi aşağıdaki gibi yazılabilir:

$$t_W = \frac{\bar{Y}_{ti} - \bar{Y}_{tj}}{\sqrt{s^2_{W\bar{X}_i} + s^2_{W\bar{X}_j}}}$$

Burada şaşırtıcı hiçbir şey yapmıyoruz. Ortalamalar yerine kırpılmış ortalamaları, varyansların yerine Winsorize edilmiş varyansları koyduk, ancak örneklem büyüklüğü $n_{\text{i}}$'yi düzeltmek için $h_j$'yi (kırpılmış örneklem büyüklüğü) kullandık. Bundan başka bu bölümdeki her yöntemde yerine koyduğumuz veya koyacağımız standart bir $t$ testimiz var. Bununla beraber karşımıza çıkan başka bir güçlük serbestlik derecesi kestirimidir. Sebestlik derecesini aşağıdaki eşitlikle kestiriyoruz:

$$sd_W = \frac{(s^2_{W\bar{X}_i} + s^2_{W\bar{X}_j})^2}{s^4_{W\bar{X}_i}(h_i - 1) + s^4_{W\bar{X}_j}(h_j - 1)}$$

Bu biraz karışık olmakla beraber çalışılması çok zor olmayan bir formüldür. Keselman ve arkadaşlarının (2005) belirttiği gibi, "Araştırmacılar normal olmayan veri ile uğraştıklarını hissederlerse, her zaman kullanılan merkezî eğilimin ve değişkenliğin en küçük kareler kestiricilerini daha dayanıklı kestiricilerle değiştirir ve bu kestiricileri daha önce tavsiye edilenlerin çoklu karşılaştırma yöntemlerinin herhangi birine uygularlar."

**NOT**

Bir kez daha vurgulamak isterim ki Bonferroni testi ve türevleri tamamen geneldir. Bunlar varyans analizinin veya diğer başka bir istatistiksel yöntemin elemanı değildir. Herhangi bir istatistiksel yöntem kullanılarak uygulanmış çeşitli testleriniz varsa (büyük olasılıkla başka testler), Bonferroni yaklaşımını AHO'yu kontrol etmek amacıyla kullanabilirsiniz. Örneğin yeni aldığım bir e-postada birisi Bonferroni testini lojistik regresyona nasıl uygulayacağını sormaktadır. Bunu varyans analizinde yaptığının aynısıyla yapabilir. Lojistik regresyonda yer alan testlerin sayısını alır, $\alpha$ bu sayıya bölünür, $\alpha/c$'den daha düşük olasılıktaki sonuçlar manidar olarak ilân edilir. Bunu yapabilmek için lojistik regresyon hakkında bile bir şey bilmenize gerek yoktur.

## 12.4 Karşılaştırmalar İçin Güven Aralıkları ve Etki Büyüklükleri

Bir deneyden elde ettiğimiz veri üzerinde yaptığımız test ve tekil karşılaştırmalara baktığımız işlem "tekil farklılıklar" olarak aldandırılır ve genellikle grup ortalamaları arasındaki farkın büyüklüğünü bilmek isteyeceğiz. Bölüm 11'de gördük ki bütün ortalamaları karşılaştıran geniş kapsamlı $F$, genellikle $\eta^2$ veya $\omega^2$ gibi $r$-ailesi ölçüleriyle kullanılır. Tekil ortalama veya ortalama kümelerinin karşılaştırılmasına baktığımızda, farklarımıza ilişkin güven aralığı hesaplamak veya etki büyüklüğü ölçüsü olarak $d$-ailesini kullanmak daha mantıklı olacaktır.

$d$-ailesi ölçülerine yaklaşmanın çeşitli yolları vardır. En basit yol için, $t$ testlerini tartıştığımız ve uyguladığımız Bölüm 7'ye gitmemiz gerekir ve orada tartıştığımız ölçüleri kullanırız. Doğrusal karşılaştırmalardaki yaklaşımı izlersek aynı yere geleceğiz. Hatırlayınız, iki grup ortalamasına bakarken, iki grup arasında $t$ testi yapmamız ve doğrusal karşılaştırma hesaplamasından $F$ elde edip sonra onun karekökünü almamız arasında bir fark yoktu. Doğrusal karşılaştırmaları kullanmanın avantajı, 'daha genel olması ve iki gruptan ziyade grup kümelerinin ortalamaları arasında karşılaştırma yapmamıza imkân vermesidir' diyebiliriz.

Siegel'ın morfin çalışmasından bir örnek alabiliriz. Beni gerçekten ilgilendiren doğrusal karşılaştırmalardan biri Grup M-M ile Grup $M_c$-M arasındaki doğrusal karşılaştırmadır. Eğer bunların ortalamaları arasındaki fark gerçekten manidar ise bu bize fizyolojik kapsamda morfin verildiği zaman önemli bir şeylerin değişiyor olduğunu söyler. Grup istatistikleri aşağıda verilmiştir.

| İstatistikler | M-M | Mc-M |
|---|---|---|
| Ortalama | 10,00 | 29,00 |
| Std. Sapma | 5,13 | 6,06 |
| Aryans | 26,32 | 37,95 |
| $KO_{\text{hata}}$ | 32,00 | |

© Cengage Learning 2013

Bu doğrusal karşılaştırmaların katsayıları M-M için "–1" , Mc-M için "+1" ve diğer durumlar için "0" olacaktır.

### Güven Aralığı

İlk önce durumlar arasındaki farklar için güven aralıklarını hesaplayalım. İki ortalama arasındaki farkın güven aralığına ilişkin genel formül

$$GA_{,95} = (\overline{X}i - \overline{X}j) \pm t_{,025} S_{\overline{X}i - \overline{X}j}$$

veya $\psi_j$ doğrusal karşılaştırma değerlerini göstermek ve $\psi_j = \sum a_j \overline{X}j$ olmak üzere,

$$GA_{,95} = (\psi_j) \pm t_{,025} S_{\overline{X}i - \overline{X}j}$$

yazılabilir. Buradan doğrusal karşılaştırmaların standart hatası ($s_{\text{hata}}$)

$$\sqrt{\frac{2KO_{\text{hata}}}{\text{n}}}$$

olur. Örneğimizdeki iki duruma ait farkın güven aralığı aşağıdaki gibi yazılabilir:

$$GA_{,95} = \big((-1)(10) + 1(29)\big) \pm 2,03\sqrt{8,00}$$
$$= 19 \pm 2,03(2,828) = 19 \pm 5,74$$
$$13,26 \leq \mu_{\text{M-M}} - \mu_{\text{Mc-M}} \leq 24,74$$

Aralığın evren ortalamaları arasındaki gerçek farklılığı içine alma olasılığı ,95'tir.

Etki büyüklüğünü hesaplama zamanı geldiğinde, kullanacağımız hata terimini seçmemiz gerekecek - denklemdeki standart sapma. Bütün analizden gelen $KO_{\text{hata}}$ teriminin karekökünü tercih edeceğim, çünkü beş grubun ortalama varyansının karekökünü temsil etmektedir. Kline (2004) bu yaklaşımı tavsiye etmektedir. Bunun dışında iki tane mükemmel alternatifim daha var. Birincisi, sorudaki iki grubun varyanslarının ortalamasını alırım (örneklem büyüklükleri eşit değilse belki ağırlıklandırılabilir). Bu durumda (26,32 + 37,95)/2 = 32,135 ve $\sqrt{32,135}$ = 5,669 olur. Bu, beş grubun tamamına ait varyans heterojenliği hakkında kuşkulanırsam anlam kazanabilir. Alternatif olarak, bir grubu kontrol grubu olarak düşünürüm ve onun standart sapmasını hata terimi olarak kullanırım. Burada M-M'nin kontrol grubu olacağını savunabilirim, çünkü deneme 4'te koşullar değişmemektedir. Bu durumda $s_{\text{hata}}$ = 5,13 olacaktır. Sanırım tercihim, sorudaki varyans ortalamalarından elde edilen kestirim yönünde olacak. Eğer beş gruptaki varyanslar homojense, sorudaki grupların ortalamaları bütün beş grubun ortalamasından fazla sapmaz, böylelikle büyük bir kaybım olmaz.

## Etki Büyüklüğü

Yukarıda gördüğümüz gibi, Mc-M ve M-M grupları arasındaki farkın güven aralığı $13,26 \leq \mu_{Mc-M} - \mu_{M-M} \leq 24,74$'tür. Bütün limitler, 0'ın aynı tarafındadır, bu farkın manidar olduğunu gösterir. Bununla beraber buradaki bağımlı değişkenimiz hayvanın pençelerini yalamadan önce geçen zamanın uzunluğudur. Hiçbirimizin bu zamanı, bu durum için, kısa ya da uzun anlayabilecek güçlü bir sezgisinin olduğunu sanmıyorum. On üç saniyelik fark oldukça uzun görünmekle beraber ne olduğunu anlayabilmenin daha iyi bir yolu olmasını isterim. Bunu hesaplamanın bir yolu ortalamalar arasındaki farkın etki büyüklüğünü hesaplamaktır.

Etki büyüklüğü ölçümüz, temelde bağımsız gruplar için daha önce yaptığımız *t* testindeki ile aynıdır. Bununla beraber bunu hafifçe farklı yazacağım çünkü daha karmaşık karşılaştırmalara genelleyebileceğiz. Gördüğümüz gibi, $\psi$ iki ortalama veya ortalama kümeleri arasındaki doğrusal farklılığı temsil etmekteydi. Yani gerçekte sadece ortalamalar arasındaki farktır. Bu farkı alıp standartlaştırırsak, basitçe gruplar arasındaki farkı standart sapma birimi cinsinden ifade etmiş oluruz. (Bölüm 7'de yaptığımız buydu.)

Bölüm 7'de $s_{\text{birleşik}}$ birleşik varyansın karekökü ve gruplar içi ortalama standart sapmanın ölçüsü olmak üzere

$$\hat{d} = \frac{\overline{X}_i - \overline{X}_j}{s_{\text{birleşik}}}$$

tanımını yapacağız. Burada da temel olarak aynı şeyi hesaplayacağız, fakat ifadeyi aşağıdaki gibi yazacağım:

$$\hat{d} = \frac{\psi}{s_{hata}} = \frac{\sum a_i \hat{X}_i}{s_{hata}}$$

Pay basit bir doğrusal farklılık ve payda gruplar içi standart sapmanın bir kestirimidir.

Yukarıdaki formülde, iki nokta öne çıkar. Birincisi, katsayılar bizim standart düzen dediğimiz şekilde olmalıdır. Bu basitçe katsayıların mutlak değerleri toplamının 2 olması anlamına gelir. Örneğin, iki grubun ortalamasını üçüncü başka bir grupla karşılaştırmak istersek bu durumda katsayılar (1/2, 1/2, –1) olmalıdır. Kabul ettiğimiz manidarlık düze-

**Tablo 12.2** Morfin örneğimizdeki durumların ortalamaları

| Gruplar: | M-S | M-M | S-S | S-M | Mc-M | | |
|---|---|---|---|---|---|---|---|
| Ortalamalar: | 4,00 | 10,00 | 11,00 | 24,00 | 29,00 | | |
| | | | Katsayılar | | | $\Sigma a_j^2$ | $\psi = \sum a_j \overline{X}_j$ |
| $a_j$ | –1/2 | 1/3 | –1/2 | 1/3 | 1/3 | 0,833 | 18,167 |
| $b_j$ | 0 | –1 | 0 | 0 | 1 | 2 | 19 |
| $c_j$ | –1 | 0 | 1 | 0 | 0 | 2 | 7 |
| $d_j$ | 0 | 1 | –1 | 0 | 0 | 2 | –1 |

**M-M, S-M, Mc-M gruplarına karşı M-S, S-S**

$$\hat{d}_1 = \frac{\sum a_i \overline{X}_i}{s_{hata}} = \frac{\left(-\frac{1}{3}\right)\overline{X}_{M-S} + \left(-\frac{1}{3}\right)\overline{X}_{M-M} + \left(-\frac{1}{3}\right)\overline{X}_{M-S} + \left(\frac{1}{2}\right)\overline{X}_{S-M} + \left(\frac{1}{2}\right)\overline{X}_{Mc-M}}{\sqrt{MS_{hata}}}$$

$$= \frac{\left(-\frac{1}{3}\right)4{,}00 + \left(-\frac{1}{3}\right)10{,}00 + \left(-\frac{1}{3}\right)11{,}00 + \left(\frac{1}{2}\right)24{,}00 + \left(\frac{1}{2}\right)29}{\sqrt{32}}$$

$$= \frac{\frac{4{,}00 + 10{,}00 + 11{,}00}{3} + \frac{24{,}00 + 29{,}00}{2}}{\sqrt{32}} = \frac{-8{,}333 + 26{,}5}{5{,}657} = \frac{18{,}167}{5{,}657} = 3{,}21$$

**M-M'ye karşı Mc-M**

$$\hat{d}_1 = \frac{\sum b_i \overline{X}_i}{s_{hata}} = \frac{(-1)\overline{X}_{M-M} + (1)\overline{X}_{Mc-M}}{\sqrt{MS_{hata}}}$$

$$= \frac{(-1)10{,}00) + (1)29{,}00}{\sqrt{32}}$$

$$= \frac{-10{,}00 + 29{,}00}{\sqrt{32}} = \frac{19}{5{,}657} = 3{,}36$$

© Cengage Learning 2013

yinde (1, 1, 2) katsayılarını kullanarak geldiğimiz nokta ile (3, 3, -6 ) kullanarak geldiğimiz *F* değeri aynıdır. Ancak birincisi bize ilk iki grup ile üçüncü grup ortalamasına ait doğrusal karşılaştırma için sayısal bir değer verir. Bunu aşağıdakini yazdığınız zaman kolaylıka görebilirsiniz.

$$\psi = (\tfrac{1}{2})\overline{X}_1 + (\tfrac{1}{2})\overline{X}_2 + (-1)\overline{X}_3$$
$$= \frac{\overline{X}_1 + \overline{X}_2}{2} - \overline{X}_3$$

Ortalamaları arasındaki farkları aldığımız grupları açıkça görebilirsiniz.

$\widehat{d}$ denklemi için diğer bir konu paydanın seçilmesindedir. Birkaç paragraf öncsinde belirttiğim gibi, en az üç olası kestirim vardır. $KO_{\text{hata}}$'nın karekökünü kullanabiliriz veya karşılaştırılan grupların ortalama varyansının karekökünü kullanabiliriz veya bir grubu kontrol grubu olarak seçip onun standart sapmasını kestirimimiz olarak kullanabiliriz. En

yaygın kullanılan yaklaşım $KO_{hata}$'nın kareköküdür. Ben de burada aynısını yapacağım, çünkü örneğimizdeki varyanslar birbirine oldukça benzerdir.

Daha önceden teorik nedenlerden dolayı ilgi duyduğumuz dört doğrusal karşılaştırma ile ilgilendik. Geleneksel Bonferroni yöntemi bu karşılaştırmalardan ikisinin manidar olduğunu, diğer ikisinin ise manidar olmadığını gösterdi. Manidar olan karşılaştırmalar için etki büyüklüğü hesaplaması Tablo 12.3'te gösterildi. Bu hesaplamalarda tutarlılık olması açısından payda olarak $KO_{hata}$ kullandım. Testlerimizin son iki karşılaştırmanın manidar olmadığını söylemesinden dolayı, bunlar için en iyi yaklaşımımız etki büyüklüklerini 0,00 olacak şekilde ele almak olabilir. Test istatistiklerinin neredeyse manidar olacak kadar büyük olması durumunda ne yapacağımız sorusu ilgi çekici olabilir. Bu durumda etki büyüklüğü ölçümünü, ilgili hipotezin manidar olmadığına dikkat çekerek belirtirim.

Diğer etki büyüklüklerinin de tatmin edici olduğunu, hepsinin en az üç standart sapma farklılığı olduğunu görebilirsiniz. Bu etkiler hakkında gelecek kesimde konuşacağım.

## 12.5 Sonuçların Rapor Edilmesi

Bu veri üzerinde çeşitli testler yaptık ve aşağıdakiler Bonferroni prosedürüne ilişkin rapordur.

> Bu deneyde sıcak yüzeye konulan tavşanlarda morfin toleransı olgusu test edilmiştir. Altta yatan hipotez; tekrarlı morfin enjeksiyonlarının hayvanlarda acıya aşırı duyarlılık geliştirdiği ve ilâcın etkisini azalttığı şeklindedir. Hayvanlar ilâçsız test edildiğinde veya ilâcı farklı bir şekilde aldıklarında bu aşırı duyarlılık kendini kısa pençe-yalama tepki zamanı ile göstermektedir.
>
> Analizin tümüne ilişkin $F$ manidar bulunmuştur, ($F(4,35) = 27,33$, $p < ,05$). Daha sonra Bonferroni testi kullanılarak yapılan önemli doğrusal karşılaştırmalar, morfinin etkilerinin öngörüldüğü şekilde olduğunu göstermiştir. Tuzlu çözelti veya 1-3 denemelerinde farklı bir şekilde morfin aldıktan sonra test ölçümünde morfin alan gruplar, test ölçümünde tuzlu çözelti alan gruplara göre daha uzun tepki zamanı ortalamasına sahip olmuşlardır. ($t(35) = 9,95$, $t_{,006;35} = 2,64$). Standartlaştırılmış etki büyüklüğü 3,21 bulunmuştur. Bu değer iki grup ortalaması arasında neredeyse 3 1/4 standart sapma fark olduğu anlamına gelir.
>
> Farklı şekilde morfin almanın etkisinin Mc-M'de ($\bar{X} = 9$), M-M'den ($\bar{X} = 10$) istatistiksel olarak daha uzun pençe-yalama tepki zamanı ortalaması olacak şekilde görülmüştür. ($t(35) = 6,72$, $t_{\alpha/3} = t_{,0167} = \pm 2,52$). Buradaki standartlaştırılmış etki büyüklüğü 3,36 olarak hesaplanmıştır.

## 12.6 Planlı-Çoklu Doğrusal Karşılaştırmalar

Önsel olarak az sayıda belirlenmiş hipotezlerde, doğrusal karşılaştırmaların ve Bonferroni $t$ testinin kullanımı çokça tavsiye edilmektedir. Aslında benim güçlü tercihim, doğrusal karşılaştırmalarda, en iyi ulaşılabilecek, amacı oldukça iyi tanımlı, az sayıda soru sormaktır. Daha geniş kapsamlı planlı-çoklu karşılaştırmalar, ilgilenmediğiniz soruları da cevaplayacağından güç kaybettirir; fakat bu yaklaşımın, çok sayıda hipotez içeren[10] veya hipotezlere, veri toplandıktan sonra ulaşılan araştırmalarda avantajlı yanları vardır. Bu durumlar için bir dizi sonradan edinilen bilgiye dayalı veya planlı-çoklu teknik vardır.

---

[10] Önceden planlanmış olsa dahi test edilmesi gereken çok sayıda hipotez varsa buradaki prosedürler genellikle Bonferroni $t$ testinden daha güçlüdür.

## Fisher'ın En Küçük Manidar Farklar (LSD) Yöntemi

**Fisher'ın En Küçük Manidar Fark (LSD)**

**Fisher'ın korunmalı t testi**

Planlı-çoklu karşılaştırma yapmak için kullanılan en eski yöntemlerden biri **Fisher'ın En Küçük Manidar Fark (LSD)** (Fisher's Least Significant Test) testidir. Bu test **Fisher'ın korunmalı t testi** (Fisher's protected mânidar testi) olarak da bilinir.

Çok sayıda ortalama olduğu zaman insanlar bazen bu yöntemi kullanmaktan (haklı olarak) yakınırlar. Ancak karşılaştırılacak az sayıda ortalamamız olduğunda (özellikle üç olduğu zaman) bu yöntem oldukça mantıklı ve kullanışlıdır. Bu yöntemi basitçe açıklamak gerekirse, ortalamalar arasındaki ikili karşılaştırmalar için standart Student *t* testini kullanır diyebiliriz. Önsel çoklu *t* testi karşılaştırmaları ile LSD arasındaki tek fark, daha önce de tartışıldığı gibi, LSD'nin varyans analizinden manidar bir *F* gerektirmesidir. Bütün sıfır hipotezleri doğru ise (bütün evren ortalamaları birbirine eşit ise) manidar bir *F* için aile-boyu hata oranının $\alpha$ olması gerekir. Eğer bütün sıfır hipotezleri *doğru değilse*, ancak bazı sınırlandırılmış sıfır hipotezi barındıran ortalama alt kümelerine ilişkin sıfır hipotezleri doğru ise, çoğu zaman durum böyledir, tüm karşılştırmaları içeren *F*, artık *AHO*'yu koruyamaz. Carmer ve Swanson'ın (1973) en popüler çoklu karşılaştırma testleri arasında en güçlü olduğunu göstermesine rağmen, bazıları belirtilen nedenden dolayı bu yöntemi tavsiye etmeyebilir. Eğer deneyinizde üç ortalama varsa LSD yöntemi mükemmeldir, çünkü *AHO*, $\alpha$ düzeyinde kalacaktır ve buna ek olarak standart *t* testinin eklenmiş gücünü kazanırsınız. (Üç ortalama ile *AH*O $\alpha$ değerinde olacaktır çünkü, eğer bütün yokluk hipotezi doğruysa I. Tip hata yapma olasılığınız bütüne ilişkin *F*'yi kullanarak $\alpha$ olur ve arkasından *t*'ye vereceğiniz bir I. Tip hata *AHO*'yu etkilemez. Eğer sıfır hipotezlerinin tamamı doğru değilse ve ancak sınırlandırılmışlarından biri doğru ise, üç ortalama için yine bir I. Tip hata yapma şansınız olacak ve tekrar olasılık $\alpha$'ya eşit olacaktır.) Sıfır hiptezlerinden birinin doğru olduğuna ilişkin iyi bir nedeniz yoksa LSD'yi üç ortalamadan fazla olan durumlarda kullanmanız tavsiye edilmez.

## Studentlaştırılmış (Studentized) Ranj İstatistiği (*q*)

**Studentlaştırılmış Ranj İstatistiği (*q*)**

Birçok çoklu karşılaştırma testi Studentlaştırılmış Ranj İstatistiği'ne ve türevlerine dayandığından ilerlemeden önce bu istatistik atlanmamalıdır. **Studentlaştırılmış Ranj İstatistiği (*q*)** aşağıdaki gibi tanımlanır:

$$q_r = \frac{\overline{X}\text{en büyük} - \overline{X}\text{en küçük}}{\sqrt{\frac{KO_{\text{hata}}}{\text{n}}}}$$

Bu eşitlikte, $\overline{X}_{\text{en büyük}}$ ve $\overline{X}_{\text{en küçük}}$ işlem ortalamalarından en büyük ve en küçük olanı, *r* de kümedeki işlem sayısını belirtir. *q*'nun formülünün *t*'ye benzediğine büyük olasılıkla dikkat etmişsinizdir. Gerçekten;

$$q_r = \frac{\overline{X}\text{en büyük} - \overline{X}\text{en küçük}}{\sqrt{\frac{KO_{hata}}{n}}}$$

$$t = \frac{\overline{X}\text{en büyük} - \overline{X}\text{en küçük}}{\sqrt{\frac{2(KO_{hata})}{n}}}$$

Burada, *t* formülünde tek farklılık paydada $\sqrt{2}$ bulunmasıdır. *q, t*'nin doğrusal bir fonksiyonudur ve $q = t\sqrt{2}$ ile *t*'den *q*'ya geçiş yapılabilir. *t* ve *q* testleri arasındaki asıl fark *q* tablolarından kaynaklanır (EK *q*). Bu tablolar analize dahil edilen ortalama sayısına göre

$q$'nun kritik değerini düzenlememizi sağlar. İki işlem (uygulama) olduğu zaman, ilgili tabloyu kullandığımız sürece, $t$ veya $q$ çözümlemesi yapmış olmamız önemsizleşir. Sadece ortalamamız olduğunda veya ortalama kümemizden karşılaştırmak için tesadüfî olarak iki ortalama seçtiğimizde, $t$ testi uygun olan testtir.[11] Bir ortalama kümesini incelediğimizi ardından da karşılaştırmak için kasıtlı olarak en yüksek ve en düşük ortalamaları seçtiğimizi düşünelim. Bu, açıkça I. Tip hata olasılığını değiştirmeye zorlar. $H_0$ doğru olsa bile, en büyük ve en küçük ortalamalar arasında "manidar fark" çıkma şansı, sıralanmış ortalamalarda komşu terimler arasında fark çıkma şansından daha yüksektir. Studentlaştırılmış ranj istatistiğini kullanışlı hale getiren bir durumdur. Bu istatistik, bu amaç için tasarlanmıştır.

$q$'yu kullanmak için önce ortalamaları en küçükten en büyüğe doğru sıralarız. Daha sonra sıralanmış ortalamalardan karşılaştırmak istediklerimiz arasındaki sıra sayıları hesaba katılır. Komşu ortalamalar için $r = 2$ olur ve herhangi bir değişiklik yapılmaz. Böylece $q_{,05} = t_{0,5}\sqrt{2}$ olur. Komşu olmayan ortalamalar için, karşılaştırılacak olan ortalamalar arasındaki sıra arttıkça, $q$'nun kritik değeri de artar.

$q$'nun kullanılmasına örnek olarak morfin toleransına dair veriyi dikkate alalım. Ortalamalar

| $\bar{X}_1$ | $\bar{X}_2$ | $\bar{X}_3$ | $\bar{X}_4$ | $\bar{X}_5$ |
|---|---|---|---|---|
| 4 | 10 | 11 | 24 | 29 |

n = 8, $sd_{\text{hata}} = 35$ ve $KO_{\text{hata}} = 32{,}00$ olsun. En büyük ortalama 29, en küçük ortalama 4'tür ve kümedeki toplam ortalama sayısı ($r$) 5'tir (bir çok tablonun terminolojisine göre, bunun için bu ortalamalar $r = 5$ basamak uzaktadır).

$$q_1 = \frac{\bar{X}_{\text{en büyük}} - \bar{X}_{\text{en küçük}}}{\sqrt{\frac{MS_{\text{hata}}}{n}}} = \frac{29 - 4}{\sqrt{\frac{32{,}00}{8}}} = \frac{25}{\sqrt{4}} = 12{,}5$$

Dikkat ederseniz $r$ hesaplamaya dahil edilmemiştir. $r$ tabloda sürece dahil olur. EK $q$'dan $r = 5$ ve $sd_{\text{hata}} = 35$ için $q_{,05}(5,35) = 4{,}07$ olur. $12{,}5 > 4{,}07$ olduğundan $H_0$ reddedilir ve "en düşük ortalama ile en yüksek ortalama arasında fark manidardır" kararına varılır.

Alternatif bir yaklaşım $q_{\text{hesaplanan}}$ için çözümleme yapmak, $q_{\text{hesaplanan}}$'ı $q$'nun örneklem dağılımı için referans almak, manidar çıkabilecek en küçük farklılığı bulmak ve bulunan en küçük farklılığı diğer farklılıklarla karşılaştırmak olabilir. Bu gereksiz bir testmiş gibi görünebilir ancak bu yaklaşım SPSS'te de olduğu gibi bilgisayar programlarındaki çoklu karşılaştırmalarda sıkça kullanılmaktadır. Bu bilgisayar dökümünüzün nasıl göründüğünü açıklayabilir. Her iki yol da aynı sonucu verir.

## 12.7 Tukey Testi

**Tukey testi, Tukey HSD (Tukey'nin Gerçek Manidar/ Manidar Fark) testi**

**WSD (Hooly Significant Difference/ Tümüyle Manidar Fark) testi**

Çoklu karşılaştırmalarda yapılan çalışmaların çoğu John Tukey'nin orijinal çalışmasına dayanır ve önemli bir test onun adını taşır. **Tukey testi**, **Tukey HSD (Tukey'nin Gerçek Manidar/Tukey's Honestly Manidar Fark) testi** veya **WSD (Hooly Significant Difference/Tümüyle Manidar Fark) testi** olarak da bilinir. Bu test karşılaştırılmaları Studentlaştırılmış $q$ istatistiği kullanarak yapar. Tek farklı tarafı $q_{\text{HSD}}$ her zaman $q_{\text{r}}$'nin en büyük değerini kullanır. Başka bir deyişle, beş ortalama varsa ve *bütün* farklılıklar test edilecekse bunlar birbirinden beş basamak ilerideymiş gibi yapılır.

11 İki ortalamanın karşılaştırılması için ihtiyacımız olan bütün bilgi varyans analizindeki $F$'den sağlanabilir. Doğrusal karşılaştırmaları hesaplamaya ihtiyacımız yoktur.

Etkisi, bir miktar güç kaybıyla bütün muhtemel yokluk hipotezlerine karşı aileboyu hata oranını $\alpha$ düzeyinde sabitlemektir, bütün hipotezleri tamamlamak değildir. Tukey HSD, $\alpha$ üzerindeki kontrolünden dolayı, birçok insanın tercih ettiği ikili karşılaştırma yöntemidir.

Morfin toleransı verimize Tukey HSD testini uygulamak gerekirse, öncelikli olarak ortalamalarımızı artan büyüklükte dizmemiz gerekir.

| M-S | M-M | S-S | S-M | Mc-M |
|---|---|---|---|---|
| 4 | 10 | 11 | 24 | 29 |

EK $q$'dan $KO_{\text{hata}}$ için 35 serbestlik derecesi ve $r = 5$ için kritik $q$ değeri 4,07 olarak bulunur. Bu kritik değeri bütün doğrusal farklılıklar için kullanacağız.

Gerekli $q$ testini her ortalama çifti için kullanabiliriz, fakat bu çoğu yazılımın izlemediği bir yoldur. $KO_{\text{hata}}$'nın, serbestlik derecesinin, $q$ değerinin ve $q$ kritik değerinin (4,07) bütün doğrusal karşılaştırmalar için aynı olduğunu biliyoruz. Bu durumda ortalamalar arasındaki minimum manidar farkın ne olacağı kolayca belirlenebilir.

$$q = \frac{\overline{X}_1 - \overline{X}_j}{\sqrt{\frac{KO_{\text{hata}}}{n}}} \qquad \text{böylece} \qquad q = \sqrt{\frac{KO_{hata}}{n}} = \overline{X}_1 - \overline{X}_j$$

$$= 4{,}07\sqrt{\frac{32{,}00}{8}} = 8{,}14$$

Bu sonuç, eğer iki ortalama arasındaki fark 8,14'ten daha büyük ise, $q$ testi manidar olacaktır, değilse olmayacaktır. Aşağıdaki ortalamaların manidar olduğunu görebiliriz. Çünkü hepsi 8,14'ten büyüktür.

M–S'ye karşı S-M = 4,00 – 24,00 = -20
M–S'ye karşı Mc-M = 4,00 – 29,00 = -25
M–M'ye karşı S-M = 10,00 – 24,00 = -14
M-M'ye karşı Mc-M = 10,00 – 29,00 = -19
S-S'ye karşı S-M = 11,00 – 24,00 = -13
S-S'ye karşı Mc-M = 11,00 – 29,00 = -18

Buradan aşağıdaki ifade yazılabilir:

(M–S = M–M = S–S) ≠ (S–M = Mc–M)

Eşit işareti yokluk hipotezini reddetmeyeceğimizi gösterir, ortalamaların eşit olduğunu kanıtladığımız anlamına gelmez.

## Eşit Olmayan Örneklem Büyüklükleri ve Varyansların Heterojenliği

Tukey yöntemi öncelikle eşit olan örneklem büyüklükleri için tasarlanmıştır ($n_1 = n_2 = \ldots = n_k \ldots = n$). Sıklıkla deneyler planlandığı şekilde ilerlemez ve kendimizi eşit olmayan sayıdaki gözlemlerle karşı karşıya buluruz. Buna rağmen ortalamaları karşılaştırmak isteriz. Bu probleme ilişkin olarak Tukey HSD testine göre birtakım iyi çalışmalar yapılmıştır (bkz: Games ve Howell, 1976; Keselman ve Rogan, 1977; Games, Keselman ve Rogan, 1981).

Tukey–Kramer yaklaşımı olarak da bilinen bir çözümlemede $\sqrt{KO_{\text{hata}}/n}$ yerine

$$\sqrt{\frac{\frac{KO_{\text{hata}}}{n_i} + \frac{KO_{\text{hata}}}{n_j}}{2}}$$

kullanılmıştır, aksi halde örneklem büyüklüklerini eşit gibi analiz edebilirsiniz. SPSS'in varsayılan çözümlemesi böyledir.

Alternatif ve genellikle tercih edilen, Games ve Howell (1976) tarafından önerilen bir yol vardır. Games ve Howell prosedürü, Bölüm 7'de $t$ testlerine uygulanan Behrens-Fisher yaklaşımı olarak bilinen bir yöntemdir. Bu yaklaşımda yazarlar her ortalama çifti için ayrı

ayrı, ortalamalar arasında ifadesi aşağıda verilen $W_r$ gibi kritik bir fark hesaplanmasını önermektedirler.

$$W_r = \overline{X}_i - \overline{X}_j = q_{,05}(r, sd')\sqrt{\frac{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}{2}}$$

Burada $q_{,05}$ Student'laştırılmış ranj istatistiği tablosundan aşağıdaki serbestlik derecesine göre elde edilecektir.

$$sd' = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}\right)^2}{\frac{\left(\frac{s_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{s_j^2}{n_j}\right)^2}{n_j - 1}}$$

Bu, temel olarak çoklu *t* testlerine ait tartışmada belirtildi. Burada *t* yerine Studentlaştırılmış ranj istatistiğini kullandık; bu SPSS'te isteğe bağlı çözümlemedir. Bu zahmetli bir iştir, ancak çalışmanın desenlenmesi ve verinin toplanmasıyla karşılaştırıldığında daha az gayret gerektirir. Özel yöntemlere olan ihtiyaç, varyans analizi ve ona eşlik eden karşılaştırmalarda, eşit olmayan örneklem büyüklüklerinde varynasların homojenliği sayıltılarının ihlâlinden kaynaklanır. Ayrıca varyans analizinde, örneklem büyüklüklerinden bağımsız olarak, örneklem varyansları hemen hemen eşit ise, $W_r$ için olan formüldeki $s_i^2$ ve $s_j^2$ ifadeleri yerine $KO_{\text{hata}}$ yazabilirsiniz. Yine örneklem büyüklüklerinden bağımsız olmak üzere, eğer varyanslar heterojen ise muhtemelen Games ve Howell prosedürünü uygulamanız gerekecek.

## Ranj Tabanlı Diğer Testler

Studentlaştırılmış Ranj İsatistiği'ne dayanarak geliştirilen çok sayıda test vardır. Bunlar bu kitabın Web sayfasında bulunabilir. Ben sadece bunların neler olduğundan ve birbirinden farklarından söz edeceğim.

### Newman-Keuls Testi

Newman–Keuls testi uzun bir süredir tartışılmakla beraber hâlâ kullanılmaktadır ve SPSS'te bulunmaktadır. Newman–Keuls ve Tukey arasındaki temel fark, Tukey testi *r*'yi bağımsız değişkenin düzey sayısına sabitlerken Newman-Keuls testi sürekli olarak *r*'yi test ettiğimiz en büyükten en küçüğe doğru sıralı ortalamaların sayısına eşit olacak şekilde ayarlamasıdır. (Buradaki mantık, birkaç sayfa önce düzelenmiş Bonferroni testinde kullanılan mantıkla birçok açıdan benzerdir.) Örneğin M-S'e karşı S-S karşılaştırmasına yönelik testte, üç ortalamadaki en yüksekler ve en düşükler olduğu için r = 3 olur. Bunu yapmamızın avantajı daha güçlü bir teste sahip olmamızı sağlamasıdır. Dezavantajı $\alpha$ üzerindeki kontrolümüzün bir kısmını kaybetmemizdir.

### Ryan Yöntemi

REGWQ testi olarak da bilinen Ryan prosedürü, Tukey testinin Newman–Keuls doğrultusunda yapılmış bir modifikasyonudur. Bu test *r* değerinde daha önce de olduğu gibi düzeltme yapar, ancak bunu FEW değerini maksimumu $\alpha$ olacak şekilde sınırlandırarak yapar. Bu testi Tukey'ye tercih ederim ama çok nadir kullanılır.

### Scheffé Testi

**Scheffé testi**

**Scheffé testi** Tukey'nin yöntemine dayanır ve en eski testlerimizden biridir. Bu test sadece ikili karşılaştırmalar için değil, herhangi ve bütün çoklu karşılaştırmalarda $\alpha$ değerini ,05 düzeyinde tutar. Scheffé, testin ikili karşılaştırmalara uygulanması durumunda güç kaybettiğini belirtmiş ve ikili karşılaştırmalarda kullanılmasını tavsiye *etmediğini* söylemiştir. Buna rağmen SPSS ve diğer programlar onun tavsiyesine uymayarak programı kullanmaktadırlar. (Scheffé'i kınamayınız, bunun yapılmaması gerektiğini söylemiştir.) Daha fazla bilgi için Howell (2010) incelenebilir.

### Dunnett Testi

**Dunnett testi**

**Dunnett testi** bir işlem ortalamasını (genellikle kontrol grubu), diğer işlem ortalamasıyla karşılaştırmak için tasarlanmıştır. Bu amaç için çok güçlü olmakla beraber, diğer testlerdeki esneklik bunda yoktur. Web sayfasında da bulunabilen, kendi tablo değerlerini kullanır.

## Benjamini–Hochberg Testi

Birbiriyle rekabet halinde olan ikili karşılaştırma testleri hakkında sadece kısaca bilgi veriyordum, ama yeni bir yaklaşımdan bahsetme ihtiyacı hissettim. Çünkü istatistikçiler bu yaklaşımı sevmektedirler ve bu yaklaşım zaman içinde davranış bilimlerinde daha önemli bir rol oynayacağa benzemektedir. Tartıştığımız çoklu karşılaştırma testlerinin her biri aileboyu hatayı (*AHO*'yu) kontrol etmeye odaklandı. Bunlardan bir kısmı da seride ilerledikçe kritik değerini değiştiren ardışık testlerdi. Benjamini ve Hocherg (1995, 2000) popülerleşmeye başlayan, ardışık ve *AHO*'ya dayanmayan bir test geliştirmişlerdir. Aileboyu hata yaklaşımı yerine **Yanlış Buluş Oranı (YBO) (False Discovery Rate (FDR))** adını verdikleri bir yöntemi savunmaktadırlar. Tukey 1950'li yıllarda *AHO*'nun savunuculuğunu yapmaya başladığı zaman, belki de bilmeden, bizim düşüncemiz olan sadece bir I. Tip hatanın kontrolüne yöneldi. Aileboyu bir hata oranı hesapladığınızda *bir veya daha fazla* I. Tip hata ile uğraşıyorsunuz demektir. Bu, bir tek I. Tip hata yapmış olsanız bile, vermiş olduğunuz kararlar kümesinin hatalı olmasına yol açarsınız demekle aynı anlama gelir. (İlginç bir biçimde, II. Tip hata yaptığımız zaman çıkarımlarımızın hatalı olduğunu düşünmeyiz.) Hocherg ve Benjamini probleme farklı bir açıdan yaklaşarak "Bulduğumuz manidar sonuçların (buluşların), yanlış buluşlar olma yüzdesi nedir?" diye bir soru sormuşlardır. Dokuz karşılaştırma (basit doğrusal karşılaştırmalar, kompleks karşılaştırmalar, tek ortalama testleri ve diğer testler olabilir) yaptığımızı düşünelim ve dört manidar etki bulmuş olalım. Bizim tarafımızdan bilinmemekle beraber bunlardan biri gerçekten I. Tip hata olsun. Bu durumda YBO aşağıdaki gibi tanımlanır:

**Yanlış Buluş Oranı (YBO)**

$$YBO = \frac{\text{Hatalı Reddetme Sayısı}}{\text{Toplam Reddetme Sayısı}} = \frac{1}{4} = ,25$$

YBO üzerine mükemmel bir tartışması olan Maxwell ve Delaney (2004) tarafından yapılan bir "düşünce deneyi"nden örnek verebilirim. 10 yokluk hipotezini test ettiğimiz bir durum düşünün. Bu yokluk hipotezlerinden üçünün yanlış, birinin doğru olduğunu biliyor olalım. Bu deneyi zihnimizden 100 defa yaptığımızı ve her deneyde bu 10 sıfır hipotezini test ettiğimizi düşünelim. Daha sonra yanlış sıfır hipotezlerini reddebilecek çok büyük bir gücümüzün olduğunu böylelikle üç yanlış sıfır hipotezini neredeyse hep reddettiğimizi varsayalım. Son olarak aileboyu hata oranını ,20'ye ayarladığımız bir kritik değerimiz olsun. (Muhtemelen ,20'nin çok yüksek olduğunu düşüneceksiniz ama bana biraz sabredin.) Sonra 100 hipotetik deneyden %80'inde (1 - %20) I. Tip hata yapmadığımızı düşünelim. (Nerdeyse her zaman gerçekten manidar olan üç sıfır hipotezini reddettik ve diğer kalan yediyi kabul ettik.) Ayrıca, aileboyu hata oranını ,20'ye ayarladığımızdan her seferin %20'sinde bir I. Tip hata yapabiliriz (herhangi bir deneyde iki I. Tip hata yapmadığımızı varsayarak). Gücümüzün çok büyük olmasından dolayı neredeyse her zaman üç yanlış

yokluk hipotezini reddettik. Belki sizi ürkütecek ama buradaki *AH*O değerimiz ,20. Peki YBO için durum nedir? Yukarıda verilen tanıma göre, deneylerin %80'inde hatamız olmamalı. Diğer 20 deneyde bir I. Tip hata yapabiliriz, 3 doğru ret kararında bulunabiliriz, 20 deney için YDO, ¼ = ,25 olur ve diğer 80 için YDO 0 olur. 100 deneme için ortalama FWE ,20 iken YDO ,05 olur. FWE için ,20 olarak ayarlanan kritik değer YDO'yu ,05'te bıraktık. Bu kritik değeri nasıl seçtiğimiz bir problemdir. Bu genel durum için oldukça karmaşık olmakla birlikte, bağımsız doğrusal karşılaştırmalarda veya ikili doğrusal karşılaştırmalarda oldukça kolaydır. Bakınız Keselman, Cribbie ve Holland (1999). Bu bölümde ikili doğrusal karşılaştırmaların şiddetli bir savunucusu olduğumdan bu kısımla kendimizi sınırlandırmamız çok sıkıcı olmamalı.

**Benjamini ve Hochberg'in Doğrusal Yükselme (Linear Step Up (LSU)) Yöntemi**

İzleyeceğimiz yöntem **Benjamini ve Hochberg'in Doğrusal Yükselme (Linear Step Up (LSU))** yöntemi adını alır.[12] Bu testin mantığına yer vermeyeceğim, fakat bu Benjamini ve Hochberg'in (1995) çalışmasında ve Maxwell ve Delenay'nin (2004) bölümünde oldukça açıktır. Thissen, Steinberg ve Kuang (2002), Benjamini ve Hochberg'in testini veri sayfasına taşımak için basit bir yöntem sunmuşlardır. Testi yapabilmemiz için bu tartışmayı basamaklar halinde yapılandıracağım.

Siegel'ın morfin verisinde 10 ikili karşılaştırma yaptığımızı varsayalım. Sonuçlar $p$ değerlerine göre sıralanmış şekilde Tablo 12.4'te gösterilmiştir. $i$ ile adlandırılmış sütun, karşılaştırmanın indeksidir ve $p$'ler en yüksekten en düşüğe doğru sıralanmıştır. Tablonun, testin kritik değerlerinin yer aldığı kritik bölümü $p_{\text{krit}}$ ile adlandırılmıştır. Bu değer, "$i$" indeks, "$k$" test sayısı (burada $k = 10$) ve $\alpha$'da istenilen YDO (burada $\alpha$=,05) değeri olmak üzere aşağıdaki şekilde tanımlanır:

$$p_{\text{krit}} = \left(\frac{i}{k}\right)\alpha$$

Testi gerçekleştirmek için tabloda aşağıya doğru ilerleriz. $p > p_{\text{crit}}$ ise yokluk hipotezini kabul ederiz ve sıradaki satıra geçeriz $p < p_{\text{crit}}$ ise yokluk hipotezi ve bütün sonra gelenleri reddederiz.

Benjamini–Hochberg testini kullanarak M-M'ye karşı S-S, S-M'ye karşı Mc-M ve M-S'ye karşı ve M-M'nin birbirinden farklı olmayan çiftler olduğu söylenebilir. Diğer bütün doğrusal farklılıklar manidardır. Tukey testi kullanıldığında, Benjamini–Hochberg testinde reddedilen M-S'ye karşı S-S reddedilmeyip altı yokluk hipotezi reddedilmişti. Bu bizim manidar bulmak istediğimiz bir doğrusal farklılıktır, çünkü morfin alındıktan sonra aynı şartlarda tuzlu çözelti alınmasının, duyarlılığı arttıracağını söylemektedir.

**Tablo 12.4** Siegel verisi üzerinde Benjamini–Hochberg testi

| *Grup* | *t* | *p* | *i* | $p_{\text{krit}}$ | Manidarlık |
|---|---|---|---|---|---|
| M-M'ye karşı S-S | -,354 | ,726 | 10 | ,05 | Hayır |
| S-M'ye karşı Mc-M | -1,768 | ,086 | 9 | ,045 | Hayır |
| M-S'ye karşı M-M | -2,121 | ,041 | 8 | ,040 | Hayır |
| M-S'ye karşı S-S | -2,475 | ,018 | 7 | ,035 | Evet |
| S-S'ye karşı S-M | -4,596 | ,00007 | 6 | ,030 | Evet |
| M-M'ye karşı S-M | -4.950 | ,00003 | 5 | 025 | Evet |
| S-S'ye karşı Mc-M | -6,364 | ,00000 | 4 | ,020 | Evet |
| M-M'ye karşı Mc-M | -6,717 | ,00000 | 3 | ,015 | Evet |
| M-S'ye karşı S-M | -7,071 | ,00000 | 2 | ,010 | Evet |
| M-S'ye karşı Mc-M | -8,839 | ,00000 | 1 | ,005 | Evet |

© Cengage Learning 2013

12 Benjamini ve Hochberg (2000) burada verilen bir testin, bazen "adapte edilmiş" olarak adlandırdıkları türevini önermişlerdir. Bu test LSU'dan daha güçlüdür ancak daha külfetlidir. Bu testler SPSS'in sağladığı Hochberg GT2'den farklıdır. R'de yazılmış bir programa (BenjaminiHochbergLSU.R) kitabın Web sayfasından ulaşılabilir.

## 12.8 Hangi Test?

Özel durumumuz için en uygun çoklu karşılaştırma yöntemini seçmek kolay değildir. Birçok test vardır ve bunlar çeşitli şekillerde farklılaşırlar. Seçim, ilk başta iki uç durum düşünülürse biraz kolaylaşabilir.

Eğer testinizi daha önceden planlamışsanız ve sadece bir karşılaştırma yapacaksanız, standart *t* testini (gerekirse heterojen varyanslar için düzelterek) uygulamanızı, kompleks bir karşılaştırmanız varsa doğrusal karşılaştırmaları öneririm. Eğer çeşitli önsel doğrusal karşılaştırmalar varsa, ikili olması şart değil, çoklu Bonferroni'nin *AHO*'yu kontrol ederken aynı zamanda gücü maksimize ederek iyi iş yapacağınızı söyleyebilirim.

Eğer çok sayıda grubunuz varsa ve çok sayıda karşılaştırma yapmak istiyorsanız, bütün karşılaştırmalarla ilgilenirsiniz veya ilgilenmezsiniz, Tukey testi (veya REGWQ) tavsiye edilebilir. Burada tavsiye ettim ama, Schefe'i önerebileceğim bir durum düşünmüyorum. Bunu taviz vermeyenler sevmektedir.

Benjamini-Hochberg testine gelince; bu farklı, tamamen hata oranlarını kontrol ettiğinden tablolaştırılması zor bir testtir. Farklı hata oranları ile çalışırken bir test diğerinden daha güçlüdür demek uygun olmaz. Benjamini-Hochberg testine ilişkin zaafım var, çünkü bir yanlış reddedişin bir karar ailesini geçersiz kılması fikrine dayanmamaktadır. Eğer nadiren oluşan I. Tip hatayı kabul ederek diğer doğrusal karşılaştırmalar için daha fazla güç kazanmaya istekliyseniz, bu testi çok tavsiye ederim.

## 12.9 Bilgisayar Çözümlemeleri

Birçok bilgisayar programı, çoklu karşılaştırmalar yapabilmektedir, ancak bütün paketlerde bütün yöntemler bulunmamaktadır. Gösterim 12.1'de morfin verisine ait SAS kullanılarak yapılan analizin sonucu yer almaktadır. SAS'ı seçmemin nedeni, çok geniş bir yöntem seçeneğine sahip olması ve kullanılan belli başlı paketlerden biri olmasıdır. Amacımız için SPSS çıktılarından daha fazla yararlı olabilecek ayrıntı barındırmasıdır. Basit ikili karşılaştırmalar için tamamen uygunsuz olduğunu söylediğim Scheffé testini de dahil ettim.

Gösterim 12.1, SAS program komutları ve genel varyans analizi ile başlamaktadır. Bu analiz Tablo 12.1'de yer alan özet bilgilerle de uyumludur. Basitçe $\eta^2$ değeri $R^2$ = ,757'dir. Görebileceğiniz gibi, deneysel düzenleme varyansın önemli bir kısmını açıklamaktadır. Kalan kısım daha önce kısaca söz ettiğim Newman-Keuls, Ryan, Tukey ve Scheffé testleri sonuçlarıdır.

En az tutucu test olan Newman-Keuls testi, koşullar arasında en çok fark bildiren test olmuştur. Eğer önce ortalamalara ve çıktının sonunda yer alan "SNK Gruplaması" kısmına bakarsanız, A, B ve C harflerini içeren bir sütun göreceksiniz. Aynı harfe sahip durumlar birbirinden farklılaşmıyor şeklinde değerlendirilmiştir. Mc-M ve S-M durumlarının ortalamaları manidar olacak şekilde birbirinden farklılaşmamaktadır. Ancak S-S, M-M, M-S'den farklı ortalamaya sahiptirler. Durum S-S ve M-M, B harfine sahiptir ve ortalamaları birbirinden manidar olarak farklı değildir, fakat diğer üç durumun ortalamasından farklıdır. Son olarak M-S durumunun ortalaması bütün diğer ortalamalardan farklıdır.

```
Data Siegel;
      Infile 'Siegel.dat';
      Input Subject Condition latency;
Run;

Proc GLM Data = Siegel;
      Class Condition;
      Model Latency = Condition/SS3;
      Means Condition/ SNK Tukey REGWQ Scheffe;
Run;
```

The SAS System 11:15 Wednesday August 18, 2010

The GLM Procedure

Dependent Variable: LATENCY

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 3497.600000 | 874.400000 | 27.33 | < .0001 |
| Error | 35 | 1120.000000 | 32.000000 | | |
| Corrected Total | 39 | 4617.600000 | | | |

| $\eta^2$ → R-Square | Coeff Var | Root MSE | LATENCY Mean |
|---|---|---|---|
| 0.757450 | 36.26189 | 5.656854 | 15.60000 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| CONDITION | 4 | 3497.600000 | 874.400000 | 27.33 ← $F$ for Condition | <.0001 |

**Student-Newman-Keuls** Test for LATENCY

NOTE: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha = 0.05 Error Degrees of Freedom = 35 Error Mean Square = 32

| Number of Means | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Critical Range | 5.7420599 | 6.9219411 | 7.6279954 | 8.1319062 ← $w_5 = q_5$ |

Means with the same letter are not significantly different.

| SNK Grouping | Mean | N | Condition |
|---|---|---|---|
| A | 29.000 | 8 | Mc-M |
| A | 24.000 | 8 | S-M |
| B | 10.000 | 8 | S-S |
| B | 10.000 | 8 | M-M |
| C | 4.000 | 8 | M-S |

Adapted from output by SAS Institute Inc.

**Gösterim 12.1** *(Devam ediyor.)*

**Ryan-Einot-Gabriel-Welsch** Multiple Range Test for LATENCY

NOTE: This test controls the Type I experimentwise error rate.

Alpha = 0.05 Error Degrees of Freedom = 35 Error Mean Square = 32

Larger than for SNK

| Number of Means | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Critical Range | 6.8765475 | 7.5391919 | 7.6279954 | 8.1319062 |

Same as SNK

Means with the same letter are not significantly different.

| REGWQ Grouping | Mean | N | CONDITION |
|---|---|---|---|
| A | 29.000 | 8 | Mc-M |
| A | 24.000 | 8 | S-M |
| B | 11.000 | 8 | S-S |
| B | 10.000 | 8 | M-M |
| B | 4.000 | 8 | M-S |

---

**Tukey's Studentized Range (HSD)** Test for LATENCY

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha = 0.05 Error Degrees of Freedom = 35 Error Mean Square = 32

Critical Value of Studentized Range = 4.06595

Minimum Significant Difference = 8.1319 ← Critical range for all differences

Means with the same letter are not significantly different.

| Tukey Grouping | Mean | N | CONDITION |
|---|---|---|---|
| A | 29.000 | 8 | Mc-M |
| A | 24.000 | 8 | S-M |
| B | 11.000 | 8 | S-S |
| B | 10.000 | 8 | M-M |
| B | 4.000 | 8 | M-S |

---

**Scheffe's** Test for LATENCY

NOTE: This test controls the Type I experimentwise error rate.

Alpha = 0.05 Error Degrees of Freedom = 35 Error Mean Square = 32

Critical Value of F = 2.64147

Minimum Significant Difference = 9.1939 ← Critical range for all differences

Means with the same letter are not significantly different.

| Scheffe Grouping | Mean | N | CONDITION |
|---|---|---|---|
| A | 29.000 | 8 | Mc-M |
| A | 24.000 | 8 | S-M |
| B | 11.000 | 8 | S-S |
| B | 10.000 | 8 | M-M |
| B | 4.000 | 8 | M-S |

---

**Gösterim 12.1** *(Devamı)*

Tablonun biraz yukarı kısmına bakarsanız, bu testin aileboyu (burada "deneyboyu" olarak adlandırılır) hata oranlarıyla nasıl ilgilendiği hakkında bir açıklama göreceksiniz. Daha önce söylediğim gibi, Newman-Keuls sıfır hipotezinin tümüne karşı aileboyu hatayı $\alpha$ düzeyinde tutar, fakat yokluk hipotezleri kümesinin bir alt kümesi doğru ise bunun artmasına izin verir. Daha sonrasında $\alpha$ = ,05 düzeyinde çalışıldığını, hata terimi olan $KO_{\text{hata}}$ = 32,00 için serbestlik derecesinin 35 olduğunu gösterir bir ifade görürsünüz. Bu ifadeyi takiben kritik ranj değerleri bilgisini görürsünüz. Bunlar *r*'nin farklı değerlerinde olabilecek manidar en küçük ortalamalar arası farklılıklardır. Kritik ranjlar şu şekildedir:

$$W_r = q_{,05}(sd_e)\sqrt{\frac{KO_{hata}}{n}}$$

Örneğin, *r* = 3 ( en büyük ve en küçük üç ortalama arasındaki fark)

$$W_r = q_{,05}(df_{hata})\sqrt{\frac{KO_{hata}}{n}} = 3,46\sqrt{\frac{32}{8}} = 3,46(2) = 6,92$$

Çünkü, üç basamaktaki farklılıkların hepsi (ör: 29-11 = 18; 24 -10 = 14; 11 - 4 = 7) 6,92 değerinden daha büyük olduğu için, hepsi manidar olarak ifade edilebilir.

Gösterim 12.1'in diğer kısmı Ryan (REGWQ) testi sonuçlarını içerir. Dikkat ederseniz *r* = 2 ve *r* = 3 için olan kritik ranjlar Newmann-Keuls'a göre olanlardan daha büyüktür (Tukey'ye göre daha küçük olduğu halde). Sonuç olarak, *r* = 3 için 7,54 fark değerini geçmemiz gerekirken, 11 ile 4 arasındaki fark sadece 7'dir. Bundan dolayı bu test Grup 1 (M-S)'yi Grup3 (S-S)'den farklı bulmaz. Bu daha liberal olan Newman-Keuls'da farklı bulunmuştu. Bununla birlikte aile-boyu hata oranı bu karşılaştırma kümesinde $\alpha$ = ,05 iken Newman-Keuls'da bu değer yaklaşık olarak $\alpha$ = ,10 olmuştur.

Tukey Testi biraz farklı sunulmuştur. Tukey testinde ortalamalar arasındaki farkın, sıralamada nerde oldukları önemsiz olacak şekilde, manidar olabilmesi için 8,1319 kritik ranj değerinin geçilmesi gerektiği görülebilir. Bu veri kümesi için olan çıkarımlarımız, Ryan ile aynıdır.

Scheffé testi diğerlerinden oldukça farklı çalışsa da, bütün ikili karşılaştırmalar için bir ranj değeri hesaplamak mümkündür. Gösterim 12.1'de bu değerin 9,1939 olduğu ve bu değerin Tukey için olan değerden neredeyse 1 puan yukarıda olduğu görülebilir. Scheffé yönteminin ikili karşılaştırmalardaki ileri derecede tutuculuğunu yansıtır ve benim bu prosedürün kullanımına olan itirazımı gösterir.

SAS ayrıca Bonferroni ve Dunn-Sidak testleri de dahil birçok çoklu karşılaştırma üretebilir. Bunları burada göstermedim çünkü, bütün olası ikili ortalama karşılaştırmaların yapılmasında bu testlerin kullanımı saçmadır. Ryan veya Tukey testleri neredeyse her zaman daha güçlüdür ve aile-boyu hata oranını koruyabilmektedirler. Daha sınırlı sayıda ikili karşılaştırmam olduğunu ve ilgilenmediğim doğrusal farklılıklara bakmama sözüyle SAS (BON)'de Bonferroni testini kullanabilirdim.

## 12.10 Eğilim Analizi

Tartıştığımız katsayılar, gruplar arasında karmaşık karşılaştırmalar veya basit ikili karşılaştırmalar olup olmadığına bakılmaksızın grup ortalamaları arasındaki karşılaştırmalarla ilgilidir. Bununla birlikte, bağımsız değişken tarafından tanımlanan grupların bir sürekli aralık boyunca sıralandığını varsayalım. Bir örnek kalp hastalığını önlemede aspirinin yararlı etkileri üzerine bir çalışma olabilir. Bireylere, genellikle "bebek aspirini" olarak adlandırılan, 81 mg'lık aspirini günlük 1, 2, 3, 4 veya 5 dozda almak isteyip istemedikleri sorulur. Bu çalışmada 4-doz hapın 2-doz haptan daha iyi olup olmadığıyla ilgilenilmemektedir. Örneğin, aspirinin yararlı etkileri ilâcın dozajı artırılarak artar mı? Başka bir

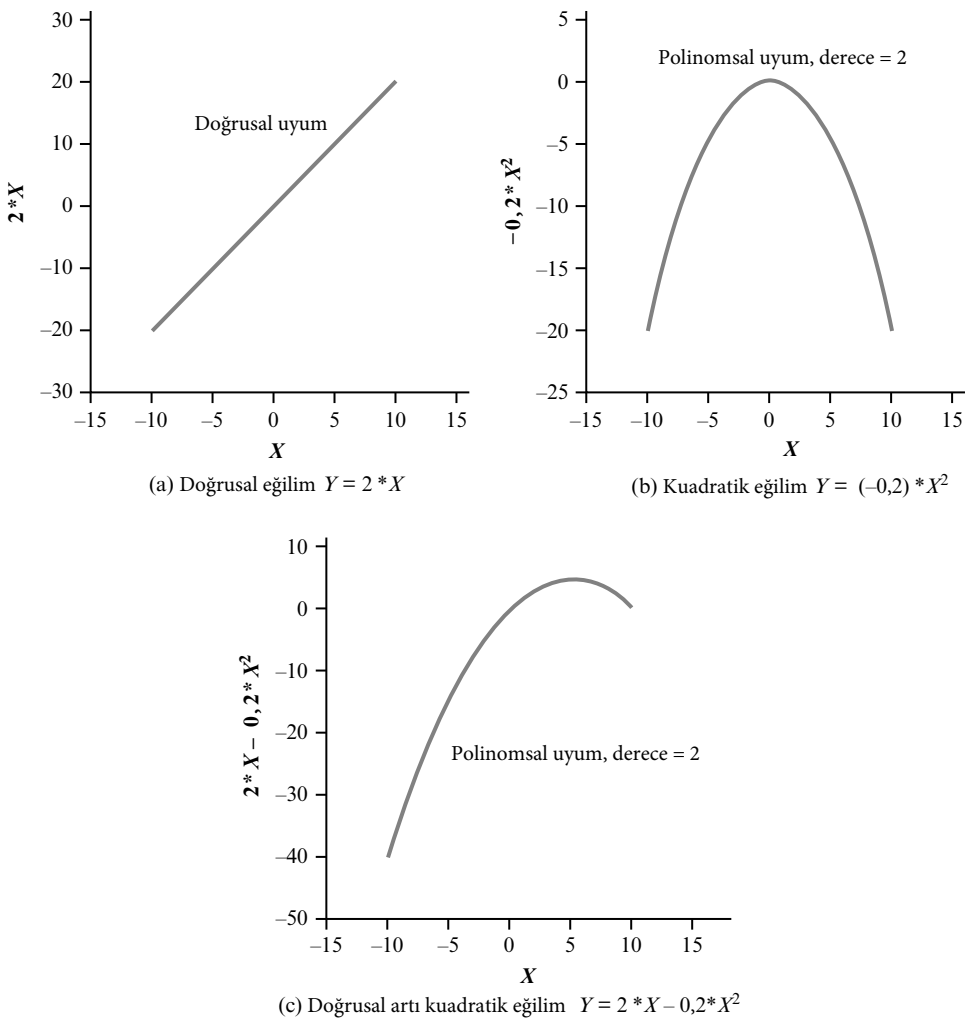
**eğilim**

ifadeyle, bu, belirli ortalamalar arasındaki çoklu karşılaştırmalardan ziyade etkililikteki **eğilim**le ilgilidir. Diğer bir deyişle, belirli ortalamalar arasındaki çoklu karşılaştırmalardan çok eğilimin etkisiyle ilgilenmekteyiz.

Aspirin örneğine devam etmek için, iki muhtemel sonucu göz önünde bulundurun. Bir sonuçta etkililiğin dozaj ile doğrusal olarak arttığını bulabilirsiniz. Bu durumda, en azından test edilen dozaj aralığı içinde, daha fazla aspirin alındığında, daha büyük etki olur. İkinci bir alternatif bulgu, bir noktaya kadar etkilik dozaj ile artar, ancak daha sonra doz etkililiğiyle ilgili eğri düzleşir ve hatta belki de azalır. Bu, ya kuadratik ilişkidir veya hem doğrusal hem de kuadratik bileşenlerle ilişkidir. Bu tür ilişkileri bulmak optimal doz tavsiyesi için önemli olacaktır.

**doğrusal ve kuadratik (ikinci dereceden) fonksiyonlar**

Şekil 12.2'de tipik **doğrusal** ve **kuadratik (ikinci dereceden) fonksiyonlar** gösterilmiştir. Kuadratik fonksiyonları düzgünce karakterize etmek zordur, çünkü fonksiyonun şekli hem $\chi^2$ katsayısının işaretine hem de $X$'in işaretine bağlıdır ($X$ negatiften pozitife geçtiğinde eğrinin yönü değişir ve $X$'in pozitif değerleri için, eğer katsayı pozitifse eğri yükselir, negatifse eğri düşer.) Ayrıca Şekil 12.2, doğrusal ve kuadratik bileşenlerin her ikisinin fonksiyonlarını içerir. Burada kuadratik bir fonksiyon tarafından belirlenen eğriliğin yükselen bir doğrusal eğilim üstüne getirilmiş olduğunu görebilirsiniz.




(a) Doğrusal eğilim $Y = 2 * X$

(b) Kuadratik eğilim $Y = (-0{,}2) * X^2$

(c) Doğrusal artı kuadratik eğilim $Y = 2 * X - 0{,}2 * X^2$

© Cengage Learning 2013

**Şekil 12.2** Tipik doğrusal ve kuadratik fonksiyonlar

Eğilim testleri, tartıştığımız karşılaştırma prosedürlerinden önemli bir şekilde farklıdır. Önceki örneklerin tümünde, bağımsız değişken genel olarak nitel idi. Bu şekilde, örneğin, herhangi bir sırada seçtiğimiz morfin-tolerans örneğindeki grupları kaydedebiliriz. Ayrıca, doğrusal farklılıklar için *F* veya *t* değerleri, sadece ortalamaların sayısal değerlerine bağlıdır, belirli gruplara ait belirli ortalamalara bağlı değildir. Şu anda düşündüğümüz analizde, *F* veya *t* değeri her iki grubun ortalamalarına ve bu ortalamaların belirli sıralamalarına bağlıdır. Aspirin örneğini kullanarak bunu biraz daha farklı açıklamak için, en büyük ve en küçük ortalamalar arasındaki test, hangi grubun hangi ortalamaya sahip olduğundan etkilenmeyecektir. Bununla birlikte, sırasıyla 1-grain (,065 gram) ve 5-grain grupları en küçük ve en büyük ortalamalara sahip olsaydı, sonra 4 ve 2-grain grupları en küçük ve en büyük ortalamaya sahip olsaydı eğilim analizi sonuçları oldukça farklı olurdu. (Benzer bir nokta ki-kare testinin yönlü olmayan tartışmasında Kesim 6.7'de yapılmıştır.)

## Alkol ve Saldırganlık

Dikkat dağıtıcı görev varlığında alkol ve saldırganlık arasındaki ilişki üzerine Giancola ve Corman (2007) tarafından yapılan çalışma örnek olarak alınacaktır. Bu, Bölüm 11'de başlayan bir çalışmadır, veri ve özet tablolar Tablo 12.5'te verilmiştir. Bağımsız değişkenin düzeyleri (D0, D2, D4, D6 ve D8) katılımcıların dikkat dağıtıcı görevlerini bildirmek zorunda oldukları ışıklandırılmış kare sayısını temsil etmektedir. Bunun 0'dan 8'e doğrusal olarak arttığına dikkat ediniz. Belirgin sorular, dikkatin dağılması ile saldırganlığın doğrusal olarak azalıp azalmayacağı, önce bazı düzeylerde azalıp sonra sabit kalıp kalmayacağı, veya önce azalıp sonra yeniden artmaya başlayıp başlamayacağı şeklindedir. Şöyle bir baktığımızda sonuncusunu öneririm ancak bunu test etmemiz gerekir.

Standart tek yönlü varyans analizi aşağıdaki özet tabloda verilmiştir:

**Özet Tablo**

| Kaynak | *sd* | KT | KO | F |
|---|---|---|---|---|
| Uygulama | 4 | 62,460 | 15,615 | 6,90* |
| Hata | 55 | 124,458 | 2,263 | |
| Toplam | 59 | 189,918 | | |

*p < ,05

**Tablo 12.5** Görev zorluğunun fonksiyonu olarak uygulanan şok düzeyi

| | D0 | D2 | D4 | D6 | D8 | Toplam |
|---|---|---|---|---|---|---|
| | 1,28 | -1,18 | -0,41 | -0,85 | 2,01 | |
| | 1,35 | 0,15 | -1,25 | 0,14 | 0,40 | |
| | 3,31 | 1,36 | -1,33 | -1,38 | 2,34 | |
| | 3,06 | 2.61 | -0,47 | 1,28 | -1,80 | |
| | 2,59 | 0,66 | -0.60 | 1,85 | 5,00 | |
| | 3,25 | 1,32 | -.1,72 | -0,59 | 2,27 | |
| | 2,98 | 0,73 | -1,74 | -1,30 | 6,47 | |
| | 1,53 | -1,06 | -0,77 | 0,38 | 2,94 | |
| | -2,68 | 0,24 | -0.41 | -0,35 | 0,47 | |
| | 2,64 | 0,27 | -1,20 | 2,29 | 3,22 | |
| | 1,26 | 0,72 | -0,31 | -0,25 | 0,01 | |
| | 1.06 | 2,28 | -0,74 | 0,51 | -0,66 | |
| **Ortalama Ortalama** | 1,802 | 0,675 | -0,912 | 0,544 | 1,889 | 0,800 |
| **St. Sap.** | 1,656 | 1,140 | 0,515 | 1,180 | 2,370 | 1,800 |
| **Var.** | 2,741 | 1,299 | 0,265 | 1,394 | 5,616 | 3,168 |

© Cengage Learning 2013

Özet tablodan beş grup arasında önemli farklılıklar olduğu görülmektedir. Bu farklılıkları incelemenin bir yolu ışıklandırılmış kare sayısının bir fonksiyonu olarak grup ortalamalarının grafiğini çizmektir. Bu durum Şekil 12.3'te gösterilmektedir ve saldırganlığın azalmasında dikkat dağılmasının etkililiğinin bir noktaya kadar arttığı ancak dikkat dağınıklığının daha yüksek düzeylerinde saldırganlığın taban seviyesine geri döndüğü açıkça görülmektedir.

Genel bir varyans analizi aslında, yatay düz doğrunun $Y = 0{,}800$ (genel ortalama) boyunca olması durumunda, veri bu doğruya yeterince uyacak mı sorusunu sorar. $F$, yokluk hipotezinin reddedilmesine yol açmıştır çünkü, örneklem ortalamaları 1,9 kadar yüksek ve -,9 kadar düşüktür. Bir sonraki soru yatay olmayan doğrunun veriye iyi uyum sağlayıp sağlamadığı şeklindedir. Şekil 12.3'e bakıldığında, bu durumun muhtemelen böyle olmadığı ya da en azından zayıf bir uyumun olduğu fikrini vermektedir. Daha sonra, veride kalan sistematik artık (hata olmayan) varyansın daha sonra bir doğrusal fonksiyona uyup uymayacağı ve eğer öyleyse, bu artık varyansın kuadratik bir fonksiyonla açıklanıp açıklanamayacağı sorularını izlemeliyiz.

Eğilim analizini çalıştırmak için, doğrusal ve dik karşılaştırmalar başlıkları altında tartışılan materyallere geri döneceğiz. (Son cümledeki *doğrusal* kelimesi karıştırılmamalıdır. Aynı yaklaşım, bir kuadratik fonksiyona gelindiğinde kullanılacaktır. Bu bağlamda doğrusalın basitçe anlamı, katsayılardan ve ortalamalardan hiçbirinin üssünün yükseltilmediği bir doğrusal kombinasyon oluşturacağız.)

Bölüm 12.3'te doğrusal karşılaştırmayı aşağıdaki gibi tanımlamıştık:

$$\psi = a_1\overline{X}_1 + a_2\overline{X}_2 + a_2\overline{X}_3 + ..... + a_k\overline{X}_k = \sum a_j\overline{X}_j$$

**polinom eğilim katsayıları**

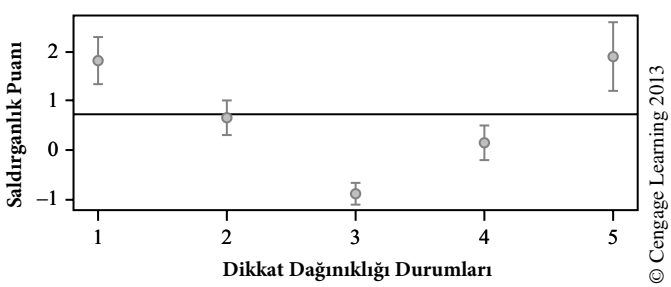
Daha önceki yapılanla, şimdiki arasındaki tek fark kullandığımız katsayılardır. Burada gruplarda eşit sayıda bireyler bulunmakta ve yatay eksen üzerindeki değerler eşit aralıklıdır; doğrusal, kuadratik ve yüksek dereceden fonksiyonlar (**polinom eğilim katsayıları**) için katsayılar kolayca tablolaştırılmakta olup bunlar EK Polinomsal'da bulunabilir. EK Polinomsal'da bulunan beş grup için doğrusal ve kuadratik katsayılar aşağıdaki gibidir:

| | | | | | |
|---|---|---|---|---|---|
| **Doğrusal:** | –2 | –1 | 0 | 1 | 2 |
| **Kuadratik:** | 2 | –1 | –2 | –1 | 2 |

Kubik ve kuadratik sayıları kullanmayacağız, fakat bunların kullanılması aşağıda daha açık hale gelecektir. Herhangi dik doğrusal katsayılar kümesinde olduğu gibi, $\sum a_j = 0$ ve $\sum a_i b_j = 0$ gereklilikleri karşılanmalıdır. Katsayılar "standart ayar" formunda olmamalıdır çünkü katsayı değerlerinin mutlak toplamı 2'ye eşit değildir. Bu, burada problem değildir.

Kesim 12.3'ten hatırlanacağı gibi, doğrusal karşılaştırma için kareler toplamını aşağıdaki gibi hesaplarız:

$$KT_{\text{karşılaştırma}} = \frac{n\psi^2}{\sum \alpha_j^2}$$




**Şekil 12.3** Dikkat dağınıklığının fonksiyonu olarak saldırganlık

Örneğimizdeki veride,

$$\psi_{\text{Doğrusal}} = (-2)1,802) + (-1)0,675 + (0) - 1,912 + (1)0,544 + (2)1,889$$
$$= 0,0425$$
$$KT_{\text{Doğrusal}} = \frac{n\psi^2}{\sum a_j^2} = \frac{12(0,0425^2)}{10} =,002$$

Bütün doğrusal farklılıklar gibi, bu farklılık da bir serbestlik derecesine sahiptir ve bu nedenle, $KT_{\text{doğrusal}} = KO_{\text{doğrusal}}$'dır. Doğrusal farklılıklar için kareler ortalamasını $KO_{\text{hata}}$'ya bölerek, $F$'e dönüştürebileceğimizi daha önceden biliyorduk.

$$F = \frac{KO_{\text{dogrusal}}}{KO_{\text{hata}}}$$
$$= \frac{,002}{2,263} = 0,0009$$

Bu, 1 ve 55 serbestlik derecesindeki $F$ değeridir. EK $F$'te bunu $F_{,05}(1,55) = 4,015$ olarak bulabilirsiniz. Çünkü doğrusal bileşenler için $F$ (0,0009), 4,015'ten çok küçüktür. $H_0$ korunacaktır ve ortalamalarımızda manidar doğrusal eğilim olmadığı sonucuna varılacaktır. Başka bir ifadeyle, saldırganlığın dikkat dağınıklığındaki seviyenin artmasıyla doğrusal olarak değişmediği sonucuna ulaşılacaktır. Burada, bulduğumuz manidar olmayan $F$, test ettiğimiz eğilim bileşeninin 0'dan manidar olarak farklı olmadığını gösterir.[13]

Verimiz doğrusal bir eğilim göstermemesine rağmen, kuadratik eğilim gösteriyor gibi görünmektedir - çizgi önce yukarıya çıkıyor, sonra aşağıya iniyor. Şekil 12.2c'deki grafikte görüldüğü gibi hem doğrusal hem de kuadratik eğilim gösteren veri bulmak da yaygındır.

Bir sonraki adım, doğrusal bileşen uyum gösterdikten sonra, kalan artık varyansın şimdiye kadar bildiğimiz hata varyansından daha büyük olup olmadığını sormaktır. Eğer $KT_{\text{doğrusal}}$ neredeyse bütün $KT_{\text{işlem}}$'den oluşuyorsa, yüksek dereceden terimleri açıklamak için orada çok az bir şey kalacak veya hiçbir şey kalmayacaktır. Diğer taraftan eğer $KT_{\text{doğrusal}}$, $KT_{\text{işlem}}$'in nispeten küçük bir parçasıysa, sonra yüksek dereceden bileşenlere bakmak anlamlı olacaktır. Önceki hesaplamalarımızdan elde ettiğimiz,

$$KT_{\text{artık}} = KT_{\text{işlem}} - KT_{\text{doğrusal}}$$
$$= 62,460 - 0,002$$
$$= 62,458$$
$$sd_{\text{artık}} = sd_{\text{işlem}} - sd_{\text{doğrusal}}$$
$$= 4 - 1$$
$$= 3$$
$$KO_{\text{artık}} = \frac{KT_{\text{artık}}}{sd_{\text{artık}}}$$
$$= \frac{62,458}{3}$$
$$= 20,819$$

13 Yakın zamanda birisinden bir mesaj aldım. Alkol etkisinin alkol dozu ile arttığı deney hipotezini çalışıyordu. Manidar olmayan bir genel $F$ bulmuştu, fakat doğrusal eğilim testini hesapladığında "yüksek bir manidarlık"-la karşılaşmıştı ve ne yapacağını soruyordu. Çünkü hipoteziyle doğrusal eğilimi test etti, oysa genel $F$ değil, tavsiyem sadece eğilim analizine güvenmesi şeklinde oldu.

$$F_{\text{artık}} = \frac{MS_{\text{artık}}}{MS_{\text{hata}}} = \frac{20,819}{2,263} = 9,20$$

olur. Çünkü artık için $F$, $F_{(,05;3;55)}$'ten büyüktür, yokluk hipotezi reddedilir ve doğrusal bileşen tarafından açıklanmayan manidar bir değişkenliğin kaldığı sonucuna ulaşılır. Kuadratik katsayıların, doğrusal katsayıların yerini alması dışında, kuadratik bileşen için olan hesaplamalar da temelde aynıdır.

$$\Psi_{Kuadratik} = (2)1,802 + (-1)0,675 + (-2)(-0,912) + (-1)0,544 + (2)1,889$$
$$= 7,989$$
$$KO_{Kuadratik} = \frac{n\Psi^2}{\sum b_j^2} = \frac{12(7,989^2)}{14} = 54,706$$
$$F = \frac{KO_{Kuadratik}}{KO_{hata}} = \frac{54,706}{2,263} = 24,17$$

Bu test açıkça manidardır, veri kümemizde kuadratik eğilim olduğunu göstermektedir. Bu bir sürpriz olarak görünmemelidir, çünkü grafikten açık olarak görülmektedir.

Bu noktada bir uyarı yapmak gerekir. Devam etmek cazip gelebilir, kubik ve kuadratik katsayıları uygulayabilirsiniz, bunu EK Polinomsal'da bulabilirsiniz. Ayrıca, bunu yaptığınızda, dört kareler toplamının ($KT_{\text{doğrusal}}$, ……, $KT_{\text{kuadratik}}$) olacağını ve hesaplanan gruplar arasındaki tüm kareler toplamının etkileneceğini gözleyebilirsiniz. Çok etkilenmeden önce, iki noktayı birleştiren doğruyu çizebildiğinizi gösterebilirseniz ne kadar gururlanacağınızı düşünün. Aynı düşünce burada geçerlidir. Veri kümesine bakmadan, $k$-1 derecedeki polinomun $k$ noktaya tam olarak uyduğunu başlamadan biliyorsunuz. Eldeki veri kümesini kuadratik bileşenlere uydurma işi, neden daha ötesine gitmekte istekli olmadığımın tek nedenidir. Üstelik, dördüncü dereceden polinoma uyum olsaydı ve bu ikinci dereceden bileşenin manidar olduğu bulunsaydı, bu sonuçlar hakkında ne söylemek gerekirdi? Doğrusal veya ikinci dereceden bileşen anlam kazanır, ancak bildiğim bir kuadratik bileşen herhangi bir teori ile açıklanamamıştır.

## Eşit Olmayan Aralıklar

Yukarıdaki bölümde bağımsız değişken düzeylerinin süreklilik boyunca eşit aralıklı olduğu varsayıldı. Eşit aralıklara sahip olmadığımızda da eğilim analizini yapmak mümkündür ve aritmetik aynıdır. Tek sorun eğilim katsayılarını elde etmeye çalıştığımızda oluşur çünkü, aralıklar eşit olmadıkça EK Polinomsal'dan katsayıları alamayız.

İkinci dereceden katsayıları hesaplamak çok zor değildir ve iyi bir açıklama Keppel'da (1973) bulunur. Yüksek dereceden polinomlar için hesaplamalar daha zahmetlidir ancak,

Robson'da (1959) sürece ilişkin açıklamalar vardır. Çoğu insan kendi analizlerini standart istatistiksel yazılımlarla yürütür ve bu yazılımlar genellikle eşit olmayan aralıklar probleminin üstesinden gelmektedir. El kitabının derinine dalmadan, yazılımın aralık probleminin üstesinden nasıl geleceğini belirlemek oldukça zordur.

Alıştırma 12.25 hem kubik hem de kuadratik bileşenleri içeren bir örnektir. Bu alıştırmayı yaparak eğilim analizi hakkında çok şey öğrenebilirsiniz.

## Anahtar Terimler

Karşılaştırma başına hata oranı (*PC*) (12.1)
Aile-boyu hata oranı (*AHO*) (12.1)
Çok amaçlı yokluk hipotezi (12.1)
Önsel karşılaştırmalar (12.1)
Çoklu karşılaştırmalar (12.1)
Karşılaştırmalar (12.3)
Doğrusal kombinasyon (12.3)
Doğrusal karşılaştırma (12.3)
Parçalara ayırma (12.3)
Dik doğrusal karşılaştırma (12.3)
Dunn testi (12.3)
Bonferroni *t* (12.3)
Bonferroni eşitsizliği (12.3)
Dunn-Sidak testi (12.3)
Fisher'ın en küçük manidar fark testi (LSD) (12.6)
Studentlaştırılmış ranj istatistiği (*q*) (12.6)
Tukey Testi (12.6)
Tukey HSD testi (12.6)
Newman–Keuls testi (12.6)
Ryan prosedürü (REGWQ) (12.6)
Sheffé testi (12.6)
Dunnet testi (12.6)
Yanlış Buluş Oranı (YBO) (12.6)
Benjamini ve Hochberg Doğrusal Adımlama (DSU) prosedürü (12.6)
Eğilim (12.9)
Kuadratik fonksiyon (12.9)
Polinomsal eğilim katsayıları (12.9)

## Alıştırmalar

12.1 Aşağıda verilen veri kümesinin yiyecek veya su yoksunluğunun bir öğrenme görevi davranışı üzerindeki etkilerini temsil ettiğini varsayalım. İşlem 1, hayvanların isteklerine göre yiyecek ve su aldığı (1), işlem 2 hayvanların günde iki kez yiyecek ve su aldığı (2) kontrol koşullarını temsil etmektedir. İşlem 3'te hayvanlar yiyecekten yoksun bırakılmıştır, işlem 4'te sudan yoksun bırakılmıştır ve işlem 5'te hem yiyecekten hem de sudan yoksun bırakılmıştır. Bağımlı değişken, önceden belirlenmiş ölçüte ulaşmadaki deneme sayısıdır. Deneye başlamadan önce, birleştirilmiş kontrol grupları (uygulama 1 ve 2) ile birleştirilmiş deney gruplarını, kontrol gruplarını birbirleriyle, tek olarak yoksun bırakılan (yiyecek veya sudan biri) uygulamalarla çift yoksun (hem yiyecek hem su) bırakılan uygulamayı ve tek olarak yoksun bırakılan uygulamaları birbirleriyle karşılaştırmak istediğimize karar verdiğimizi düşünelim.

| İsteğe Göre Kontrol | Günde İki Kez Kontrol | Yiyecek Yoksunluğu | Su Yoksunluğu | Yiyecek ve Su Yoksunluğu |
|---|---|---|---|---|
| 18 | 20 | 6 | 15 | 12 |
| 20 | 25 | 9 | 10 | 11 |
| 21 | 23 | 8 | 9 | 8 |
| 16 | 27 | 6 | 12 | 13 |
| 15 | 25 | 11 | 14 | 11 |
| 90 | 120 | 40 | 60 | 55 |

a. Doğrusal karşılaştırmaları kullanarak veriyi analiz ediniz.

b. Doğrusal karşılaştırmaların dik olduğunu gösteriniz.

c. Doğrusal karşılaştırmaların toplamı için kareler toplamını $KT_{işlem}$ gösteriniz.

© Cengage Learning 2013

12.2 Bölüm 11'de ilk örnekteki veriyi kullanarak, Sayma ve Kafiyelemeye karşı Sıfat ve İmgeleme için doğrusal karşılaştırmaları hesaplayınız ve sonra Sıfata karşı İmgelemeyi karşılaştırınız. Her doğrusal karşılaştırma için $\alpha$ = ,05 alınız. (*Bu ve sonraki alıştırmaların Bölüm 11'deki alıştırmalara gönderme yaptığını unutmayınız.*)

12.3 Alıştırma 12.2'deki karşılaştırma başına düşen ve aile-boyu hata oranları ne olurdu? (*İpucu*: Doğrusal karşılaştırmalar dik mi?)

12.4 Alıştırma 11.2'deki iki grup üzerinden doğrusal karşılaştırma için *F*'yi hesaplayınız. Bu zaman kaybı mıdır? Neden?

12.5 Alıştırma 11.2'deki iki grup için Student'laştırılmış aralığı hesaplayınız ve bunun $t\sqrt{2}$ 'ye eşit olduğunu gösteriniz. (Alıştırma 11.2b'den alınan *t*'dir.)

12.6 Alıştırma 11.3'te, aşağıdaki doğrusal farklılıklar için *F* değerlerini hesaplayınız. Bölüm 13'te kullanılmak üzere sonuçları kaydediniz.

a. 1 ve 2'ye karşı 3 ve 4

b. 1 ve 3'e karşı 2 ve 4

c. 1 ve 4'e karşı 2 ve 3

d. (a), (b) ve (c)'deki doğrusal karşılaştırma hangi sorulara işaret etmektedir?

12.7 Alıştırma 12.2'de verilen doğrusalları kullanarak, Alıştırma 11.1'deki veri üzerinde Bonferroni *t* testini uygulayınız. Maksimum *AHO*'yu ,05 olarak ayarlayınız.

12.8 Tablo 11.2'de (sayfa 332) verilen örneğe Tukey testini uygulayınız ve sonuçları yorumlayınız.

12.9 Alıştırma 12.8'de belirtilen veri üzerinde belirli doğrusal karşılaştırmaları yapmaya neden daha fazla ilgi duyuyor olabilirsiniz?

12.10 Games ve Howell (1976) yaklaşımına Tukey'nin HSD prosedürünü eşit olmayan örneklemler için aşağıdaki veriye uygulayınız.

| Grup | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\overline{X}_j$ | 10 | 18 | 19 | 21 | 29 |
| $n_j$ | 8 | 5 | 8 | 7 | 9 |
| $s_j^2$ | 7,4 | 8,9 | 8,6 | 7,2 | 9,3 |

© Cengage Learning 2013

12.11 Alıştırma 12.10'daki veri üzerinde grup 1, 2 ve 3 (birleştirilmiş) ile grup 4 ve 5'i (birleştirilmiş) karşılaştırmak için Scheffé testini kullanınız. Daha sonra grup 1 ile grup 2, 3 ve 4'ü (birleştirilmiş) karşılaştırınız.

12.12 Tablo 11.6'daki (sayfa 348) log dönüşümlü THC veri kümesine Tukey yöntemini uygulayınız. Bu yöntem için maksimum *AHO* ne olmalıdır?

12.13 Tablo 11.6'daki log dönüşümlü veriye Dunnett testini SPSS kullanarak uygulayınız.

12.14 Alıştırma 12.7'deki veri üzerinde Bonferroni *t* testini uygulamak için kullanılabilir bir Bonferroni komutuna sahip olmayan istatistiksel paket nasıl çalıştırılabilir?

12.15 Bonferroni çok aşamalı testi bazen değiştirilmiş ardışık reddedici yöntem olarak adlandırılır. Neden?

12.16 Tablo 11.6'daki Conti ve Musty (1986) log dönüşümlü veriyi doğrusal ve kuadratik eğilim bileşenlerine uydurunuz. THC'nin 0 $\mu$g'ı kontrol koşulu olarak alınmıştır. Bu örnek için, bütün gruplarda 10 bireyin olduğunu varsayınız. (Sonuçları hissedilir bir şekilde değiştirmeden 0,5 $\mu$g grubuna 2,56 ve 1 $\mu$g grubuna 2,36 ekleyebilirsiniz.) Doğrusal katsayılar (bağımsız değişken üzerindeki eşit olmayan aralıkla hesaplanan) -0,72, -0,62, -0,22, 0,28 ve 1,28'dir. İkinci dereceden katsayılar 0,389, 0,199, -0,362, -0,612, 0,387'dir.

Eğer varsa, SPSS ONEWAY'i kullanarak cevaplarınızı doğrulayınız.

12.17 Alıştırma 12.17 için hesaplanan sonuçların kısa bir raporunu yazınız.

12.18 Bölüm 11'deki (sayfa 366) alıştırmada açıklanan Introini-Collison ve McGaugh'ın (1986) verisine herhangi bir istatistiksel paket kullanarak Tukey ve Scheffé testlerini uygulayınız. Bu analizleri kitabın Web sitesinde bulunan Epineq.dat ve Epinuneq.dat'ın her ikisi için yapınız. Aralık değişkenlerinin düzeylerini birleştirmeyiniz.

12.19 Alıştırma 12.18'de üç aralık boyunca veri birleştirilse de birleştirilmese de bu çok fark yaratmazdı. Hangi koşullar altında büyük bir fark olmasını beklersiniz?

12.20 Epineq.dat verisini kullanarak, üç ilâç dozu üzerinde hem doğrusal hem de ikinci derece eğilim testlerini uygulayınız. Bunu her bir üç aralık için sırayla yapınız. (*İpucu*: Doğrusal katsayılar, -0,597110, -0,183726, 0,780836 ve kuadratik katsayılar, 0,556890, -0,795557, 0,238667'dir.)

12.21 Alıştırma 12.20'deki sonuçları yorumlayınız.

12.22 Stone, Rudd, Ragozzino ve Gold (1992) bellekte glikozun oynadığı rolü araştırmışlardır. Sabah 6'da başlayarak 12 saatlik gece/gündüz döngüsüyle fareleri büyütmüşlerdir. Eğitim sırasında fareler, yarısı aydınlık olan deney kutusuna yerleştirilmiş ve karanlık yarıya geçtiklerinde farelere ayaklarından şok verilmiştir. Fareler hızlı bir şekilde aydınlık yarıda kalmayı öğrenmişlerdir. Daha sonra gece/gündüz döngüsü tüm fareler için ilerletilmiştir, bu durum orijinal eğitimde belleğe engel olmak olarak bilinir. Üç gün sonra, 0, 1, 10, 100, 250 ve 500 mg/kg sakaroz farelere enjekte edildikten 30 dakika sonra fareler yeniden teste alınmıştır. Amaç sakarozun değişen gündüz döngüsünün yıkıcı etkilerini azaltıp azaltmadığını ve farklı dozların farklı etkilere sahip olup olmadığını görmektir. Stone arkadaşlarının sonuçlarına kabaca benzeyen üretilmiş veri aşağıda verilmiştir ve buradaki bağımlı değişken, karanlık bölmeye girmek için gecikme süresidir.

| Glukoz Düzeyi mg/kg | | | | | |
|---|---|---|---|---|---|
| **0** | **1** | **10** | **100** | **250** | **500** |
| 295 | 129 | 393 | 653 | 379 | 521 |
| 287 | 248 | 484 | 732 | 530 | 241 |
| 91 | 350 | 308 | 570 | 364 | 162 |
| 260 | 278 | 112 | 434 | 385 | 197 |
| 193 | 150 | 132 | 690 | 355 | 156 |
| 52 | 195 | 414 | 679 | 558 | 384 |

© Cengage Learning 2013

a. *X* değerlerini 1, 2, 3, 4, 5, 6 alarak ve bütün mevcut dozajları kullanarak bu verinin grafiğini çiziniz.
b. Bağımsız değişken olarak mevcut dozajları alıp eğer uygunsa, SPSS One-way'i kullanarak eğilim analizini çalıştırınız.
c. Bağımsız değişken olarak 1, 2, 3, 4, 5, 6 kodlamasını yaparak (b) seçeneğinde yapılanı tekrarlayınız.
d. Sonuçlarınızı yorumlayınız. Sınav için bütün gece ayakta kalan öğrencilere bu sonuçların nasıl olduğunu söylersiniz?
e. Sizce Stone ve arkadaşları neden bir kodlama sistemini diğerine tercih etmiş olabilirler?

12.23 Alıştırma 12.1'deki veriyi kullanarak, o soruda açıklanan ilk karşılaştırma (doğrusal) için güven aralığını bulunuz. Cevabınızı yorumlayınız. (Eğer SPSS kullanacaksanız Compare Means/One-Way ANOVA kullanınız, bu, katsayıları belirlemenizi sağlar.)

12.24 Alıştırma 12.1'deki veriyi kullanarak, o soruyla çalıştırdığınız bütün karşılaştırmalar için etki büyüklüklerini hesaplayınız. Bu etki büyüklüklerini nasıl yorumlarsınız?

Bunlar neden standartlaştırılmış etki büyüklüğü şeklinde isimlendirilmiştir ve standartlaştırılmış olmayan etki büyüklüğü nedir?

12.25 Alıştırma 11.27'deki veriyi kullanarak, Davey ve arkadaşlarının (2003) çalışmasındaki manidar sonuçlara ulaşmak için uygun test(ler)i uygulayınız.

12.26 Alıştırma 11.8'deki Foa ve arkadaşlarının tecavüz mağdurları için terapi konusundaki çalışmasını dikkate alalım. Ham veri Ex12.26.dat Web sitesinden bulunabilir. Bu veriye, Benjamini ve Hochberg LSU prosedürünü uygulayınız.

## Tartışma Soruları

12.27 Öğrenciler sık sık önsel ve çoklu karşılaştırma testlerinin aileboyu hata oranlarının neden farklı olduğunu görmede zorlanırlar. Başkalarına bu farkı anlatacak bir örnek (istatistikten olması şart değil) hazırlayınız.

12.28 Araştırma literatüründe en az beş farklı koşulun kullanıldığı bir çalışma örneği bulunuz ve bu deneyden gelmiş olabilecek bir veri kümesi oluşturunuz. Tartışılan tekniklerin birkaçını uygulayınız, bunların kullanımlarını doğrulayınız ve sonuçları yorumlayınız. (Bu gibi bir örneğin dışında veri kümesi için asla birkaç farklı tekniği uygulamayın.) *İpucu*: Herhangi sayı kümesini alarak verilen ortalama ve varyansla veri üretebilirsiniz [onları en azından tek modlu ve simetrik yapabilirsiniz], onları standartlaştıtrabilirsiniz, istenilen standart sapmayla standart puanları çarpabilirsiniz ve sonuçlara istenilen ortalamayı ekleyebilirsiniz. Bunu ayrı ayrı her bir grup için yapınız ve kendi verinizi oluşturunuz.



*Bölüm 13*

# Faktöriyel Varyans Analizi

## Hedefler

İki veya daha çok bağımsız değişken durumu için varyans analizinin tartışılması. Bu bölüm aynı zamanda iç içe desenleri de kapsamaktadır.

## Konular


**13.1** Eysenck'ın Çalışmasının Genişletilmesi
**13.2** Yapısal Model ve Beklenen Kareler Ortalamaları
**13.3** Etkileşimler
**13.4** Basit Etkiler
**13.5** Varyans Analizinin Sigara İçmenin Etkilerine Uygulanması
**13.6** Ortalamalar Arası Karşılaştırmalar
**13.7** Faktöriyel Deneyler İçin Güç Analizi
**13.8** Alternatif Deneysel Desenler
**13.9** İlişki ve Etki Büyüklüğü Ölçümleri
**13.10** Sonuçların Rapor Edilmesi
**13.11** Eşit Olmayan Örneklem Büyüklükleri
**13.12** Yüksek Mertebeden Faktöriyel Desenler
**13.13** Bir Bilgisayar Örneği

ÖNCEKİ İKİ BÖLÜMDE, sadece bir bağımsız değişkenimiz varken tek yönlü varyans analizi üzerinde çalıştık. Bu bölümde, varyans analizini, iki veya daha çok değişken içeren deneysel desenlerdeki işlemlere genişleteceğiz. Basitlik amacıyla önce iki veya üç değişken içeren deneyleri ele alacağız, buna rağmen tartışılan teknikler daha karmaşık desenlere genişletilebilir.

Bölüm 11'deki alıştırmada Eysenck (1974) tarafından yapılan, katılımcıların birkaç farklı koşuldan birine maruz bırakılarak kelime listelerini hatırlamalarının istendiği çalışma ile ilgilenmiştik. Alıştırma 11.1'de, beş grup, öğrenilmesi istenilen kelime listesinin gerektirdiği zihinsel işlemlerin çokluğu bakımından farklılaşmaktaydı. Bu farklılık, kelimenin zihinsel görüntüsünü oluşturmak için gereken, o kelimenin içindeki harf sayısının basitçe sayılmasından kaynaklanıyordu. Daha sonra hatırlamak için kelimeleri basitçe ezberlenmesi söylenen bir grup daha vardı. Hatırlamanın, öncelikle materyalin veriliş düzeyi ile ilişkili olup olmadığının belirlenmesiyle ilgilendik. Eysenck'ın çalışması aslında daha karmaşıktı. O, kavramları işleme düzeylerinin, yaşlı ve genç katılımcılar arasındaki hatırlanma farklılıklarını açıklayabilir olup olmadığını merak ediyordu. Özellikle koşullar daha fazla işlem gerektirdiğinde, yaşlı katılımcılar bilgiyi derin bir şekilde işleyemiyorlarsa, onların genç katılımcıların hatırlayacağından daha az öğe hatırlamaları beklenmektedir. Bu çalışma **faktör** olarak tanımlayacağımız iki bağımsız değişken içeriyor: Yaş ve Hatırlama durumu (bundan sonra basitçe Durum denilecektir). Deney böylece **iki yönlü faktöriyel desenin** bir örneği olarak gösterilebilir.

faktör

iki yönlü faktöriyel desen

Her faktörün her düzeyinin, diğer faktörlerin her düzeyi ile eşleştirildiği deneysel desene **faktöriyel desen** denir. Diğer bir deyişle, faktöriyel desen bağımsız değişkenlerin düzeylerinin tüm *kombinasyonlarını* içerir. Bu bölümde ele alınan faktöriyel desende, her işlem kombinasyonunda çalışılan farklı katılımcıların durumları dikkate alınacaktır. Örneğin, bizim örneğimizde bir grup genç katılımcı Sayma durumunda, başka bir grup genç katılımcı Kafiye durumunda vb. Bizim iki faktörümüzün (5 hatırlama durumu x 2 Yaş) 10 kombinasyonu olduğundan 10 farklı katılımcı grubu elde edeceğiz. Araştırma planı, birden fazla işlem kombinasyonu için *aynı* katılımcıları içeriyorsa, bu deseni **tekrarlı-ölçüm deseni** olarak adlandıracağız, bunu Bölüm 14'te ayrıntılı olarak tartışacağız.

faktöriyel desen

tekrarlı-ölçüm deseni

Faktöriyel desenlerin tek yönlü desenlere göre önemli avantajları vardır. Bunlardan birincisi faktöriyel desenlerin, sonuçları daha geniş genellemelere izin vermesidir. Bir an için Eysenck'ın çalışmasını düşünelim. Eğer Alıştırma 11.1'deki gibi sadece yaşlı katılımcılar üzerinde beş Durum için tek yönlü analiz yapmış olsaydık, sonuçlarımız sadece yaşlı katılımcılar için geçerli olacaktı. Genç ve yaşlı katılımcılar için faktöriyel desen kullandığımızda, farklı durumlar altında genç ve yaşlı katılımcılar arasındaki farklılığın hangi durumlarda daha fazla olduğunu saptama şansımız vardır. Aynı zamanda, tüm görevlerde yaşlara bağlı hatırlama farklılıkları olup olmadığını veya belirli durumlarda genç (veya yaşlı) katılımcıların sadece belirli görevlerde üstün olup olmadığını saptayabilme şansımız da vardır. Böylece faktöriyel desenler, sonuçların daha geniş bir şekilde yorumlanmasına ve bağımsız değişkenlerin her birinin ayrı ayrı sonuçları için anlamlı yorumlar yapılabilmesine imkân sağlar.

Faktöriyel desenlerin ikinci bir önemli özelliği, değişkenlerin **etkileşim**ini görmeye imkân vermesidir. Hatırlama etkisi durumunun yaştan bağımsız olup olmadığını veya Durum ve Yaş arasında bir etkileşim olup olmadığını sorabiliriz. Örneğin, beş hatırlama durumunda genç katılımcılar yaşlı katılımcılara göre daha fazla (veya daha az) fark gösteriyorsa bir etkileşim var diyebiliriz. Etkileşimin etkileri genelde elde ettiğimiz en ilginç sonuçlar arasında yer alır.

etkileşim

Faktöriyel desenlerin üçüncü avantajı ekonomik olmasıdır. Bir değişkenin diğer değişkenin düzeyleri üzerinden ortalamasını alacağımız için, aynı güç derecesinde, iki değişkenli faktöriyel desen, iki tane tek yönlü desene göre daha az katılımcı gerektirir. Aslında hiç çaba sarf etmeden getiri elde edebiliriz. Yaş ve Durum etkileşimi için bir neden olmadığını varsayalım. Beş durumun her biri için, her durum için 10 yaşlı ve 10 genç katılımcıdan

olmak üzere 20 puanımız olacaktır. Eğer genç katılımcılar için tek yönlü analiz ve yaşlı katılımcılar için başka bir tek yönlü analiz yapmış olsaydık, her deneyde Durumdan kaynaklanan farklılıkları aynı güçte saptayabilmek için iki kat katılımcıya ihtiyaç duyardık. Bu da iki deney ve her deney koşulu için 20 katılımcıyı gerektirirdi.

Faktöriyel desenler dikkate alınan faktör sayıları ile adlandırılır. İki bağımsız değişken veya faktör içeren faktöriyel desen iki yönlü faktöriyel, üç faktör içeren desen ise üç yönlü faktöriyel desen olarak adlandırılır. Desenleri adlandırmanın alternatif bir metodu da, her faktör düzeyinin sayısı açısından adlandırmaktır. Eysenck'ın çalışması, Yaş faktöründen iki düzey ve Durum faktöründen de beş düzey içerir, **2 x 5 faktöriyel** gibi. Üç faktörlü bir çalışmada, üçer düzey içeren iki faktör ve dört düzey içeren bir faktör varsa bu 3 x 3 x 4 faktöriyel olarak adlandırılır. "İki yönlü" ve "2 x 5" terimlerinin ikisi de aynı deseni tanımlamanın ortak yollarıdır ve bu kitap boyunca ikisi de kullanılacaktır.

**2 x 5 faktöriyel**

İzleyen kesimde öncelikli olarak daha çok iki yönlü analiz ile ilgilenecek olsak da, ilgi sadece bununla sınırlı kalmayacaktır. İki yönlü analizi anladığımız zaman yüksek mertebeden analizler otomatik olarak daha anlaşılır hale gelmeye başlayacak ve bununla ilişkili problemler iki faktör açısından daha kolayca tartışılabilecektir. Bu bölümün çoğunda tartışmalarımızı Bölüm 11'de değindiğimiz sabit modellerle –tesadüfî desenin tersi- sınırlayacağız. Sabit faktörün, deneycinin sadece ilgilendiği düzeylerine göre özellikle seçilen bir model olduğunu hatırlamak gerekir. Tesadüfî model, düzeyleri tesadüfî süreçlerle belirlenen ve ilgi odakları faktörün muhtemel tüm düzeyleri olan modelleri içerir. Yaş veya "terapi türü", sabit modeller için iyi örneklerdir. Eğer isimleri ve fiilleri hatırlama arasındaki farkı çalışmak istiyorsak, kullandığımız belirli fiiller tesadüfî değişkeni temsil eder, çünkü ilgilendiğimiz fiilleri tüm fiillere genellemiş oluruz.

## Notasyon

A ve B değişkenlerini içeren iki değişkenli hipotetik bir deneyi göz önüne alalım. Bu türde bir desen Tablo 13.1'de görülmektedir. *A*'nın düzeylerinin sayısı *a* ile, *B*'nin düzeylerinin sayısı da *b* ile ifade edilmiştir. *A*'nın bir düzeyi ile *B*'nin bir düzeyinin herhangi bir kombi-

**Tablo 13.1** Faktöriyel desenin gösterimi

| | $B_1$ | $B_2$ | ... | $B_b$ | |
|---|---|---|---|---|---|
| $A_1$ | $X_{111}$<br>$X_{112}$<br>...<br>$X_{11n}$<br>$\overline{X}_{11}$ | $X_{121}$<br>$X_{122}$<br>...<br>$X_{12n}$<br>$\overline{X}_{12}$ | ... | $X_{1b1}$<br>$X_{1b2}$<br>...<br>$X_{1bn}$<br>$\overline{X}_{1b}$ | $\overline{X}_{1.}$ |
| $A_2$ | $X_{211}$<br>$X_{212}$<br>...<br>$X_{21n}$<br>$\overline{X}_{21}$ | $X_{221}$<br>$X_{222}$<br>...<br>$X_{22n}$<br>$\overline{X}_{22}$ | ... | $X_{2b1}$<br>$X_{2b2}$<br>...<br>$X_{2bn}$<br>$\overline{X}_{2b}$ | $\overline{X}_{2.}$ |
| | ... | ... | ... | ... | ... |
| $A_a$ | $X_{a11}$<br>$X_{a12}$<br>...<br>$X_{a1n}$<br>$\overline{X}_{a1}$ | $X_{a21}$<br>$X_{a22}$<br>...<br>$X_{a2n}$<br>$\overline{X}_{a2}$ | | $X_{ab1}$<br>$X_{ab2}$<br>...<br>$X_{abn}$<br>$\overline{X}_{ab}$ | $\overline{X}_{a.}$ |
| | $\overline{X}_{.1}$ | $\overline{X}_{.2}$ | ... | $\overline{X}_{.b}$ | $\overline{X}_{..}$ |

© Cengage Learning 2013

**hücre**

nasyonu **hücre** terimi ile adlandırılır ve hücre başına düşen gözlem sayısı $n$ veya daha açık olarak $n_{ij}$ ile gösterilir. Toplam gözlem sayısı $N = \sum n_{ij} = abn$'dir. Herhangi bir karışıklık olmaması için tek bir gözlem ($X$), üç alt indisle, $X_{ijk}$, ile belirtilir. $i$ satır numarasını ($A$'nın düzeyi), $j$ sütun numarasını ($B$'nin düzeyi) ve $k$ ise $ij$ hücresindeki $k$'yıncı gözlemi belirtir. Böylece, $X_{234}$, ikinci satır ve üçüncü sütunnun kesiştiği hücredeki dört numaralı katılımcıyı gösterir. $A$'nın düzeylerindeki kişilerin ortalaması $\bar{X}_A$ veya $\bar{X}_i$, $B$'nin düzeyleri için $\bar{X}_B$ veya $\bar{X}_j$ ile belirtilir. Hücre ortalamaları $\bar{X}_{ij}$ ve genel ortalama da $\bar{X}_{..}$ ile gösterir. Gereksiz alt indisler genellikle karışıklık kaynağıdır ve bu nedenle mümkün olduğu durumlarda indis kullanılmayacaktır.

Bu gösterimler varyans analizi tartışmaları boyunca kullanılacaktır. Mevcut sistemin en büyük avantajı, daha karışık desenler için genelleştirilmiş olmasıdır. Yani, katılımcılar günün üç farklı zamanını hatırladıysa, $\bar{X}_{\text{Zaman 1}}$'in ne anlama geldiği açık olarak anlaşılmalıdır.

## 13.1 Eysenk'ın Çalışmasının Genişletilmesi

Daha önce değindiğimiz gibi, Eysenck Yaş ve Hatırlama Durumu değişkenleri ile bir çalışma yapmıştı. Bu çalışma, 18-30 ve 55-65 yaş aralıklarında 50'şer katılımcıyı içermekteydi. Tablo 13.2'deki veri kümesi Eysenck'ın çalışmasında rapor ettiği aynı ortalamalar ve standart sapmalar ele alınarak hazırlandı. Tablo, standart varyans analizi için yapılan tüm hesaplamaları içermektedir ve bunların her biri tek tek tartışılacaktır. Analize başlamadan önce veri kümesinin yaklaşık olarak normal dağılım gösterdiğinin ve eşit varyanslara sahip olduğunun kabul edilmesi önemlidir. Kişilere ait veri kümeleri yapay olduğu için kutu grafikleri tabloda verilmemiştir ama gerçek veri kümelerinde hesaplanması için çaba gösterilmesi iyi olur. Hücrelerden ve marjinal ortalamalardan hareketle hatırlama daha çok işlem gerektiren durumlarda artmaktadır ve genç katılımcıların yaşlı katılımcılara göre daha çok madde hatırladığı görülmektedir. Dikkat edilirse genç ve yaşlı katılımcılar arasındaki farkın verilen görevlere bağlı olduğu görünmekte, görevin işlem boyutu ne kadar çok ise fark da o kadar büyük olmaktadır. Analizin kendisini göz önüne aldıktan sonra, bu sonuçlar hakkında daha çok şey söyleyeceğiz.

İleride oluşabilecek karışıklığı önlemek için iki önemli terimi ele almak faydalı olacaktır. Daha önce belirttiğimiz gibi bu deneyde yaş ve durum olmak üzere iki faktör vardır. Belirli durumlar *göz ardı edildiğinde*, yaşlı ve genç katılımcıların ortalamaları arasındaki farka bakarsak, yaş değişkenin **ana etki**si denilen kavramla karşı karşıya kalırız. Aynı şekilde katılımcıların yaşlarını göz ardı ederek beş durum ortalamaları arasındaki farkı ele alırsak, durum değişkeninin temel etkisinden söz edebiliriz.

**ana etki**

Veriyi incelemenin alternatif bir yönteme, sadece Sayma görevindeki yaşlı ve genç katılımcıların ortalamalarının karşılaştırılması örnek olabilir. Yaşlı ve geç katılımcıların Amaçlı görevdeki ortalamalarını da karşılaştırabiliriz. Son olarak sadece yaşlı katılımcıların beş durumdaki ortalamalarını da karşılaştırabiliriz. Bu üç örneğin her birinde görüldüğü gibi, bir faktörün, diğer faktörün tek bir düzeyi üzerindeki etkisini incelemiş oluruz. Bu durumda **basit etki** başka bir deyişle **şartlı etki** kavramlarıyla karşı karşıya kalırız, bu da bir faktörün diğer faktörün tek bir düzeyindeki etkisi anlamına gelir. Yukarıda belirtildiği gibi, basit etki sıkça *durumsal etki*dir, diğer değişkenin bir düzeyi üzerinedir. Diğer taraftan ana etkide, bir faktör diğer faktörü göz ardı etmektedir. Eğer daha fazla işlem içeren görevler daha iyi hatırlamaya yol açtığı şeklinde bir ifade kullanıyorsak, burada ana etkiden bahsedilmektedir. Eğer genç katılımcıların daha fazla işlem gerektiren görevlerde daha iyi hatırlama gösterdiğini söylersek bu bir basit etkidir. Belirtildiği gibi basit etkiler diğer değişkenin düzeyi üzerine durumsal olarak etki eder. Basit etki ve kısaca hesaplanması üzerinde daha sonra konuşacağız, şimdilik terminolojiyi öğrenmemiz yeterlidir.

**basit etki**

**şartlı etki**

**Tablo 13.2** Eyesenk (1974) örneği için veri ve hesaplamalar

**(a) Veri:**

| | Hatırlama Durumları | | | | | Ortalama$_{i.}$ |
|---|---|---|---|---|---|---|
| | Sayma | Kafiyeleme | Sıfat | Mecazî | Amaçlı | |
| **Yaşlı** | 9 | 7 | 11 | 12 | 10 | |
| | 8 | 9 | 13 | 11 | 19 | |
| | 6 | 6 | 8 | 16 | 14 | |
| | 8 | 6 | 6 | 11 | 5 | |
| | 10 | 6 | 14 | 9 | 10 | |
| | 4 | 11 | 11 | 23 | 11 | |
| | 6 | 6 | 13 | 12 | 14 | |
| | 5 | 3 | 13 | 10 | 15 | |
| | 7 | 8 | 10 | 19 | 11 | |
| | 7 | 7 | 11 | 11 | 11 | |
| Ortalama$_{1j}$ | 7,0 | 6,9 | 11,0 | 13,4 | 12,0 | 10,06 |
| **Genç** | 8 | 10 | 14 | 20 | 21 | |
| | 6 | 7 | 11 | 16 | 19 | |
| | 4 | 8 | 18 | 16 | 17 | |
| | 6 | 10 | 14 | 15 | 15 | |
| | 7 | 4 | 13 | 18 | 22 | |
| | 6 | 7 | 22 | 16 | 16 | |
| | 5 | 10 | 17 | 20 | 22 | |
| | 7 | 6 | 16 | 22 | 22 | |
| | 9 | 7 | 12 | 14 | 18 | |
| | 7 | 7 | 11 | 19 | 21 | |
| Ortalama$_{2j}$ | 6,5 | 7,6 | 14,8 | 17,6 | 19,3 | 13,16 |
| **Ortalama**$_{.j}$ | 6,75 | 7,25 | 12,9 | 15,5 | 15,65 | 11,61 |

**(b) Hesaplamalar:**

$$\begin{aligned} KT_{\text{toplam}} &= \sum (X - \overline{X}..)^2 \\ &= (9 - 11{,}61)^2 + (8 - 11{,}61)^2 + \cdots + (21 - 11{,}61)^2 \\ &= 2667{,}79 \end{aligned}$$

$$\begin{aligned} KT_A &= nc \sum (\overline{X}_{i\cdot} - \overline{X}..)^2 \\ &= 10 \times 5[(10{,}06 - 11{,}61)^2 + (13{,}16 - 11{,}61)^2] \\ &= 240{,}25 \end{aligned}$$

$$\begin{aligned} KT_C &= na \sum (\overline{X}_{\cdot j} - \overline{X}..)^2 \\ &= 10 \times 2[(6{,}75 - 11{,}61)^2 + (7{,}25 - 11{,}61)^2 + \cdots + (15{,}65 - 11{,}61)^2] \\ &= 1514{,}94 \end{aligned}$$

$$\begin{aligned} KT_{\text{hücreler}} &= n \sum (\overline{X}_{ij} - \overline{X}..)^2 \\ &= 10[(7{,}0 - 11{,}61)^2 + (6{,}9 - 11{,}61)^2 + \cdots + (19{,}3 - 11{,}61)^2] \\ &= 1945{,}49 \end{aligned}$$

$$KT_{AC} = KT_{\text{hücreler}} - KT_A - KT_C = 1945{,}49 - 240{,}25 - 1514{,}94 = 190{,}30$$

$$KT_{\text{hata}} = KT_{\text{toplam}} - KT_{\text{hücreler}} = 2667{,}79 - 1945{,}49 = 722{,}30$$

*(devam edecek)*

© Cengage Learning 2013

**Tablo 13.2** *(Devamı)*

**(c) Özet tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* (Yaş) | 1 | 240,25 | 240,250 | 29,94* |
| *C* (Durum) | 4 | 1514,94 | 378,735 | 47,19* |
| *AC* | 4 | 190,30 | 47,575 | 5,93* |
| Hata | 90 | 722,30 | 8,026 | |
| Toplam | 99 | 2667,79 | | |

* $p < ,05$

## Hesaplamalar

Kareler toplamının hesaplanması Tablo 13.2b'de görülmektedir. Bu hesaplamaların çoğu tek yönlü varyans analizinde yaptıklarımıza benzer olduğundan aşina olmanız beklenir. Örneğin, $KT_{\text{toplam}}$, Bölüm 11'deki yöntemle aynı şekilde hesaplanır. Gözlemlerin genel ortalamadan sapmalarının karelerinin toplamını alırız.

Durum faktörünün olmadığını varsayarsak burada Yaş faktörü için elde ettiğimiz kareler toplamı, tek yönlü varyans analizdeki $KT_{\text{işlem}}$'den başka bir şey değildir. Diğer bir deyişle, yaş faktörünün ortalamalarının genel ortalamadan olan sapmalarının karelerinin toplamını alırız ve *nc* ile çarparız. Çarpım yaparken *nc* kullanmamızın sebebi her yaşta, her *c* düzeyinde *n* katılımcının olmasıdır. (Formülde bunun çarpan olarak hatırlanmasına ihtiyaç yoktur. Sadece ilgili ortalamanın hesaplanmasında dikkate alındığının bilinmesi yeterlidir.) $KT_{\text{C}}$'nin hesaplanması Yaş değişkeni göz ardı edilerek aynı şekilde yapılır.

$KT_{\text{toplam}}$ terimi hesaplanırken tanıdık olmayan $KT_{\text{hücre}}$ terimiyle karşılaşırız**.** Bu terim, tek bir **hücre**deki değişkenliği temsil eder ve aslında yapay bir terimdir, özet tabloda gözükmez. Diğer kare toplamları gibi hesaplanır. Hücre ortalamalarının genel ortalamadan sapmalarının kareleri toplamı alınarak ortalama başına düşen gözlem sayısı olan *n* ile çarpılır. Bu terimi neden istediğimiz açık olmayabilir, fakat Yaş ve Durum değişkenlerinin etkileşimini incelerken oldukça kullanışlı olacaktır. (Hesaplanmasını anlamak için 10 grup ve $KT_{\text{grup}}$ hesaplaması gerektiren bir çalışmayı düşünmek faydalı olacaktır.)

$KT_{\text{hücreler}}$, hücre ortalamalarının ne kadar farklılaştığının ölçüsüdür. İki hücre ortalaması örnekleme hatası haricinde üç farklı nedenle farklılaşabilir: (1) farklı yaş düzeylerinden (*A* yaş); (2) durumun farklı düzeylerinden (C durum) veya; (3) *A* ve *C* arasındaki etkileşimden. $KT_{\text{hücreler}}$'i bildiğimizden hücrelerin ne kadar farklılaştığının ölçüsünü bilebiliriz. Bize farklılığın ne kadarının yaş faktöründen kaynaklandığını söyler ve aynı şekilde bize farklılığın ne kadarının durum/koşul faktöründen kaynaklandığını belirtir. Yaş ve durum etkileşiminden kaynaklanan farklılık ($KT_{\text{AC}}$) da yaş veya durum ile açıklanamayan farklılıkları açıklarken karşımıza çıkabilir. Böylece $KT_{\text{hücreler}}$ üç bileşene ayrılabilir: -$KT_{\text{A}}$, $KT_{\text{C}}$ ve $KT_{\text{AC}}$. $KT_{\text{AC}}$'yi elde etmek için $KT_{\text{A}}$, $KT_{\text{C}}$'yi $KT_{\text{hücreler}}$'den çıkarırız, kalan $KT_{\text{AC}}$'yi verir. Örneğimiz için,

$$\begin{aligned} KT_{\text{AC}} &= KT_{\text{hücreler}} - KT_{\text{A}} - KT_{\text{C}} \\ &= 1945{,}49 - 240{,}25 - 1514{,}94 = 190{,}30 \end{aligned}$$

Geriye sadece hata nedeniyle oluşan kare toplamlarını hesaplamak kalır. Tek yönlü varyans analizinde olduğu gibi, bunu çıkarma işlemiyle elde ederiz. Toplam değişkenlik

$KT_{\text{toplam}}$ olarak gösterilir. Bu toplamın ne kadarının *A*, *C* ve *AC*'den geldiğini biliyoruz. Geriye kalan, hesaplanamayan değişkenlikler veya hatadır. Böylece,

$$KT_{\text{hata}} = KT_{\text{toplam}} - (KT_{\text{A}} + KT_{\text{C}} + KT_{\text{AC}})$$

ve $KT_{\text{A}} + KT_{\text{C}} + KT_{\text{AC}} = KT_{\text{hücreler}}$ olduğundan hata için aşağıdaki eşitlik yazılabilir:

$$KT_{\text{hata}} = KT_{\text{toplam}} - KT_{\text{hücreler}}$$

Bu bize, hata kareleri toplamını verir ve analiz için gereken tüm kare toplamlarını elde etmiş oluruz.

$KT_{\text{hata}}$'yı hesaplamanın, hata kareleri toplamının neyin göstergesi olduğunu daha açık ifade edecek daha doğrudan ama yorucu bir yolu vardır. $KT_{\text{hata}}$, her hücre içindeki değişimi temsil eder ve her hücredeki kare toplamları elde edilerek de hesaplanabilir. Örneğin;

$$\text{KT}_{\text{hücre 11}} = (9 - 7)^2 + (8 - 7)^2 + \ldots + (7 - 7)^2 = 30$$

Diğer hücrelere de benzer işlemleri yaparak aşağıdaki eşitlikleri elde edebiliriz:

$$KT_{\text{hücre 11}} = 30{,}0$$
$$KT_{\text{hücre 12}} = 40{,}9$$
$$\ldots \quad = \ldots$$
$$\frac{KT_{\text{hücre 25}}}{KT_{\text{hata}}} = \frac{64{,}1}{722{,}30}$$

Her hücreye ait karelerin toplamı, daha sonra 10 hücre üzerinden toplanarak elde edilir. Hücre içi değişimleri kullanmak, $KT_{\text{hata}}$ ile gösterilen hata terimini elde etmenin zor yoludur. Ortalama karelere gelirsek 2 x 5 = 10 hücrenin her birinin hücre içi varyanslarının ortalaması olacaktır.

Tablo 13.2c varyans analizinin özet tablosunu gösterir. Daha önce bildiklerimizden, kaynak sütunu ve kare toplamları sütunu açıkça bellidir. Şimdi serbestlik derecelerine bakalım: Hesaplamak basit ve kolaydır. Serbestlik derecelerinin toplam sayısı ($sd_{\text{toplam}}$) her zaman $N - 1$'e eşittir. Yaş ve durum değişkenleri için serbestlik derecesi, değişkenin düzey sayısı eksi 1'dir. Böylece, $sd_{\text{A}} = a - 1 = 1$, $sd_{\text{C}} = c - 1 = 4$ olur. Etkileşim için serbestlik derecesi, o etkileşimi oluşturan değişkenlere ait serbestlik derecelerinin çarpımıdır. Yani, $sd_{\text{AxC}} = sd_{\text{A}}$ x $sd_{\text{C}} = (a - 1)(c - 1) = 1$ x 4 = 4 olur. Bu üç kural ne kadar karışık olursa olsun tüm varyans analizlerinde kullanılır. Hatalar için serbestlik derecesi, çıkarma işlemi ile ($sd_{\text{hata}} = sd_{\text{toplam}} - sd_{\text{A}} - sd_{\text{C}} - sd_{\text{AC}}$) veya hücre içi değişkenliği temsil eden hata terimleri akılda tutularak bulunabilir. Çünkü her hücre $(n - 1)$ *sd* içerir ve *ac* sayıda hücre olduğu için $sd_{\text{hata}} = ac(n - 1) = 2$ x 5 x 9 = 90'dır.

Tek yönlü varyans analizinde kareler ortalaması, kareler toplamının ilgili serbestlik derecelerine bölünmesiyle bulunur. Aynı yöntem tüm varyans analizlerinde kullanılır.

Son olarak *F*'yi hesaplamak için, her *KO*'yu $KO_{\text{hata}}$'ya bölmeliyiz. Böylece, Yaş için F*A* = $KO_{\text{A}} / KO_{\text{hata}}$; Durum için $F_C = KO_{\text{C}} / KO_{\text{hata}}$ ve *AC* Etkileşimi için $F_{AC} = KO_{\text{AC}} / KO_{\text{hata}}$ olarak elde edilir. $KO_{\text{hata}}$'nın neden her durumda bölen olduğunu anlayabilmek için konudan kısa bir an saparak yapısal modeli ve beklenen ortalama kareleri göz önüne almalıyız. Fakat, öncelikle analiz sonuçlarımızın ne söylediği ile ilgilenmemiz gerekiyor.

## Yorumlama

Tablo 13.2c'deki özet tablodan anlaşılacağı gibi Yaş, Durum ve ikisinin etkileşiminden kaynaklanan manidar etkiler vardır. Ortalamalar bağlamında genç katılımcıların yaşlı katılımcılara göre daha fazla öğe hatırladığı açıktır.

Açık olan bir diğer sonuç, daha çok işlem gerektiren görevlerin daha az işlem gerektirenlere göre daha iyi hatırlamaya yol açtığıdır. Bu, Bölüm 11'de bulduğumuz farklılıktır. Manidar etkileşim bir değişkenin etkisinin, diğer değişkenin düzeyine bağlı olduğunu gösterir. Örneğin, yaşlı ve genç gruplar arasındaki fark, Sayma ve Kafiye gibi daha kolay görevlerde, Resim ve Amaç gibi daha zor görevlere göre daha azdır. Başka bir açıdan, beş koşul için yaşlı katılımcılarda olan farklılaşma, genç katılımcılarda olan farklılaşmadan daha azdır.

Bu sonuçlar Eysenck'ın hipotezini desteklemektedir, daha çok bilgi işleme gerektiren görevlerde genç katılımcılar yaşlılara göre daha iyi performans gösterirken, daha kolay görevlerde hemen hemen aynı performansı göstermektedirler. Bu sonuçlar yaşlı katılımcıların bilgi işleme kapasitelerinin olmadığı anlamına gelmez. Sadece yaşlı katılımcılar gençlere göre daha düşük performans göstermektedir.

## 13.2 Yapısal Model ve Beklenen Kareler Ortalamaları

Tek-yönlü varyans analizini tartışırken yapısal modeli ele almış ve eşitliğini aşağıdaki gibi yazmıştık;

$$X_{ij} = \mu + \tau_j + e_{ijk}$$

Bu eşitlikte, $\tau_j = \mu_j - \mu$, $j$'inci işlemin etkisini temsil eder. İki yönlü desenlerde iki "işlem" değişkenlerimiz ($A$ ve $B$ diyelim) ve ikisinin etkileşimi vardır. Bunlar, daha karmaşık modeller üretilirken $\alpha$, $\beta$, ve $\alpha\beta$ ile gösterilir ve model aşağıdaki gibi yazılabilir:

$$X_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

burada

$X_{ij}$ = herhangi bir gözlem
$\mu$ = genel ortalama
$\alpha_i$ = $A_i$ faktörünün etkisi $A_i = \mu_{Ai} - \mu$
$\beta_j$ = $B_j$ faktörünün etkisi $B_j = \mu_{Bj} - \mu$
$\alpha\beta_{ij}$ = $A_i$ ve $B_j$ faktörlerinin etkileşiminin etkisi
$= \mu - \mu_{A_i} - \mu_{B_j} + \mu_{ij}$; $\sum_i \alpha\beta_{ij} = \sum_j \alpha\beta_{ij} = 0$
$e_{ijk}$ = $X_{ij}$ gözlemine ait birim hata
$= N(0, \sigma_e^2)$

Bu modelden sabit değişkenlerin beklenen kareler ortalaması Tablo 13.3.'teki gibi gösterilebilir. Her $F$ oranı için hata teriminin paydada olmasının uygunluğu açıktır, çünkü $E(KO)$'de, $\sigma_e^2$'den başka sadece bir terim vardır.

**Tablo 13.3** İki Yönlü Varyans Analizi İçin Beklenen Ortalama Kareler (Sabit)

| Kaynak | *E(MS)* |
|---|---|
| $A$ | $\sigma_e^2 + nb\theta_\alpha^2$ |
| $B$ | $\sigma_e^2 + na\theta_\beta^2$ |
| $AB$ | $\sigma_e^2 + n\theta_{\alpha\beta}^2$ |
| Hata | $\sigma_e^2$ |

© Cengage Learning 2013

burada $\theta_\alpha^2 = \dfrac{\Sigma\alpha_j^2}{a-1} = \dfrac{\Sigma(\mu_i - \mu)^2}{a-1}$

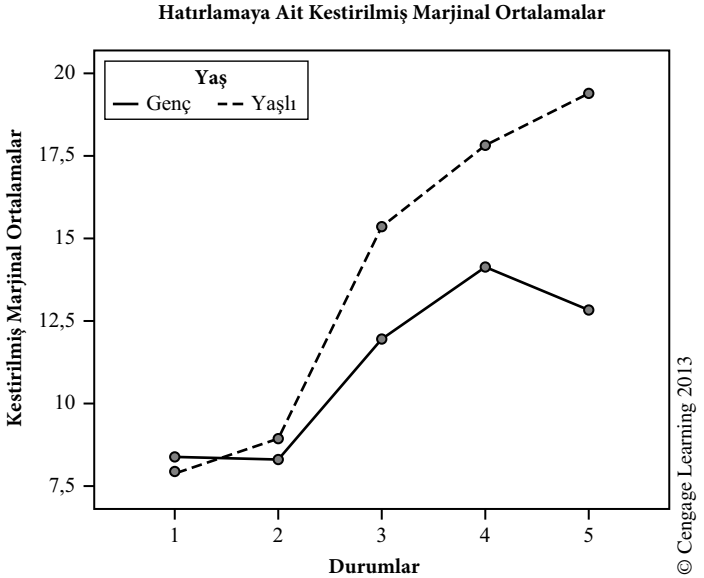
Faktör A'nın etkisini test etmeyi ele alalım:

$$\frac{E(MS_A)}{E(MS_{hata})} = \frac{\sigma_e^2 + nb\theta_\alpha^2}{\sigma_e^2}$$

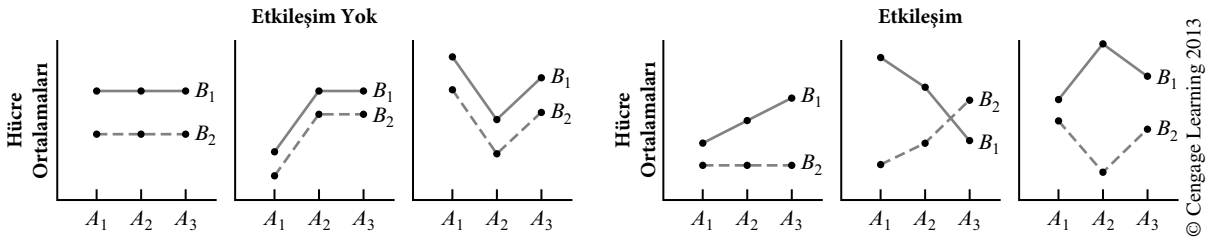
Eğer, $H_0$ hipotezi doğru ise, $\mu_{A_1} = \mu_{A_2} = \mu$ ve $\theta_\alpha^2$ ve böylece $nb\theta_\alpha^2$ değeri 0 olacaktır. Bu durumda $F$, hatanın iki farklı kestiriminin oranı olacaktır ve standart $F$ dağılımına göre yaklaşık olarak 1 olması beklenir. Eğer, $H_0$ hipotezi yanlış ise $\theta_\alpha^2$ 0 olmayacaktır ve standart (merkezî) $F$ dağılımının dışına çıkarak $F$'in 1'den büyük olması beklenir. Aynı mantık $B$ ve $AB$'nin etkilerini test ederken de kullanılabilir. Kullanılabilecek alternatif yöntemleri tartışırken Kesim 13.8'de yapısal modele ve beklenen kareler ortalamasına geri döneceğiz. Beklenen kareler ortalaması daha karışık gelebilir, fakat belirli etkilerdeki hata terimlerine karar verirken daha mantıklı sonuçlar verilebileceğini göreceğiz.

## 13.3 Etkileşimler

Faktöriyel desenlerin en önemli faydalarından biri, bize değişkenlerin etkileşimlerini inceleme şansı vermesidir. Birçok durumda etkileşim terimi temel etkiden daha ilginç olabilir (faktör etkileri tek tek ele alındığında). Eysenck'ın çalışmasını ele alalım. Her yaş grubu için ayrı ayrı ortalamaların grafiği Şekil 13.1'de görülmektedir. Daha önce sonuçların yorumunda belirttiğim gibi, Durum nedeniyle genç gruplar arasındaki farklılıkların, yaşlı gruplara ait farklılıklardan daha büyük olduğu açıkça görülmektedir. İki çizginin paralel olmaması bize etkileşim ile ilgili bir ipucu verir. Eğer koşul farklılıkları her iki yaş grubu için de aynı olsaydı, iki çizginin paralel olması beklenirdi – ancak genç katılımcılar için Durumlar arası farklılık varken, yaşlı katılımcılar için eşitlik vardır. Bu, genç katılımcıların yaşlılardan genel olarak daha üstün olup olmamasından ve iki yaş grubunun karşılaştırılıp karşılaştırılmamasından bağımsız olarak doğru olabilir. Genç katılımcılara ait çizginin yükselmesi veya alçalması Yaşın temel etkisini değiştirir, fakat etkileşimin varlığını veya yokluğunu gösteren hücre ortalamalarının gösterildiği grafiklere baktığımızda durum daha




**Şekil 13.1** Tablo 13.2'deki veri için hücre ortalamaları




**Şekil 13.2** Muhtemel etkileşim olmayan ve etkileşim olan durumların gösterimi

anlaşılır hale gelir. Şekil 13.2'de, ilk üç grafik etkileşimin olmadığı durumu temsil ediyor. Üç durumda da, çizgiler kırık oldukları halde bile paraleller. Başka bir açıdan baktığımızda, *B* faktörünün $A_1, A_2$ ve $A_3$ üzerindeki basit etkilerinin aynı olduğunu görebiliriz. İkinci üç grafikte ise, doğru parçaları açık bir şekilde paralel değil. İlkinde bir doğru parçası yatay, diğeri yükseliyor. İkincisinde çizgiler çaprazdır. Üçüncüsünde çizgiler çakışmıyor, ancak zıt yönde hareket etmektedirler. Her durumda, *B*'nin basit etkisi *A*'nın farklı düzeyleri için aynı değil. Yani, çizgiler (manidar şekilde) paralel değil ve bir etkileşimin olduğunu söyleyebiliriz.

Birçok insan eğer bir etkileşim bulunduysa temel etkinin göz ardı edilmesi gerektiğini savunuyor. Bu görüşü Web by Gary McClelland'ı inceledikten sonra kısmen kabul edebilirim. (McClelland, 2008, 2/3/2011 tarihinde http://finzi.psych.upenn.edu/Rhelp10/2008-February/153837.html adresinden indirilmiştir.). O dengelenmiş desenlerde (hücreleri eşit frekanslı), *A*'nın temel etkisinin *A*'nın basit etkilerinin bir ortalaması olduğunu savunuyor. Ama eğer basit etkilerin birbirinden farklı olduğunu söyleyen manidar bir etkileşimimiz varsa, bunların ortalamaları çok küçük veya anlamsız olur. Ayrıca, bir etkileşim gördüğümde ilk düşüncem bir değişkenin diğer bir değişkenin düzeylerinde olan basit etkilerine bakmak olur. Bu temel etkiler için kaygılanmaktan daha çok bilgilendiricidir.

Etkileşimin etkileri üzerindeki bu tartışmalar hücre ortalamalarının incelenmesine odaklanır. Bu yolu seçmemin sebebi, daha kolay olması ve sonuçlar hakkında daha fazla şey söyleme imkânı vermesidir. Rosnow ve Rosenthal (1989) etkileşimi incelemek için daha uygun bir yol önermişlerdir, bu yolda ilk olarak veri üzerindeki satır ve sütun etkilerini ortadan kaldırmak gerekir. İlginç bir noktaya dikkat çekiyorlar, ama yukarıdaki açıklamalarla etkileşimleri anlamak muhtemelen daha kolay hale gelecektir.

## 13.4 Basit Etkiler

Daha önce basit etkiyi, bir faktörün (bağımsız değişken) diğer faktörün bir düzeyi üzerindeki etkisi olarak tanımladım - örneğin, genç katılımcılar için durumdan kaynaklanan farklar. Basit etkileri analiz etmek, manidar etkileşim içeren veri gruplarının analizleri açısından oldukça önemli olabilir.

Basit etkinin nasıl hesaplandığını ve yorumlandığını göstermek için Eysenck'ın veri kümesini kullanacağım. Tablo 13.4 hücre ortalamalarını ve Tablo 13.2'den elde edilen özet tabloyu göstermektedir. Tablo ayrıca tüm basit etkilerin nasıl elde edildiğini içeren hesaplamaları da kapsamaktadır.

Tablo 13.4c.'deki özet tablo Yaş, Durum ve ikisinin etkileşimi sonucu oluşan etkileri açıklar. Biz daha önce bu sonuçları orijinal analiz ile birlikte tartışmıştık. Orada da söylediğimiz gibi etkileşim durumu, iki yaş değişkeni için farklı Durum etkilerinin, beş Durum için farklı Yaş etkilerinin olması anlamına gelir. Burada şunu sormamız önemli olur: Genel Durum etkimiz gençlerde olduğu gibi, yaşlı katılımcılar için gerçekten uygun olur mu

**Tablo 13.2** Basit etkilerin hesaplanmasının gösterimi (Veri, Tablo 13.2'den alınmıştır.)

**(a) Hücre ortalamaları ($n$ = 10)**

| | Sayma | Kafiye | Sıfat | Mecazî | Amaç | Ortalama |
|---|---|---|---|---|---|---|
| **Yaşlı** | 7,0 | 6,9 | 11,0 | 13,4 | 12,0 | 10,06 |
| **Genç** | 6,5 | 7,6 | 14,8 | 17,6 | 19,3 | 13,16 |
| **Ortalama** | 6,75 | 7,25 | 12,90 | 15,50 | 15,65 | 11,61 |

**(b) Hesaplamalar**

**Her Yaş İçin Durumlar/Şartlar**

$$KT_{\text{Yaşlıda}} = 10 \times [(7{,}0 - 10{,}06)^2 + (6{,}9 - 10{,}06)^2 + \cdots + (12 - 10{,}06)^2] = 351{,}52$$

$$KT_{\text{Gençte}} = 10 \times [(6.5 - 13.16)^2 + (7.6 - 13.16)^2 + \cdots + (19{,}3 - 13{,}16)^2] = 1353{,}72$$

**Her Durum İçin Yaş**

$$KT_{\text{Saymada}} = 10 \times [(7{,}0 - 6{,}75)^2 + (6{,}5 - 6{,}75)^2] = 1{,}25$$

$$KT_{\text{Kafiyede}} = 10 \times [(6{,}9 - 7{,}25)^2 + (7{,}6 - 7{,}25)^2] = 2{,}45$$

$$KT_{\text{Sıfatta}} = 10 \times [(11{,}0 - 12{,}9)^2 + (14{,}8 - 12{,}9)^2] = 72{,}2$$

$$KT_{\text{Mecazlıda}} = 10 \times [(13{,}4 - 15{,}5)^2 + (17{,}6 - 6{,}75)^2] = 88{,}20$$

$$KT_{\text{Amaçlıda}} = 10 \times [(12{,}0 - 15{,}65)^2 + (19{,}3 - 15{,}65)^2] = 266{,}45$$

**(c) Özet Tablolar**

**Genel Analiz**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* (Yaş) | 1 | 240,25 | 240,25 | 29,94* |
| *C* (Durum) | 4 | 1514,94 | 378,735 | 47,19* |
| *AC* | 4 | 190,30 | 47,575 | 5,93* |
| Hata | 90 | 722,30 | 8,026 | |
| Toplam | 99 | 2667,79 | | |

* $p < ,05$

**Basit Etkiler**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| **Durumlar** | | | | |
| *C* Yaşlıda | 4 | 351,52 | 87,88 | 10,95* |
| *C* Gençte | 4 | 1353,72 | 338,43 | 42,15* |
| **Yaş** | | | | |
| *A* Saymada | 1 | 1,25 | 1,25 | <1 |
| *A* Kafiyede | 1 | 2,45 | 2,45 | <1 |
| *A* Sıfatta | 1 | 72,20 | 72,20 | 9,00* |
| *A* Mecazda | 1 | 88,20 | 88,20 | 10,99* |
| *A* Amaçlıda | 1 | 266,45 | 266,45 | 33,20* |
| Hata | 90 | 722,30 | 8,03 | |

* $p < ,05$

© Cengage Learning 2013

veya tüm durumlar altında gerçekten Yaş farklılıkları var mı? Basit etkinin analizleri Tablo 13.4b'de ve Tablo 13.4c'nin ikinci yarısında bulunmaktadır. Örneğin, bütünlüğünü bozmamak için muhtemel tüm basit etkileri hesapladım, fakat genelde ilgi alanına göre istenilen basit etkileri hesaplamalısınız. Çok sayıda basit etki test edildiğinde ya aileboyu hata oranı kabul edilemez düzeyde yükselecektir ya da aileboyu hata oranı makul sebeplerle kontrol altına alınacak ve basit etki testinin gücü düşecektir. Basit bir kural, "Sonuçları rapor ederken sizin konumunuzla ilgili olmadıkça doğrusal karşılaştırmaları veya basit etkiyi hesaplamayınız." Ne kadar çok etki test ederseniz aileboyu hata oranı o kadar çok yükselir.

## Basit Etkilerin Hesaplanması

Tablo 13.4b'den görebileceğiniz gibi, $KT_{\text{yaşlıda C}}$ herhangi bir kareler toplamıyla aynı şekilde hesaplanır. Sadece *yaşlı katılımcıların veri kümesini kullanarak* $KT_C$ *basitçe hesaplanabilir*. Biz sadece bu veriyi beş Durum için ele alırsak, ortalamalar sırasıyla 7,0; 6,9; 11,0; 13,4 ve 12,0 hesaplanır. Kareler toplamı da aşağıdaki gibidir:

$$KT_{\text{Yaşlıda C}} = n\sum(\bar{X}_{1j} - \bar{X}_{1.})^2$$
$$= 10 \times [(7 - 10{,}06)^2 + (6{,}9 - 10{,}06)^2 + \cdots + (12 - 10{,}06)^2] = 351{,}52$$

Diğer basit etkiler de, o anda ilgilenmediğimiz veri kümesi yok sayılarak aynı şekilde hesaplanabilir. Yaşlı katılımcılar için Duruma ait basit etkinin kareleri toplamı (351,52), Alıştırma 11.1'de sadece yaşlı katılımcılara yapılan tek yönlü varyans analizi sonucunda elde edilen değer ile aynıdır.

Basit etkilerin serbestlik derecesi ilgili oldukları temel etki ile aynı şekilde hesaplanır. Bu mantıklıdır, çünkü karşılaştırdığımız ortalamaların sayısı aynıdır. Bütün katılımcıları veya bunların bir kısmını kullanıyor olsak da, hâlâ beş Durumu karşılaştırıyoruz ve Durum için 5 – 1 = 4 *sd* vardır.

Basit etkileri test etmek için genellikle tüm analizin hata terimleri kullanılır ($KO_{\text{hata}}$). Beklenen kareler ortalaması Tablo 13.5'te gösterilmiştir ve bunlar hata terimlerinin neden uygun olduğunu açık hale getirir. Her basit etki için beklenen kareler ortalaması, hatadan farklı olarak sadece bir etki daha içerir (örneğin, $n\sigma^2_{\beta_j\text{'de }a}$), oysa $KO_{\text{hata}}$ hata varyansının $\left(\sigma_e^2\right)$ bir kestiricisidir. Tablo 13.4'te yaptığım şey ile yaşlı katılımcıların verisi (genelde tek-yönlü varyans analizi bilgisayar yazılımı kullandığımda basit etkiler için kareler toplamı hesaplarım) için yapacağım standart tek-yönlü varyans analizi arasındaki tek fark hata terimidir. $KO_{\text{hata}}$ normalde tüm veri kümesine dayanır çünkü daha büyük serbestlik derecelerinde daha iyi kestirimler yapılabilir.

**Tablo 13.5** Basit etkiler için beklenen ortalama kareler

| Kaynak | *E(KO)* |
|---|---|
| ***A'nın Basit Etkileri*** | |
| $A, B_1$'de | $\sigma_e^2 + n\theta^2_{\alpha \text{ at } \beta_1}$ |
| $A, B_2$'de | $\sigma_e^2 + n\theta^2_{\alpha \text{ at } \beta_2}$ |
| $A, B_3$'te | $\sigma_e^2 + n\theta^2_{\alpha \text{ at } \beta_3}$ |
| ***B'nin Basit Etkileri*** | |
| $B, A_1$'de | $\sigma_e^2 + n\theta^2_{\beta \text{ at } \alpha_1}$ |
| $B, A_2$'de | $\sigma_e^2 + n\theta^2_{\beta \text{ at } \alpha_2}$ |
| **Hata** | $\sigma_e^2$ |

© Cengage Learning 2013

### Yorum

Yaşlı katılımcılara ait kareler toplamı genç katılımcılara ait kareler toplamının dörtte biri kadar olmakla birlikte, Tablo 13.4c'nin altındaki $F$ olarak adlandırılan sütundan, Durum nedeniyle her iki yaş için ortaya çıkan farklılıklar açıkça görülmektedir. Yaş etkisiyle ilgili olarak sayma ve kafiye gibi düşük düzeyli görevlerde bir fark oluşmazken, yüksek düzeyli görevlerde fark meydana gelmektedir. Başka bir deyişle, yaş gruplarında yüksek düzeyli görevlerde farklılaşmalar meydana gelmektedir. Bu, temelde Eysenck'ın göstermek için yola çıktığı konudur.

Genel olarak, manidar bir etkileşim olmadıkça nadiren basit etkilere bakarız. Etkileşim manidar değilse, olsa bile, basit etki analizi için veri hayal etmek zor değildir veya deneyin birinci probleminin basit etkiler olduğu bir deney de hayal edilebilir.

### Basit Etkilerin Toplanabilirliği

Varyans analizindeki bütün kareler toplamı ($KT_{\text{toplam}}$ dışındakiler) daha büyük bir kareler toplamının parçalarını temsil eder ve basit etkiler de bundan istisna değildir. Durumun her Yaş düzeyindeki basit etkisi $KT_{\text{C}}$ ve $KT_{\text{AxC}}$ şeklinde bir parçalanma gösterirken, Yaşın her Durum düzeyindeki basit etkisi $KT_{\text{A}}$ ve $KT_{\text{AxC}}$ şeklinde bir parçalanma gösterir. Böylece,

$$\sum KT_{\text{A için C}} = 351,52 + 1353,72 = 1705,24$$

$$KT_C + KT_{AxC} = 1514,94 + 190,30 = 1705,24$$

ve

$$\sum KT_{\text{C için A}} = 1,25 + 2,45 + 72,20 + 88,20 + 266,45 = 430,55$$

$$KT_A + KT_{AxC} = 240,25 + 190,30 = 430,55$$

Benzer bir toplanabilirlik ilişkisi serbestlik derecesi için de geçerlidir. Basit etkilerin kareler toplamı, özetle ilgili ana etkinin kareler toplamı ile etkileşimin birleşimi olarak düşünülürse hesaplamalarımızı kontrol etmenin basit ve hızlı bir yolu bulunur.

## 13.5 Varyans Analizinin Sigara İçmenin Etkilerine Uygulanması

Bu örnek Spilich, June ve Renner (1992) çalışmasında araştırılan sigara içmenin performans üzerindeki etkilerine dayanmaktadır. Farklı katılımcılara üç farklı görev verilmiştir. Bu görevler, bilişsel işleme performansı gerektirme düzeylerinde farklılaşmaktadırlar. Birinci görev, katılımcıların ekrana bir hedefi yerleştirmesini gerektiren Yapı tanıma görevidir. İkincisi Bilişsel bir görevdir ve bir pasajı okuduktan belirli bir süre sonra hatırlamayı gerektirir. Üçüncü görev ise Sürücülük simülasyonu içeren bir video oyunudur. Her durum için bağımlı değişken katılımcıların yaptığı hata sayısıdır. (Bu orijinal çalışmadaki bütün görevler için tam anlamıyla doğru olmamakla beraber, bana, Görevlere bağımsız değişkenmiş gibi davranmaya olanak vermektedir. Spilich ve arkadaşlarının yaptığı çalışmayı ciddi anlamda çarpıtmıyorum.)

Daha sonra katılımcılar üç Sigara içme grubuna ayrılmışlardır. Grup AS görev öncesinde veya sırasında sigara içen kişilerdir. Grup DS'de yer alan kişiler, normalde sigara içen bireyler olduğu halde, görevin üç saat öncesinden beri sigara içmemişlerdir. Grup NS ise, sigara içmeyen kişilerden oluşmaktadır.

Veri aşağıdadır, ancak veriye bakmadan önce Görev, Sigara içme ve bunların etkileşimine ilişkin bir yordamada bulunabilirsiniz.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Yapı Tanıma | | | | | | | | | |
| **NS:** | 9 | 8 | 12 | 10 | 7 | 10 | 9 | 11 | 8 | 10 | 8 | 10 | 8 | 11 | 10 |
| **DS:** | 12 | 7 | 14 | 4 | 8 | 11 | 16 | 17 | 5 | 6 | 9 | 6 | 6 | 7 | 16 |
| **AS:** | 8 | 8 | 9 | 1 | 9 | 7 | 16 | 19 | 1 | 1 | 22 | 12 | 18 | 8 | 10 |
| | | | | | | Bilişsel Görev | | | | | | | | | |
| **NS:** | 27 | 34 | 19 | 20 | 56 | 35 | 23 | 37 | 4 | 30 | 4 | 42 | 34 | 19 | 49 |
| **DS:** | 48 | 29 | 34 | 6 | 18 | 63 | 9 | 54 | 28 | 71 | 60 | 54 | 51 | 25 | 49 |
| **AS:** | 34 | 65 | 55 | 33 | 42 | 54 | 21 | 44 | 61 | 38 | 75 | 61 | 51 | 32 | 47 |
| | | | | | | Otomobil Sürme Similasyonu | | | | | | | | | |
| **NS:** | 3 | 2 | 0 | 0 | 6 | 2 | 0 | 6 | 4 | 1 | 0 | 0 | 6 | 2 | 3 |
| **DS:** | 7 | 0 | 6 | 0 | 12 | 17 | 1 | 11 | 4 | 4 | 3 | 5 | 16 | 5 | 11 |
| **AS:** | 15 | 2 | 2 | 14 | 5 | 0 | 16 | 14 | 9 | 17 | 15 | 9 | 3 | 15 | 13 |

Burada sayıltılar için kendinizin elle yapabileceği hesaplamaları atlıyorum. Aslında bunu yapmak iyi bir alıştırma olabilir. Gösterim 13.1'de bu veri kümesine ait SPSS'de analizinin yapılmış halini bulabilirsiniz.

**(a) Betimsel İstatistikler**

Bağımlı Değişken: Hatalar

| Sigara İçme Grp. | Görev | Ortalama | Std. Sapma | N |
|---|---|---|---|---|
| Sigara İçmeyenler | Yapı Hatırlama | 9,40 | 1,404 | 15 |
| | Bilişsel Görev | 28,87 | 14,687 | 15 |
| | Oto. Sürme Simül. | 2,33 | 2,289 | 15 |
| | Toplam | 13,53 | 14,130 | 45 |
| Sigara İçenler | Yapı Hatırlama | 9,60 | 4,405 | 15 |
| | Bilişsel Görev | 39,93 | 20,133 | 15 |
| | Oto. Sürme Simül. | 6,80 | 5,441 | 15 |
| | Toplam | 18,78 | 19,359 | 45 |
| Aktif İçiciler | Yapı Hatırlama | 9,93 | 6,519 | 15 |
| | Bilişsel Görev | 47,53 | 14,652 | 15 |
| | Oto. Sürme Simül. | 9,93 | 6,006 | 15 |
| | Toplam | 22,47 | 20,362 | 45 |
| Toplam | Yapı Hatırlama | 9,64 | 4,513 | 45 |
| | Bilişsel Görev | 38,78 | 18,055 | 45 |
| | Oto. Sürme Simül. | 6,36 | 5,701 | 45 |
| | Toplam | 18,26 | 18,393 | 135 |

SPSS, Inc. çıktısından adapte edilmiştir.

**Gösterim 13.1** Spilich ve arkadaşlarına ait verinin analizi

**(b) Özet Tablo**

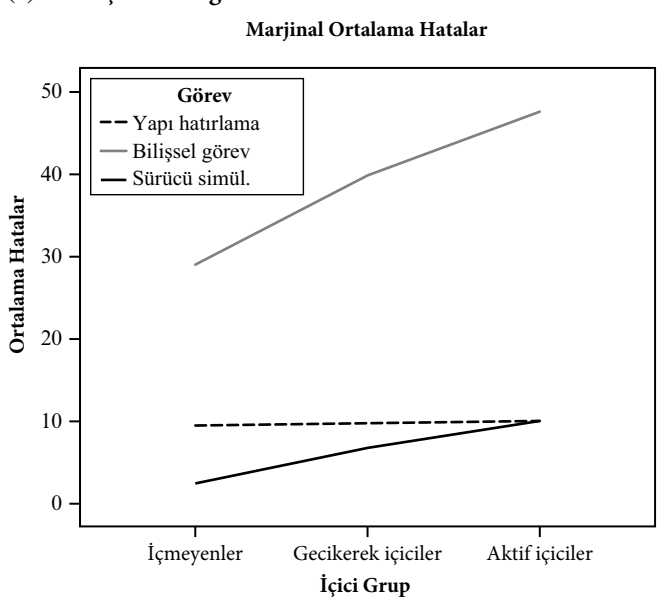
**Bireyler Arası Etkilerin Test Edilmesi**

Bağımlı Değişken: Hatalar

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Man. | Kısmî Eta Kare | Merkezî Olmayan Parametre | Gözlenen Güç[b] |
|---|---|---|---|---|---|---|---|---|
| Düzeltilmiş Model | 31744,726[a] | 8 | 3968,091 | 36,798 | ,000 | ,700 | 294,383 | 1,000 |
| Sabit (Kesim nok.) | 45009,074 | 1 | 45009,074 | 417,389 | ,000 | ,768 | 417,389 | 1,000 |
| Görev | 28661,526 | 2 | 14330,763 | 132,895 | ,000 | ,678 | 265,791 | 1,000 |
| İçici Grup | 1813,748 | 2 | 906,874 | 8,410 | ,000 | ,118 | 16,820 | ,961 |
| Görev*İçici Grup | 1269,452 | 4 | 317,363 | 2,943 | ,023 | ,085 | 11,772 | ,776 |
| Hata | 13587,200 | 126 | 107,835 | | | | | |
| Toplam | 90341,000 | 135 | | | | | | |
| Düzeltilmiş Top. | 45331,926 | 134 | | | | | | |

[a] R Kare = ,700 (Düzeltilmiş R kare = ,681)
[b] Alfa = ,05 kullanılarak hesaplanmıştır.

**(c) Etkileşim Grafiği**




**Gösterim 13.1** *(Devamı)*

Faktöriyel desen için olan SPSS özet tablosunun, ek bilgiler içerdiğinden, diğer tablolardan farklı olduğunu görebilirsiniz. "Düzeltilmiş model" olarak adlandırılan satır, ana etkilerin ve etkileşimin toplamıdır. Aslında daha önce kareler toplamı dediğimiz $KT_{\text{hücre}}$'dir. Sabit (Intercept) olarak adlandırılan satır, genel ortalama üzerinde bir testtir. Burada sürpriz olmayacak şekilde genel ortalama 0,00'dan manidar olarak farklıdır. Tablonun en altına yakın yerde bulunan "Düzeltilmiş toplam" satırı, bizim normalde toplam" dediğimiz satırdır ve "Toplam olarak adlandırılan satır ise ($\sum X^2/N$)'dir. Bu fazla satırlar nadiren ilgi çekici ek şeyler getirir.

Özet tablo Görev, Görev ve Sigara İçme Grubu ortak etkisi için manidar etkilerin olduğunu, ancak Sigara içmeyenler Grubu için manidar bir etkinin olmadığını ortaya koyar.

Görev etkisinin bir önemi yoktur, çünkü basitçe, insanlar bazı görevleri yaparken diğer görevlerden daha fazla hata yapmaktadır demektedir. Bu dünkü oyunlarda basketbol takımı futbol takımından daha fazla sayı yaptı demeye benzer. Etkileri, ortak etki grafiğinden görebilirsiniz. Betimsel istatistik tablosundaki değerlere dikkat edersek, Bilişsel görevdeki standart sapmanın ve haliyle varyansın, diğerlerinden çok daha yüksek olduğuna dikkat ediniz. Bunu daha sonraki analizler için aklımızda tutalım.

## 13.6 Ortalamalar Arası Karşılaştırmalar

Bölüm 12'de tartışılan yöntemlerin hepsi faktöriyel desenlere uygulanabilir. Eysenck örneğindeki beş Durum ortalamaları arasındaki farklılığı test edebiliriz veya Spilich örneğindeki üç Sigara içme grubu ortalamalarını standart doğrusal farklılıklar, Bonferroni *t* testi, Tukey testi veya diğer yöntemleri kullanarak karşılaştırabiliriz. Akılda tutulması gereken bir konu, Bölüm 12'deki formüllerde görülen "*n*", her işlem ortalamasının dayandığı gözlem sayısı olarak yorumlanmalıdır. Çünkü Durum ortalamalarının dayandığı (*a* x *n*) gözlem, formüle girilmesi gereken değerdir, *n* değildir. Görev ve Sigara içme grubunun etkileşimi manidar olduğundan, sonraki temel etkileri analiz etmek istemeyebilirim. Bununla beraber basit etkilerin analizi çok yararlı olabilir.

Spilich'in sigara örneğinde, Sigara içme grubunun manidar bir etkisi yoktur. O zaman, bu değişkenin üç düzeyi arasındaki doğrusal farklılıkları incelemek istemezsiniz. Çünkü bağımlı değişken (hata sayıları), görevler arasında doğrudan karşılaştırılabilir nitelikte değildir, Görev grup farklılıklarına özel olarak bakmak bir anlam ifade temez. Bunu yapabiliriz ancak kimse umursamayacaktır. (Yukarıda yapılan basketbol ve futbol takımlarına ilişkin atıfı hatırlayınız.) Bununla beraber, sigara içmenin esas etkisini, bilişsel görevler üzerinde göstermesini beklerim ve bunun için tek bir doğrusal karşılaştırma (aktif içiciler ve içmeyenler) yapabileceğiniz gibi bu basit etki için çoklu karşılaştırmalar da yapabilirsiniz. Sigara içme grubunun Bilişsel görev üzerindeki basit etkisini ele almak ve sigara içmeyenlere karşı içenleri karşılaştırmak istediğimizi varsayalım. Bu test SPSS'te Bilişsel görevler verisinde, Data/Select Cases'ten istenilen veri seçilerek yapılabilir. Compare Means/One-Way ANOVA yöntemi doğrusal karşılaştırmaları tanımlamanıza imkân verirken General Linear Model/Univariate yöntemi buna izin vermez.

SPSS'ten aşağıdaki çıktılar elde edilebilir:

**ANOVA**

**Hatalar**

| | Kareler Toplamı | sd | Kareler Ortalaması | F | Man. |
|---|---|---|---|---|---|
| Gruplar Arası | 2643,378 | 2 | 1321,689 | 4,744 | ,014 |
| Grup İçi | 11700,400 | 42 | 278,581 | | |
| Toplam | 14343,778 | 44 | | | |

**Ortalamaların Eşitliği İçin Dayanıklı Testler**

**Hatalar**

| | İstatistik[a] | sd1 | sd2 | Mnd. |
|---|---|---|---|---|
| Welch | 5,970 | 2 | 27,527 | ,007 |
| Brown-Forsythe | 4,744 | 2 | 38,059 | ,014 |

[a]Asimtotik F Dağılımı

SPSS, Inc. çıktısından adapte edilmiştir.

Basit etkiyi test etmek için olan hata teriminin (278,581), bütün analiz için olan hata teriminden (107,835) çok daha büyük olduğuna dikkat ediniz. Bu daha önce söz edilen varyans heterojenliğini yansıtmaktadır. İster standart analizdeki hata terimini ister Welch (veya Brown-Forsythe) heterojen varyans düzeltmesini kullanalım, sonuç açık bir şekilde manidardır.

Basit etkiyi, aktif sigara içiciler ve sigara içmeyenler olacak şekilde böldüğümüzde testimiz aşağıdaki gibi olur:

**Karşılaştırma Katsayıları**

| | Sigara İçici Grup | | |
|---|---|---|---|
| Karşılaşt. | Sigara İçmeyenler | Gecikmeli İçiciler | Aktif İçiciler |
| 1 | 1 | 0 | – 1 |

**Karşılaştırma Katsayıları**

| | Karşılaştırma | | Karşılaş. Değeri | Std. Hata | t | sd | Man. (2-kuyruk) |
|---|---|---|---|---|---|---|---|
| Hatalar | Varyanslar eşit kabul edildi | 1 | – 18,67 | 6,095 | – 3,063 | 42 | ,004 |
| | Varyanslar eşit kabul edilmedi | 1 | – 18,67 | 5,357 | – 3,485 | 28,000 | ,002 |

SPSS, Inc. çıktısından adapte edilmiştir.

Ve tekrar belirtelim, varyansları birleştirsek de birleştirmesek de manidar bir farklılık görmekteyiz. Görsel incelemeye dayanarak sigara içmenin Yapı tanımlamada veya Sürücü durumunda önemli bir etkisi yokken, bilişsel görevlerdeki performasta bir etkiye sahip olduğu söylenebilir.

Eğer, sigara içme değişkenindeki iki uç grubu karşılaştırmak yerine, Tukey gibi standart bir çoklu karşılaştırma testi kullanırsak sık görülen ve hoş olmayan bir sonuç alırız. Sigara içmeyen grubun Aktif içici gruptan manidar bir şekilde daha iyi performans gösterdiğini ama Gecikmeli gruptan manidar olarak daha iyi olmadığını görürsünüz. Gecikmeli grup da Aktif gruptan manidar bir şekilde farklılaşmamaktadır. Bunu, farklılaşmayan grupların altını çizerek görselleştirmemiz gerekirse aşağıdaki gibi belirtebiliriz:

İçmeyen Geciktirilmiş Aktif

Eğer mantık dersinden hatırlarsanız, $A = B$, $B = C$ ama $A \neq C$'nin bir anlam ifade etmeyeceğini bilirsiniz. Ama kesin olan Mantık ile olasılıksal olan İstatistiği birbirine karıştırmayalım. $H_0$ hipotezinin reddedilmesindeki başarısızlığın, ortalamaların eşit olduğu anlamına gelmediği unutulmamalıdır. Bu sadece ortalamaların biri daha büyük olduğu halde, bizim için yeterince farklılaşmadığı anlamına gelir. Burada Geciktirilmiş grubun Sigara içmeyen gruptan farklı olduğunu iddia edecek kadar kanıtımızın olmadığı, ama Aktif içicilerle İçmeyenlerin manidar bir şekilde farklılaştığını iddia edebilecek yeterince kanıt olduğu (örneğin, güç) bir durumdur. Bu tür durumlar çoklu karşılaştırma yöntemlerinde sıklıkla oluşur ve bizim bir miktar belirsizlikle yaşamayı öğrenmemiz gerekir.

## 13.7 Faktöriyel Deneyler İçin Güç Analizi

Sabit değişkenli faktöriyel desenlerde güç hesaplamak, esas olarak tek-yönlü desenlerde hesaplama ile ayndır. Tek-yönlü desenlerde bunu aşağıdaki şekilde tanımlamıştık.

$$\phi' = \sqrt{\frac{\sum \tau_j^2}{k\sigma_e^2}}$$

ve

$$\phi = \phi'\sqrt{n}$$

Burada $\sum \tau_j^2 = \sum(\mu_j - \mu)^2$, $k$ = işlem sayısı ve $n$ = her işlemdeki gözlem sayısıdır. Tek yönlüde olduğu gibi $\phi'$ genellikle $f$ olarak gösterilir ki adını G*Power'dan alır. İki yönlü veya yüksek mertebeli desenlerde birden fazla "işlem" vardır, fakat bu, yöntemi önemli bir şekilde değiştirmez. $\alpha_2 = \mu_j - \mu$ ve $\beta_j = \mu_j - \mu$ eşitliklerimiz olduğunu varsayalım. Burada $\mu_i$, $A_i$ işleminin parametrik ortalamasını ($B$'nin bütün düzeyleri boyunca) ve $\mu_j$ $B_j$ parametrik ortalamasını (A'nın bütün düzeyleri boyunca) gösterir. Buradan aşağıdaki terimler tanımlanabilir:

$$\phi'_{\alpha} = \sqrt{\frac{\sum \alpha_j^2}{a\sigma_{hata}^2}}$$

$$\phi_{\alpha} = \phi'_{\alpha}\sqrt{nb}$$

ve

$$\phi'_{\beta} = \sqrt{\frac{\sum \beta_j^2}{b\sigma_{hata}^2}}$$

$$\phi_{\beta} = \phi'_{\beta}\sqrt{na}$$

Bu formüllerin incelenmesi, $B$ yokmuş gibi davranıldığında, $A$'yı içeren bir yokluk hipezine karşı güç hesaplaması getirir. $B$'ye ilişkin yokluk hipotezinin hesaplanması ise, $A$'ya aynı şekilde yokmuş gibi davranılmasını gerektirir.

Ortak etkiyi kapsayacak şekilde yokluk hipotezine karşı güç hesaplaması aynı mantığı izler. Bunu aşağıdaki gibi tanımlarız:

$$\phi'_{\alpha\beta} = \sqrt{\frac{\sum \alpha_j^2}{ab\sigma_{hata}^2}}$$

$$\phi_{\alpha\beta} = \phi'_{\alpha\beta}\sqrt{n}$$

Burada $\alpha\beta_{ij}$ altında yatan yapısal modele göre $(\alpha\beta_{ij} = \mu - \mu_{i.} - \mu_{.j} + \mu_{ij})$ ifadesiyle tanımlanır. Verilen $\phi_{a\beta}$ için, testin gücünü, tek-yönlü desende yaptığımız gibi kolayca elde edebiliriz.

Bir ortak etkiye ait güç hesaplamasını göstermek için Spilich ve arkadaşlarının çalışmasının hücre ve marjinal ortalamalarını kullanabiliriz. Bu ortalamalar aşağıdaki gibidir:

| | Yapı | Bilişsel | Sürüş | Ortalama |
|---|---|---|---|---|
| **İçmeyen** | 9,400 | 28,867 | 2,333 | 13,533 |
| **Geciktirilmiş** | 9,600 | 39,933 | 6,800 | 18,778 |
| **İçici** | 9,933 | 47,533 | 9,933 | 22,467 |
| **Ortalama** | 9,644 | 38,778 | 6,356 | 18,259 |

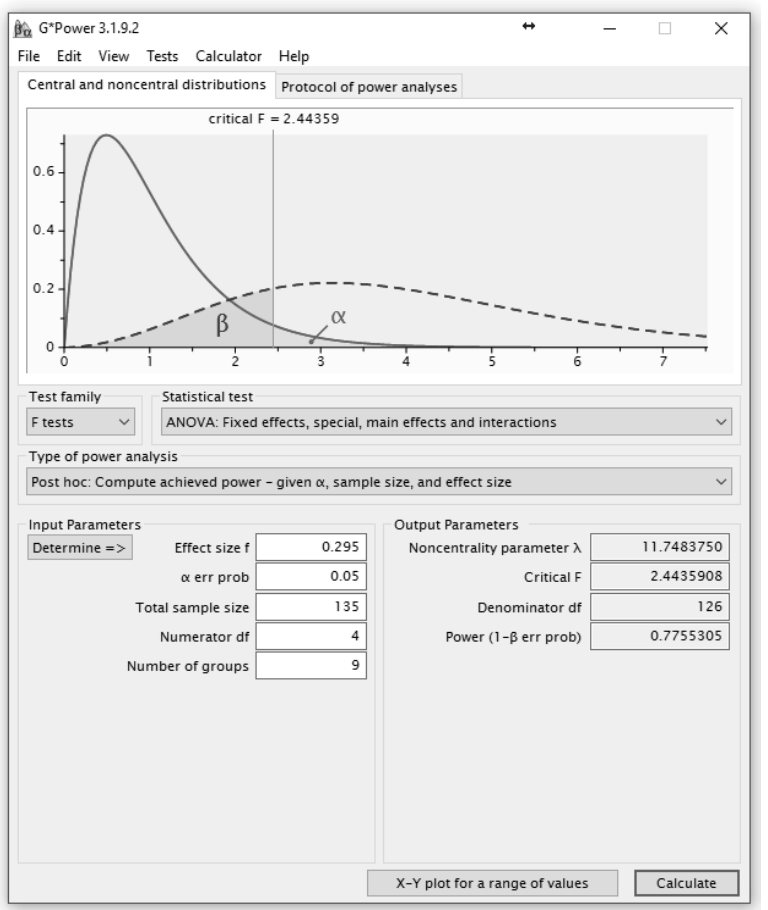
© Cengage Learning 2013

$$\phi' = \sqrt{\frac{\sum \alpha\beta_{ij}^2}{ab\sigma_{hata}^2}}$$

$$= \sqrt{\frac{(9{,}40 - 13{,}533 - 9{,}644 + 18{,}259)^2 + \ldots + (9{,}933 - 22{,}467 - 6{,}356 + 18{,}259)^2}{3 \times 3 \times 107{,}835}}$$

$$= \sqrt{\frac{20{,}088 + \ldots + 0{,}398}{970{,}515}} = \sqrt{\frac{84{,}626}{970{,}515}} = \sqrt{0{,}087} = {,}295$$

Spilich'in evren parametrelerini tam olarak kestirdiği varsayımı altında, Spilich'in deneyinin tam bir tekrarlamasıyla beklenen güç düzeyini hesaplamak istediğimizi varsayalım. (O, böyle bir evren parametresi kestirmesi kesin olarak yapmadı ama bu kestirmeler parametre değerleri için elimizde olan en iyi tahminlerdir.) Etki büyüklüğü değeri olarak 0,295 kullanılırsa (G*Power *f* adıyla anılır) aşağıdaki sonuçlar elde edilir.




Franz Faul, Edgar Erdfelder, Albert-Georg Lang and Axel Buchner/G*Power'in nazik izinleriyle

Spilich'in elde ettiği ortalamalar ve varyanslar ilgili evren ortalamasını uygun bir şekilde yansıtıyorsa, bu çalışmanın tekrarlanmasında manidar bir ortak etki elde etme olasılığı ,776'dır. Bu SPSS'ten elde edilen sonuçlarla uyuşmaktadır.

Yukarıda yer alan grafiği hatırlatmak gerekirse, şekildeki kesiksiz çizgi *F* dağılımını göstermektedir. Kesikli dağılım ise verilen evren ortalamalarına göre beklenen *F* dağılımını (merkezî olmayan) temsil etmektedir. Dikey çizgi yokluk hipotezi altında *F*'nin kritik değerini ifade eder. β ile gösterilen gölgeli alan, parametre kestirimleri doğru ise, *F* kritik değerinden az olan ve yokluk hipotezini reddedilmesine yol açmayacak şekilde elde edilebilecek *F* değerlerini belirtir. Bu durumda güç 1– β olur.

Faktöriyel desenlerin belirli avantajlarına ek olarak, belirli durumlarda iki-yönlü faktöriyel bir desen, iki ayrı tek-yönlü desenden daha güçlüdür. Her desende her işlem için katılımcı sayısının sabit tutulduğu iki hipotetik çalışma düşünün.

Deney1'de bir araştırmacı tecavüz kurbanlarına yönelik dört farklı işlemin Travma Sonrası Stres Bozukluğu için etkililiğini değerlendiriyor olsun. Araştırmacı hem erkek hem de kadın terapistler kullansın. Deneycimiz iki seçenekle karşılaşmıştır. Ya terapistin cinsi-

yet değişkenini (SexTher) göz ardı ederek dört işlem türü için analiz yapacak ya da dört işlem türü ve iki cinsiyet için 4 x 2'lik faktöriyel analiz yapacaktır. Bu durumda iki yönlü analiz tek yönlü analizden daha güçlüdür. Tek yönlü analizde SexTher'den ve SexTher'in İşlem ile etkileşiminden gelecek olan farklılaşmaları atlamış oluruz, bu da hata terimini arttırır. İki yönlü analizde ise SexTher'den ve SexTher'in İşlem ile etkileşiminden gelecek değişkenlik hesaba katılır ve bu hata teriminden eksiltilir. Böylelikle iki yönlü için olan hata terimi tek yönlü için olan hata teriminden daha düşük olur, bu da bize daha fazla güç sağlar.

Deney 2'de araştırmacının sadece kadın terapistler kullandığını düşünelim. Araştırmacının hata terimi SexTher'den ve etkileşimden kaynaklı olarak artış göstermeyecektir. Eğer deneyci çalışmasını erkek terapistler de kullanarak genişletmek isterse $KT_{\text{toplam}}$ yeni bağımsız değişkenden kaynaklı ek etkilerden dolayı artacaktır, ancak fazladan olan değişkenlik yeni terimlere yükleneceğinden hata terimi sabit kalacaktır. Bu durumda hata terimi sabit kaldığından, araştırmacı $n$'in artışından gelenin dışında, var olan orijinal çalışmasının gücünde bir güç artışına sahip olmayacaktır.

Genel bir kural olarak, ancak eklenen faktörler hata terimini sadeleştirdiğinde faktöriyel desenlerin tek-yönlü desenlere göre daha güçlü olduğu söylenebilir. Bir başka deyişle, ekstra faktörler veya değişkenler, hata teriminden kaynaklı varyansı yüklenirse, faktöriyel desenler daha güçlü olur. Diğer durumda, ortak etkileri veya basit etkileri analiz etme avantajı sağlasa da diğer her şey aynı olur.

Bir konu hakkında dikkatli olunması gerekir. Bir tesadüfî faktör eklediğinizde (örneğin, derslik) testinizin gücünü düşürebilirsiniz. Kısa bir süre içinde göreceğiniz gibi, tesadüfî faktörlü modellerde, en çok ilgilendiğiniz faktör olan sabit faktörler, muhtemelen $KO_{\text{hata}}$ yerine $KO_{\text{etkileşim}}$'i hata olarak kullanarak test edilmelidir. Bu size bir miktar güce mal olacaktır. Derslik faktörü yokmuş gibi davranamazsınız, çünkü sonrasında hataların bağımsızlığı ile ilgili problemlerin içine düşersiniz. Bu konu hakkındaki tartışma için Judd, McClelland ve Culhane (1995) incelenebilir.

Güç hakkında tartışmamız gereken ek başka bir konu daha vardır. McClelland ve Judd (1993) "optimal" adını verdikleri desen ile gücün büyük ölçüde artırılabileceğini göstermişlerdir. Bu desenlerde gücü maksimize etmek için örneklem büyüklüğü hücrelere eşit olmayacak şekilde bölüştürülmektedir. McClelland genellikle bağımsız değişkenin ihtiyacımızdan fazla olan düzeyini kullandığımızı ve bazı hücrelerde daha az veya hiç katılımcının olmamasının (özellikle sıralı düzeydeki bağımsız değişkenin merkezindeki) daha iyi olacağı halde hücrelere eşit sayıda katılımcı atadığımızı iddia etmektedir. Örneğin, iki bağımsız değişken düşünün. Faktör $A$'nın düzeyleri $A_1$, $A_2$, $A_3$, $A_4$ ve $A_5$ olsun ve Faktör $B$'nin düzeyleri $B_1$, $B_2$, $B_3$, $B_4$ ve $B_5$ olsun. McClelland ve Judd (1993) her değişkenin her düzeyinin kullanıldığı 5 x 5'lik bir desenin, sadece $A_1$ ve $A_5$ ile $B_1$ ve $B_5$'in kullanıldığı bir durumda %25'i, $A_1$, $A_3$ ve $A_5$ ile $B_1$, $B_3$ ve $B_5$'inin kullanıldığı bir 3 x 3'lük desenin ise %44'ü kadar etkili olduğunu göstermiştir. Çalışmalarını dikkatle okumanızı tavsiye ederim.

## 13.8 Alternatif Deneysel Desenler

Psikolojideki geleneksel deneysel araştırmalar için, çaprazlanmış bağımsız değişkenlerle olan sabit modeller baskın bir yaklaşım olarak kullanılagelmiştir ve öyle de devam edeceğe benzemektedir. Bu desenlerde araştırmacı her bağımsız değişkenin birkaç sabit düzeyini seçer. Bunlar öncelikli olarak ilgilenilen düzeylerdir ve araştırmacı tekrarlama yapmak isterse bu düzeyler aynı kalabilecektir. Bir faktöriyel desende her bağımsız değişkenin her düzeyi diğer bağımsız değişkenlerin bütün düzeyleriyle **çaprazlanır**.

**çaprazlanır**

tesadüfî faktör

Ancak, psikolojide ve eğitimde bu geleneksel deneysel desenlerin uygun olmadığı durumlar vardır, geleneksel deneysel desenlerde sadece birkaç durum vardır. Birçok durumda, bir veya daha fazla bağımsız değişkenin düzeyleri tesadüfî olarak örneklenir (örneğin, verilen bir okulda 10 sınıf örneklersek, sınıf bir faktördür), bu bize **tesadüfî faktör**ü verir. Başka bir durum da bir bağımsız değişkenin başka bir bağımsız değişken içinde yuvalanmasıdır. *A* bölgesindeki bir okuldan 10 sınıf ve *B* bölgesindeki başka bir okuldan bir 10 sınıf örneklersek, sözünü ettiğimiz ikinci durumun bir örneğini vermiş oluruz. Bu durumda, *A* bölgesindeki okullar *B* bölgesinde bulunmaz ve aynı şekilde *B* bölgesindeki okullar da *A* bölgesinde bulunmaz, biz buna **iç-içe/yuvalanmış desenler (nested)** deriz. Tesadüfî faktörler ve iç-içe desenler birlikte değerlendirilebilir, ancak her zaman birlikte ele alınmak zorunda değildir.

iç-içe /yuvalanmış (nested) desenler

Tesadüfî veya iç-içe bir desenimiz olduğunda, bazı değişiklikler yapmadan genel varyans analizi yapmak uygun olmaz. Birinci problem, hata terimlerinin, hesaplamak istediğimiz bir veya daha fazla *F* için doğru olmamasıdır. Bu bölümde muhtemel üç farklı desen ile çalışacağız, çapraz faktörlü geleneksel sabit desen ile başlayıp, iç-içe faktörlü tesadüfî desenler ile bitireceğiz. Mümkün olan tüm desenlerin beklentilerini tüm yönleriyle ele alamayacağım, ama tartıştığım desenlerden diğer desenlere genelleme yapmak kolay olacaktır. Bunu iki farklı sebepten yapıyorum. İlk olarak, aşağıda açıklanan düzeltilmiş geleneksel varyans analizleri bu tür durumlar için oldukça uygundur. Buna ek olarak, analizimizde **hiyerarşik modeller** veya **karma model** olarak adlandırılan genel eğilimler vardır ve bu modelleri anlayabilmek için tartıştığımız kavramların iyi anlaşılması gerekir.

hiyerarşik modeller

karma model

Devam eden kesimlerde, verinin nasıl toplandığına ilişkin farklı sayıltılarla ve farklı bağımsız değişken isimleriyle, aynı veri kümesini kullanacağım. Tüm örneklerde kullanacağım ham veri, Tablo 13.2'de gördüğümüz, Eysenck'in yaş ve değişik işlemler altındaki hatırlama durumlarına ait veri kümeleriyle aynı olacaktır. Örneğime uygun olması için değişken isimlerini değiştireceğim.

Sıkıca akılda tutulması gereken bir şey, tüm istatistiksel testlerin sanal olarak deneylerin sonsuz sayıda tekrarlanması sonrası verdiği sonuçlara göre işlemesidir. Böylece elimizdeki iki temel etki ve etkileşime ilişkin *F*'ler "Eğer sıfır hipotezi doğru ise ve deneyi 10.000 kez tekrar edersek, bu çalışmamızdaki *F* istatistiği kadar uç *F* istatistikleri ne kadar sıklıkla elde edilecektir?" sorusunu adres gösterir. Eğer bu olasılık küçükse, sıfır hipotezini reddederiz. Burada yeni bir şey yoktur. Ama bir saniyeliğine, aynı temel çalışmada 10.000 tekrarlamadan elde edilen *F* değerlerindeki değişime neyin sebep olduğunu düşünmemiz gerekir. Eysenck'ın çalışmasındaki desende, her seferinde çalışma bir grup yaşlı katılımcı ve bir grup genç katılımcı ile tekrar ediliyor. Bu bağımsız değişkende bir değişkenlik yoktur. Benzer şekilde, çalışmayı aynı beş Hatırlama Durumu (Sayma, Hece, Sıfat, Mecazî ve Amaçlı) ile tekrar ettik. Bu bağımsız değişkende de herhangi bir değişkenlik yoktur. Bağımsız değişkenin düzeyleri tekrarlamadan tekrarlamaya sabit kaldığı için bu deseni sabit etki deseni olarak adlandırdık. Bir tekrardan diğerine farklı *F* değerleri almamızın tek nedeni her tekrarda farklı katılımcıların kullanılmasından kaynaklanan örnekleme hatasıdır. (Bu sonucun tesadüfî faktörler için geçerli olmadığını biraz sonra göreceksiniz.)

Sabit-model analizinin arkasındaki temel yapısal modeli gözden geçirmek için modeli hatırlayalım;

$$X_{ijk} = \mu + \alpha_i + \beta_i + \alpha\beta_{ij} + e_{ijk}$$

Tekrarlamalar üzerindeki tek değişkenlilik son terimden ($e_{ijk}$) gelir, bu da bize $KO_{hata}$'nın neden tüm üç *F* testi için bölen görevi gördüğünü (paydada olduğunu) açıklar. Devam edebilmemiz için bu çok önemlidir.

**Tablo 13.6** Eysenk'ın temel sabit değişkenli deseninin varyans analizi

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* (Yaş) | 1 | 240,25 | 240,250 | 29,94* |
| *C* (Durum) | 4 | 1514,94 | 378,735 | 47,19* |
| *AC* | 4 | 190,30 | 47,575 | 5,93* |
| Hata | 90 | 722,30 | 8,026 | |
| Toplam | 99 | 2667,79 | | |

* $p < ,05$

© Cengage Learning 2013

## Sabit Değişkenli Çapraz Deneysel Desen

**çapraz deneysel desen**

Orijinal örneğimiz sabit faktörlerle **çapraz deneysel desen**ler olarak neyi sınıflandırdığımız üzerine olacaktır. Çapraz desende bir bağımsız değişkenin her düzeyi, diğer bağımsız değişkenin her düzeyi ile eşleştirilmiştir. Örneğin, yaşlı ve genç katılımcıların ikisi de her beş hatırlama durumu altında test edilmişlerdir. Ayrıca faktör düzeyleri sabittir, çünkü bunlar bizim gerçekten çalışmak istediğimiz düzeylerdir. (Eğer öyle olmasaydı yaşın veya bilgi işlemenin olası yöntemlerinin tesadüfî örneklemi olurdu.)

Basitçe bir referans çerçevesi olarak, bu çalışmanın analiz sonuçları Tablo 13.6'da tekrarlanmıştır. Burada 90 *sd*'ye dayalı $KO_{\text{hata}}$'nın her etki için test terimi olduğunu göreceğiz ve her etki $p < ,05$ düzeyinde manidardır.

## Tesadüfî Değişkenli Çapraz Deneysel Desen

Şimdi biraz önce analiz ettiğimiz, faktörlerden birinin tesadüfî olduğu ve bu faktörün diğer (sabit) faktör ile çaprazlandığı çalışmadan hareket edeceğiz. Judd ve McClelland (1989) tarafından yapılan bir çalışmayı örnek olarak ele alacağım. Deneklerin büyük harfleri küçük harflere göre daha hızlı tanıyıp tanımadığını test ettiğimizi düşünelim. Bu değişkeni "Durum" olarak adlandıralım. Durum burada sabit değişkendir. Birkaç farklı harf kullanmak istiyoruz, beş tanesini tesadüfî olarak örnekleyelim (örneğin; *A*, *G*, *D*, *K*, *W*) ve bunları büyük ya da küçük olarak gösterelim. Böylece Harfler değişken Durumlar ile çaprazlanmış olur (örneğin, her harf her olayda ortaya çıkıyor), elimizde çaprazlanmış bir desenimiz var, ancak bu desende Harfleri tesadüfî olarak örneklediğimizden bu bize tesadüfî faktörü verir. Her denek sadece bir harfi görecek ve bağımlı değişken o harfi tanımak için geçen tepki süresi olacaktır.

Bu çalışmada Durumlar, Eysenck'ın çalışmasındaki Yaş'ın ve Harfler de Durumların yerine geçmiştir. Bu deneyin birçok tekrarını yaparken Durumların aynı düzeylerini kullanmayı (sadece iki durum var), ama her deney için muhtemelen farklı tesadüfî örneklenmiş Harflerin olmasını beklersiniz. Bu, hesaplanan *F* değerlerindeki değişkenliğin sadece örnekleme hatasından kaynaklanmadığını, aynı zamanda örneklediğimiz harflerin de bir sonucu olduğunu gösterir. Durum ve Harfler arasındaki herhangi bir etkileşim, sabit faktör (Durum) için beklenen ortalama kareleri gösterecektir, ancak burada bunu cebirsel olarak kanıtlamayı düşünmüyorum. Bu, Durumların etkisi için beklenen ortalama kareleri etkileyecektir ve *F* oranlarını kurarken bunu dikkate almamız gerekir. (Maxwell ve Delaney, 2004, s. 475, bu olguyu anlatabilmek için mükemmel bir iş yapmışlardır.)

**beklenen kareler ortalaması**

Tesadüfî faktörlerin etkilerini görebilmek için, Kesim 11.4'te sözünü ettiğimiz **beklenen kareler ortalaması**nı göz önüne almamız gerekir. Beklenen kareler ortalaması bize bir *F* istatistiğinde pay veya payda olarak neyin kestirildiğini söyler. Beklenen kareler ortalaması için bir türetme sağlamak yerine, geçmişte sözünü ettiğim gibi (bu geliştirme için bakınız Howell, 2007), sabit, tesadüfî ve karma modeller için beklenen kareler ortalamasını basitçe bir tablo ile göstereceğim. Burada davranış bilimlerinde çok görülmeyen iki faktörün de tesadüfî olduğu bir tesadüfî model bulunmaktadır. Karma model tesadüfî ve sabit faktörlerin

**Tablo 13.7** Sabit, tesadüfî ve karma modeller için beklenen ortalama kareler ve çaprazlanmış desenler

| | Sabit | Tesadüfî | Karma |
|---|---|---|---|
| Kaynak | A sabit<br>B sabit | A tesadüfî<br>B tesadüfî | A karma<br>B karma |
| $A$ | $\sigma_e^2 + nb\theta_\alpha^2$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\theta_\alpha^2$ |
| $B$ | $\sigma_e^2 + na\theta_\beta^2$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$ | $\sigma_e^2 + na\sigma_\beta^2$ |
| $AB$ | $\sigma_e^2 + n\theta_{\alpha\beta}^2$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2$ |
| Hata | $\sigma_e^2$ | $\sigma_e^2$ | $\sigma_e^2$ |

© Cengage Learning 2013

her ikisinin de olduğu bir modeldir, çok daha yaygın olarak kullanılır. (Tesadüfî modeller için beklenen kareler ortalamasını sadece tamamlanması için koyuyorum.) Dikkat edilirse sabit faktörler için "varyans" terimi $\sigma^2$ yerine $\theta^2$ ile gösterilir. Bunun sebebi terimin sapmaların kareleri toplamının serbestlik derecesine bölünmesidir. Örneğin,

$$\theta_\alpha^2 = \frac{\Sigma\alpha_j^2}{a - 1}$$

Sabit modelde, faktörün düzeylerine, faktörde ilgilendiğimiz tüm evreni içeriyormuş gibi davranırız. Bu, parametrede olduğu gibi, tam olarak bir varyans değildir, çünkü bunun *A*'nın *sd*'sine değil *A*'nın düzey sayısına bölünmesi gerekirdi. Bu durum herhangi bir fark yaratmaz ama doğruluğunun kesin olarak ayırt edilmesi için gerekir. Tesadüfî faktörün varyans terimleri $\sigma^2$ ile gösterilir. Böylece, Harf ortalamaları varyansı $\sigma_\beta^2$ ve daima tesadüfî bir terim olarak düşünülen deneklerden kaynaklanan hata varyansı da $\sigma_e^2$ ile ifade edilir.

Eğer tamamen sabit modeli içeren sütuna bakarsak, temel etkiler için beklenen kareler ortalamasının, ana etkilerin ve etkileşimin hataya ilişkin bir bileşen içerdiğini ve bir tek bileşenin ana etki veya etkileşim için ortalama farklılıklarını yansıttığını görebiliriz. Diğer taraftan hata terimi, sadece hata bileşenini içerir. Böylece *A*, *B* veya *AB* için kareler ortalamasının oranlarını $KO_{\text{hata}}$'ya bölerek bir fomülleştirmeye gidersek, *F*'nin beklenen değerinin 1'den ayrılmasının tek nedeni, sorudaki terim için bir etki teriminin olmasıdır. (Buna benzeyen bir şeyi *F* istatistiğini ilk oluşturduğumuz Kesim 11.4.'te görmüştük.) Bu, $KO_{\text{hata}}$'nın sabit modelin tüm terimleri için uygun hata terimi olduğu anlamına gelir.

Şimdi örneğimize uygulanan karma modelin gösterildiği sütuna bakalım. Sabit etkiyi (*A*) bir kenara bırakarak diğer iki etkiye odaklanalım. Eğer oranı aşağıdaki gibi gösterirsek,

$$E(F) = E\left(\frac{MS_{\text{B}}}{MS_{\text{hata}}}\right) = \frac{\sigma_e^2 + nb\sigma_\beta^2}{\sigma_e^2}$$

bu oran tesadüfî *B* etkisi ($nb\sigma_b^2$), sadece sıfırdan farklı iken manidar bir şekilde 1'den farklı olacaktır. Böylelikle $KO_{\text{hata}}$ *B*'de yapılan *F* testi için uygun bir payda olacaktır. Bu durumda test $KO_{\text{harf}}$, $KO_{\text{hata}}$'ya bölünerek mâkul hale getirilmiş olur.

Aynı görüş etkileşim için yaptığımız test için de geçerlidir, çünkü

$$E(F) = E\left(\frac{MS_{AB}}{MS_{\text{hata}}}\right) = \frac{\sigma_e^2 + n\sigma_{\alpha\beta}^2}{\sigma_e^2}$$

ve sonuç sadece etkileşim bileşenleri manidar olduğunda manidardır.[1]

1 Eğer bir etkileşim sabit ve tesadüfî faktörlerden geliyorsa, o etkileşime tesadüfî faktör gibi davranır.

Ama şimdi $A$'da yapılan testteki sabit etkiye bakalım. Eğer her zamanki $F$ oranını oluşturursak

$$E(F) = E\left(\frac{\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_a^2}{\sigma_e^2}\right)$$

artık $A$ üzerinde yapılan mâkul bir testimiz olmaz. Etkileşim manidar olduğunda veya $A$'nın etkisi manidar olduğunda bu oran büyük olacaktır ve bu sonuca hangisinin sebep olduğunu söyleyemeyiz. Bu bir probleme sebep olur. $A$ için yapılan $F$ testini geçerli kılmanın tek yolu $KO_A$'yı $KO_{AB}$'ye bölmektir, bu bize aşağıdakini verir:

$$E(F) = \frac{KO_A}{KO_{AB}} = E\left(\frac{\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_a^2}{\sigma_e^2 + n\sigma_{\alpha\beta}^2}\right)$$

Deneyimlerime dayanarak, insanların, tesadüfî etkiler yerine sabit etkiler üzerinde test yaptığım için hata yaptığımı ve bunun problemlere yol açtığını söyleyeceklerini biliyorum. Yanılıyor olmayı isterdim ama yanılmıyorum. Tesadüfî etkilere sahip olduğumuzda bu, sabit etkilerin yerine geçer. Bunun neden olduğu ile ilgili güzel bir açıklama için Maxwell ve Delaney (2004, s. 75) çalışmalarına bakmanızı öneririm.

Örneğimizde F testini aşağıdaki gibi oluşturabiliriz:

$$F_{Durum} = \frac{K_{Durum}}{KO_{HxD}} = \frac{240,25}{47,575} = 5,05$$

$$F_{Harf} = \frac{KO_{Harf}}{KO_{Hata}} = \frac{378,735}{8,026} = 47,19$$

$$F_{HxD} = \frac{KO_{HxD}}{KO_{Hata}} = \frac{47,575}{8,026} = 5,93$$

Analizin sonuçları Tablo 13.8'de gösterilmiştir.[2]

## İç-içe (Nested) Desenler

Şimdi temel çalışmamızı bağımlı değişkenin değerlerini koruyarak değiştirelim, böylece sonuçları karşılaştırabiliriz. Klinik psikoloji programının, bayan öğrencilerin erkek öğrencilere göre daha iyi terapistler olup olmadığıyla ilgilendiğini düşünelim. Çalışmamız için, tesadüfî olarak 10 lisansüstü öğrencisi seçelim, cinsiyete göre iki gruba ayıralım ve bunların her birini

**Tablo 13.8** Bir sabit ve bir seçkisiz değişkenli varyans analizi

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Durum | 1 | 240,25 | 240,250 | 5,05* |
| Harf | 4 | 1514,94 | 378,735 | 47,19* |
| Durum×Harf | 4 | 190,30 | 47,575 | 5,93* |
| Hata | 90 | 722,30 | 8.026 | |
| Toplam | 99 | 2667,79 | | |

* $p < ,05$

Çıktı, SPSS. Inc'dan uyarlanmıştır.

[2] (Bu sonuçlar paydadaki $F$ için, karma modellere tesadüfî model gibi davranan yazılımlarca üretilen sonuçlardan farklılaşacaktır. Fakat, verilen, beklenen ortalamalar ve diğerleri için tutarlıdır. Bu sonuçları SPSS'ten aşağıdaki syntax aracılığıyla yeniden üretebilirsiniz.

```
Manova dv by Case (1,2) Letter(1,5)
    /design= Case vs 1
        Case by Letter = 1 vs within
        Letter vs within.
```

10 hasta ile çalıştırıp tedavilerinin etkililiğini ölçelim. Bu durumda her tekrar Kadın ve Erkek terapistleri içerdiğinden cinsiyet kesinlikle sabit değişkendir. Fakat, terapistler, en iyi, tesadüfî faktör olarak çalışılır; çünkü terapistler tesadüfî olarak örneklenmiştir ve sadece çalıştığımız belirli terapistlere değil, geneldeki erkek ve kadın terapistlere genelleme yapılmak istenmektedir. Terapistler aynı zamanda iç-içe (nested) faktörlerdir çünkü Cinsiyet ile Terapistleri çaprazlayamayız - Mary bir erkek terapist olarak ve Bob da bir kadın terapist olarak hizmet veremez. Çalışmanın çok defa tekrarlanması sonucu *F*'deki değişimler tesadüfî hatadan ($KO_{\text{hata}}$) ve ayrıca kullanılan terapistlerden de kaynaklanabilir. Bu değişkenlik *F* istatistikleri hesaplanırken dikkate alınmalıdır.[3]

Açıkladığım bu çalışma, daha önce sözünü ettiğimiz Harf ve Durum örneğimize benzese de tam olarak aynı değildir. Bu çalışmada terapistler cinsiyet içine yerleştirilmiştir (cinsiyet ile iç-içedir). (İlk örneğimizde her şart -Harf vb.- her durum ile eşleştirilmişti, fakat burada böyle değildir.) İç-içe bir desene sahip olduğumuz gerçeği veriyi analiz ederken oldukça önemli olacaktır. Etkileşimi hesaplayamıyor olmamız önemli bir ayrıntıdır. Açıkça Barbara, Lynda, Stephanie, Susan ve Joan kadın ve erkek olduklarında nasıl göründükleriyle ilgili ayrımı soramayız. Beş kadın kendi arasında ve beş erkek kendi arasında daha farklı görünebilir, ama bu bize bir etkileşim olduğunu göstermez.

Analizin devamında tıpkı daha önceki örneğimizde Durum etkileri için hesapladığımız gibi, cinsiyetten kaynaklanan farklılıkları da bu veri için hesaplayabiliriz. Ancak, Terapistlere geldiğimizde, sadece *kadın terapistler* içindeki farklılıkları ya da sadece *erkek terapistler* içindeki farklılıkları hesaplayabiliriz. Bunlar her cinsiyetteki terapistlerin basit etkileridir. Biz bunu "Cinsiyet içinde Terapist" olarak ifade edeceğiz ve Terapist (Cinsiyet) olarak göstereceğiz. Daha önce söylediğim gibi, bu desende etkileşim hesaplanamadığı için özet tabloda etkileşim görünmeyecektir. Sonuç olarak önceki örneğimizdeki aynı tesadüfî hata kaynağına sahibiz, bu da Cinsiyet/Terapist hücrelerin her birinde hasta puanlarının değişkenliğinin bir ölçüsüdür.

İç-içe desenler için modelimizi aşağıdaki gibi gösterebiliriz:

$$X_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{ijk}$$

Dikkat edilirse bu model genel ortalama için ($\mu$), cinsiyetler arasındaki farklılık için ($\alpha_i$), terapistler arasındaki farklılık terimi Cinsiyet içinde Terapist olduğunu ifade eden belirteçlerle birlikte ($\beta_{j(i)}$) terimi içermektedir. Etkileşim hesaplanamadığı için modelde yoktur, ama klâsik hata terimi ($e_{ijk}$) mevcuttur.

## İç-içe (Nested) Desenler İçin Hesaplamalar

Daha önce gördüğümüz desenlerden biraz farklı olsa da, iç-içe desenler üzerinde hesaplamalar yapmak kolaydır. Cinsiyet için kare toplamlarını her zaman yaptığımız şekilde hesaplayabiliriz - her cinsiyet için sapmaların kareleri toplamı, her cinsiyetteki gözlem sayısı ile çarpılır. İç-içe etki için, her cinsiyet için terapistlerin basit etkisini hesaplayabiliriz ve sonra basit etkileri toplarız. Hata terimi için, her Terapist/Cinsiyet hücresindeki hata kareleri toplamlarını hesaplarız ve bunların hepsini toplarız. Hesaplamalar Tablo 13.9'da gösterilmiştir. Ancak, bu desen için *F* istatistiğini hesaplamadan önce, sabit değişken içine geçmiş tesadüfî bir değişken varsa beklenen kareler ortalamasına da bakmamız gerekir. Bu beklenen kareler ortalaması Tablo 13.10'da gösterilmiştir, sadece tek tesadüfî faktörlü iç-içe geçmiş desen üzerine konuşuyor olsak da, bu tabloda beklenen kareler ortalamasını sabit ve tesadüfî model olarak parçaladım.

---

[3] İç-içe değişken sabit bir değişken olacak şekilde bir desen yapmak mümkündür, ancak bu davranış bilimlerinde ender olarak görülür. Burada, bu deseni, tablodaki beklenen ortalama karelerin gösterimi haricinde tartışmayacağız.

**Tablo 13.9** Tesadüfî etkilerle iç-içe (nested) desen

$$
\begin{aligned}
KT_{\text{toplam}} &= \sum(X - \bar{X}..)^2 \\
&= (9 - 11{,}61)^2 + (8 - 11{,}61)^2 + \cdots + (21 - 11{,}61)^2 \\
&= 2667{,}79 \\
KT_{\text{G}} &= nc\sum(\bar{X}_{i.} - \bar{X}..)^2 \\
&= 5 \times 4[(10{,}06 - 11{,}61)^2 + (13{,}16 - 11{,}61)^2] \\
&= 240{,}25 \\
KT_{\text{T(Erkek)}} &= n\sum(\bar{X}_{.j} - \bar{X}..)^2 \\
&= 10[(7{,}0 - 10{,}06)^2 + (6{,}9 - 10{,}06)^2 + \cdots + (12{,}0 - 10{,}06)^2] \\
&= 10(35{,}152) = 351{,}52 \\
KT_{\text{T(Kadın)}} &= n\sum(\bar{X}_{.j} - \bar{X}..)^2 \\
&= 10[(6{,}5 - 13{,}16)^2 + (7{,}6 - 13{,}16)^2 + \cdots + (19{,}3 - 13{,}16)^2] \\
&= 10(135{,}372) = 1353{,}72 \\
SS_{\text{Terapist(Cinsiyet)}} &= SS_{\text{Terapist(Erkek)}} + SS_{\text{Terapist(Kadın)}} = 351{,}52 + 1353{,}72 = 1705{,}24 \\
KT_{\text{hata}} &= KT_{\text{toplam}} - KT_{\text{G}} - KT_{\text{T(G)}} = 2667{,}79 - 240{,}25 - 1705{,}24 = 722{,}30
\end{aligned}
$$

© Cengage Learning 2013

Genellikle SPSS ve SAS syntax'larını dahil etmem ama, iç-içe desenler doğrudan SPSS menülerinden yapılamadığından bu veri kümelerinin analizi için syntax'ları vermek istedim.

```
SPSS Code
UNIANOVA
    dv BY Gender Therapist
    /RANDOM = Therapist
    /METHOD = SSTYPE(3)
    /INTERCEPT = INCLUDE
    /CRITERIA = ALPHA(.05)
    /DESIGN = Gender Therapist(Gender).

SAS Code
data GenderTherapist;
    infile 'C:\Documents and Settings\David Howell\
    My Documents\Methods8\Chapters\Chapter13\GenderTherapist.dat';
input Gender Therapist dv;

Proc GLM data = GenderTherapist;
    Class Gender Therapist;
    Model dv = Gender Therapist(Gender);
    Random Therapist(Gender)/test ;
    Test H = Gender E = Therapist(Gender);
run;
```

**Tablo 13.10** İç-içe desenler için beklenen ortalamalar

| | Sabit | Tesadüfî | Karma |
|---|---|---|---|
| Kaynak | A sabit<br>B sabit | A tesadüfî<br>B tesadüfî | A karma<br>B tesadüfî |
| *A* | $\sigma_e^2 + nb\theta_\alpha^2$ | $\sigma_e^2 + n\sigma_\beta^2 + nb\sigma_\alpha^2$ | $\sigma_e^2 + n\sigma_\beta^2 + nb\theta_\alpha^2$ |
| *B(A)* | $\sigma_e^2 + na\theta_\beta^2$ | $\sigma_e^2 + n\sigma_\beta^2$ | $\sigma_e^2 + n\sigma_\beta^2$ |
| Hata | $\sigma_e^2$ | $\sigma_e^2$ | $\sigma_e^2$ |

© Cengage Learning 2013

**Tablo 13.11** Tesadüfî iç-içe faktör ile iç-içe desen için testler

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Cinsiyet | 1 | 240.25 | 240.250 | 1.127 |
| $Hata_1$ | 8 | 1705.24 | 213.155 | 26.56* |
| Terapist(Cinsiyet) | 8 | 1705.24 | 213.155 | |
| $Hata_2$ | 90 | 722.30 | 8.026 | |
| Toplam | 99 | 2667.79 | | |

* $p < .05$

© Cengage Learning 2013

Tablo 13.10'a dikkat edilirse, sabit değişkenin içine yuvalanan tesadüfî bir değişkenimizin olduğu iç-içe bir desenimiz olduğunda, $F$ istatistiği daha farklı hesaplanır. Terapist(Cinsiyet) etkisini $KO_{TC}$'yi $KO_{hata}$'ya bölerek test edebiliriz, fakat Cinsiyeti test etmek istiyorsak $KO_C$'yi $KO_{TC}$'ye bölmemiz gerekir. $F$'lerin sonuçları Tablo 13.11'de görülmektedir, $F$'lerin nasıl yapılandırıldığının daha iyi anlaşılabilmesi için hata terimlerine alt indis ekledim.

Dikkat edilirse Cinsiyet etkileri orijinal çalışmadakiyle aynı kareler toplamına sahiptir, fakat $F$ oldukça farklıdır, çünkü Terapist(Cinsiyet) hata terimi gibi çalışmakta ve terapistler arasında göz ardı edilemez bir değişkenlik vardır. Dikkat edilirse ilk örneğimizde olduğu gibi $KT_{Terapist\ (Cinsiyet)}$, $KT_{Durum}$ ve $KT_{YaşxDurum}$'un toplamlarına eşittir, buna rağmen ben iki basit etkinin toplamı olarak düşünmeyi tercih ediyorum.

Terapistler gibi tesadüfî değişkenlere sahip olmak bazen problem yaratır. Biz cinsiyet farklılıkları ile çalışmak için yola çıktık, bu bizim en çok önemsediğimiz şeydir. Terapistlerin farklılıklarıyla çok ilgilenmeyiz, çünkü bu farklılıkların ortada olduğu açıktır. Ama Terapistlerin tesadüfî faktör olması Cinsiyet üzerinde yaptığımız testi çarpıcı şekilde değiştirir. $F$ istatistiği, yaklaşık 30'dan yaklaşık 1'e düşmüştür. Veri için aynı değerlere sahip olsak bile, çalışmanın deseninin güç üzerindeki dramatik etkisi oldukça açıktır. Maxwell ve Delaney (2004) tesadüfî faktörlü desenlerde, gücün hem denek sayısına (burada hastalar) hem de tesadüfî değişkenin düzey sayısına (burada terapist) bağlı olduğu konusuna dikkat çekmişlerdir. Genelde tesadüfî değişkenin düzey sayısı daha önemlidir.

## Özet

Üç farklı desen üzerinde çalıştık. Sabit faktörlü çapraz desenler klâsik deneysel çalışmalarda en çok kullanılan desenlerdir. Tesadüfî değişkenli iç-içe desenler, eğitim araştırmalarında kullanıldığı ve psikolojinin farklı alanlarında uygulanabilir olduğu için oldukça önemli bir desendir. Tesadüfî değişkenli çapraz desenler çok yaygın olarak kullanılmasa da karşımıza çıkar. Genelde çapraz etkiler varsa, genel olarak bunlar sabit terimdir ve iç-içe etkiler varsa onlar da genelde teadüfî terimdir. Bu, iç-içe (veya tesadüfî) desenleri açıklamak için başka kaynaklara gittiğinizde ikisinin birlikte ele alınmasının nedeni açıklamaya yardım eder. Son olarak aklınızda bulunması gereken nokta, bu kitapta ele alınan tüm denekler arası desenlerde, denekler bir diğer faktör içinde iç-içedirler (yuvalanmışlardır) ve bunlar tesadüfî bir faktör olarak kabul edilir. Tüm $F$ istatistiklerimiz bunu dikkate alarak hesaplanır.

# 13.9 İlişki ve Etki Büyüklüğü Ölçüleri

Etki büyüklüğüne iki farklı şekilde bakabiliriz, daha önce bunu tek-yönlü analiz ile yapmıştık. $\eta^2$ gibi $r$-ailesi ya da $d$ gibi $d$-ailesi ölçümlerini hesaplayabiliriz. Normalde çok amaçlı $F$'yi hesaplarken $r$-ailesi ölçümlerini kullanırız. Ayrıca, ortalamalar arasındaki doğrusal farklılığı hesaplarken etki büyüklüğü kestirimi ($d$)'yi hesaplamak daha mantıklıdır. İki ölçümü de ilerleyen kesimlerde göreceğiz.

## *r*-Ailesi Ölçümleri

Tek-yönlü analizde olduğu gibi, her bağımsız değişkenle ilişkili olan etki büyüklüğünü hesaplamak mümkündür. Bunu yapmanın en kolay fakat en yanlı yolu $\eta^2$ hesaplamaktır. Burada basitçe ilgili kareler toplamını alacağız ve $KT_{\text{toplam}}$'a böleceğiz. Böylece *A* değişkeni için etki büyüklüğü $\eta_a^2 = KT_{\text{A}}/KT_{\text{toplam}}$, *B* değişkeni için $\eta_\beta^2 = KT_B/KT_{\text{toplam}}$ ve etkileşimin etki için büyüklüğü $\eta_{a\beta}^2 = KT_{AB}/KT_{\text{toplam}}$ olur.

Bu hesaplamaları yaparken iki zorluk karşımıza çıkar. İlki, $\eta^2$, etki büyüklüğünün evrene ait değeri için yanlı bir kestirimdir. Bunu bir kenara bıraktığımızda, $\eta^2$ iyi bir betimleyici, fakat zayıf bir vardamsal istatistiktir. İkincisi, burada hesapladığımız $\eta^2$ ölçmek istediğimizi ölçmeyebilir. Bu konudan **kısmî** $\eta^2$'yi tartışırken kısaca söz edeceğiz.

**kısmî $\eta^2$**

$\omega^2$ de yanlıdır, fakat $\eta^2$'den daha az yanlıdır. Üstelik $\omega^2$'nin altında yatan istatistiksel teori bize sabit, tesadüfî ve karma modellerin ayrımını yapma ve buna göre hareket etme imkânı sağlar.

İki-yönlü ve yüksek mertebeden desenler için $\omega^2$ geliştirilirken, Tablo 13.8'de verilen beklenen kareler ortalamasıyla başlayacağız, $\sigma_\alpha^2$, $\sigma_\beta^2$, $\sigma_{\alpha\beta}^2$ ve $\sigma_e^2$ kestrimleri çıkarılacak ve sonra toplam varyansla bu bileşenlerin oranları alınacaktır. Üç farklı model için $\omega^2$ hesaplama amacıyla formül türetmek yerine, daha önceki basımda yaptığım gibi sonuçları basit bir tabloda göstereceğim. Herhangi bir öğrencinin türetimi okuduktan beş dakika sonra bunu hatırlayacağından şüphe ederim.

Faktöriyel bir varyans analizinde sabit ve tesadüfî değişken ayrımı yapmaksızın, $\omega^2$'yi kestirmek için kullanılan temel formül aynı kalır. Tek fark bu formüldeki bileşenleri nasıl hesapladığımızdır. Şu anda ilgilendiğimiz A, B ve AB gibi bağımsız değişkenlerin varyanslarını kestirirken $\hat{\sigma}_{\text{etki}}^2$ kullanarak başlayacağız ve tüm varyans kaynaklarının toplamını $\hat{\sigma}_{\text{toplam}}^2$ olarak adlandıracağız (eğer etki sabitse $\sigma^2$ yerine $\theta^2$ kullanılır. Bu terimlerin değerleri bilindiğinde $\omega_{\text{etki}}^2$ aşağıdaki gibi kestirilebilir:

$$\hat{\omega}_{\text{etki}}^2 = \frac{\hat{\sigma}_{\text{etki}}^2}{\hat{\sigma}_{\text{toplam}}^2}$$

*A*'nın ana etkisi için, örneğin bu ifade aşağıdaki hali alır:

$$\hat{\omega}_{\text{etki}}^2 = \frac{\hat{\sigma}_{\text{etki}}^2}{\hat{\sigma}_{\text{toplam}}^2} = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}_e^2}$$

Bilmemiz gereken tek şey, varyans bileşenlerini nasıl hesaplamamız gerektiğidir ($\hat{\sigma}_{\text{etki}}^2$).

Tablo 13.12, iki-yönlü analizlerde, sabit ve tesadüfî değişkenler için varyans bileşenlerini içerir, en soldaki sütundaki alt indisler sabit (*f*) veya tesadüfî (*r*) değişkenleri içindir.[4] Verilen terimlerin her birisi kolayca hesaplanarak uygun oranlar elde edilebilir. Tablo 13.8'deki desen için olan özet tablo kullanılarak bu yöntem görselleştirilmiştir. Deneklerden büyük ya da küçük harfleri tanımlamaları istenmiş ve harfler tesadüfî olarak seçilmiştir.[5]

Eğer *a* Durumun sabit etkisini ve $\beta$ Harflerin tesadüfî etkisini temsil ederse, Tablo 13.9'daki formülleri kullanarak aşağıdaki eşitlikleri elde ederiz.

$$\begin{aligned}\hat{\sigma}_\alpha^2 &= (a-1)(MS_A - MS_{AB})/nab \\ &= (2-1)(240{,}25 - 47{,}575)/(10 \times 2 \times 5) = 1{,}927\end{aligned}$$

---

4 Eğer daha yüksek dereceden desenlere ihtiyacınız varsa bunlardan birini www.uvm.edu/~dhowell/StatPages/More_Stuff/Effect_size_components.html adresinde bulabilirsiniz.

5 Bazı yazarlar benim yaptığım gibi yaparlar ve hem tesadüfî hem de sabit faktör etkileri için $\omega^2$ kullanırlar. Bazıları ise sabit faktörler $\omega^2$ ve tesadüfî faktörler için $\rho^2$ (sınıf içi korelasyon katsayısının karesi) kullanırlar.

**Tablo 13.12** İki yönlü faktöriyel desenlerdeki varyans bileşenleri kestirimleri

| Model | Varyans Bileşenleri |
|---|---|
| $A_f B_f$ | $\hat{\theta}^2_{\alpha} = (a-1)(MS_A - MS_e)/nab$ |
| | $\hat{\theta}^2_{\beta} = (b-1)(MS_B - MS_e)/nab$ |
| | $\hat{\theta}^2_{\alpha\beta} = (a-1)(b-1)(MS_{AB} - MS_e)/nab$ |
| | $\sigma^2_e = MS_e$ |
| $A_f B_r$ | $\hat{\theta}^2_{a} = (a-1)(MS_A - MS_{AB})/nab$ |
| | $\sigma^2_b = (MS_B - MS_e)/na$ |
| | $\hat{\theta}^2_{\alpha\beta} = (a-1)(MS_{AB} - MS_e)/na$ |
| | $\sigma^2_e = MS_e$ |
| $A_r B_r$ | $\sigma^2_{\alpha} = (MS_A - MS_{AB})/nb$ |
| | $\sigma^2_{\beta} = (MS_B - MS_{AB})/na$ |
| | $\sigma^2_{\alpha\beta} = (MS_{AB} - MS_e)/n$ |
| | $\sigma^2_e = MS_e$ |

Özet tablo Eysenck'ın verisi için uygun değişken isimleri kullanılarak kolaylık olması yönünden aşağıda yeniden üretilmiştir.

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *C* (Durum) | 1 | 240.25 | 240.250 | 29.94* |
| L (Harf) | 4 | 1514.94 | 378.735 | 47.19* |
| CL | 4 | 190.30 | 47.575 | 5.93* |
| Hata | 90 | 722.30 | 8.026 | |
| Toplam | 99 | 2667.79 | | |

*$p < .05$

© Cengage Learning 2013

$$\hat{\sigma}^2_{\beta} = \left(KO_{\beta} - KO_{hata}\right)/na$$
$$= (378{,}735 - 8{,}026)/10 \times 5 = 7{,}414$$
$$\hat{\sigma}^2_{\alpha\beta} = (\alpha - 1)\left(KO_{\beta} - KO_{hata}\right)/na$$
$$= (2-1)(47{,}575 - 8{,}026)/(10 \times 2) = 1{,}977$$
$$\hat{\sigma}^2_{e} = KO_{hata} = 8{,}026$$

Böylece

$$\hat{\sigma}^2_{toplam} = \hat{\sigma}^2_{\alpha} + \hat{\sigma}^2_{\beta} + \hat{\sigma}^2_{\alpha\beta} + \hat{\sigma}^2_{e}$$
$$= 1{,}927 + 7{,}414 + 1{,}977 + 8{,}026 = 19{,}344$$

Şimdi her etki için $\omega^2$ kestirimleri yapılabilir.

$$\omega^2_{Durum} = \frac{\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2_{toplam}} = \frac{1{,}927}{19{,}344} = 0{,}10$$

$$\omega^2_{Hata} = \frac{\hat{\sigma}^2_{\beta}}{\hat{\sigma}^2_{toplam}} = \frac{7{,}414}{19{,}344} = 0{,}38$$

$$\omega^2_{Durum \times Hata} = \frac{\hat{\sigma}^2_{\alpha\beta}}{\hat{\sigma}^2_{toplam}} = \frac{1{,}977}{19{,}344} = 0{,}10$$

## Kısmî Etkiler

$\omega^2$ ve $\eta^2$'nin ikisi de deneydeki toplam değişkenlikle ($KT_{\text{toplam}}$) ilişkili olan etki boyutunu ($KT_{\text{etki}}$) temsil eder. Genellikle sadece bir faktörü diğerinden ayrı ele almak daha mantıklı olur. Örneğin, Spilich ve arkadaşları (1992) çalışmalarında, farklı tür görevler altında sigara içmenin etkilerini ele almışlardır, görev farklılıkları oldukça fazladır ve kendilerince sınırlı ilgileri vardır. Eğer sigara içmenin etkilerini ölçmek istersek, ilgisiz varyanslarla bu ölçümü bozmak istemeyiz. Böylece sadece sigara içmenin etkisini, toplam değişkenlikte, sadece sigara içme ve hataya dayanacak şekilde kestirmek isteriz. Bunu şu şekilde ifade edilebiliriz:

$$\text{kısm}\square\omega^2 = \frac{\hat{\sigma}^2_{etki}}{\hat{\sigma}^2_{etki} + \hat{\sigma}^2_e}$$

Sonra gerekli terimleri hesaplarız ve böleriz. Örneğin, sigara içmenin görev ile etkileşiminin **kısmi etkisi** durumunda, her iki değişken de sabit görülerek aşağıdaki eşitliği elde ederiz:

$$\hat{\sigma}^2_{\text{SXT}} = (\text{s} - 1)(\text{t} - 1)(KO_{ST} - KO_e)/nst$$

$$= (3 - 1)(3 - 1)(682 - 108)/(15)(3)(3) = \frac{5166}{135} = 38{,}26$$

$$\hat{\sigma}_{\text{e}} = \text{Ms}_{\text{eror}} = 108$$

$$\omega^2_{\text{kısmî}} = \frac{\hat{\sigma}_{ST}}{\hat{\sigma}_{ST} + \hat{\sigma}_{eror}} = \frac{38{,}26}{38{,}26 + 108} = 0{,}26$$

Elde edilen değer mâkul ölçüde bir etki büyüklüğüdür.

## *d*-Ailesi Ölçüleri

**dengelenmemiş desenler**

*r*-ailesi ölçüleri ($\eta^2$ ve $\omega^2$ ), bağımsız değişkenlerden birinin düzeylerini içeren çok amaçlı *F* testi için konuşurken mantıklı gelebilir, fakat kişi grupları veya grup kümeleri arasındaki farka yakından baktığımızda, *d*-ailesi ölçüleri daha kullanışlı ve yorumlanabilirdir. Etki büyüklükleri (*d*), faktöriyel deneyler için biraz daha karışık gelebilir, çünkü öncelikli olarak "hata"nın ne olduğunu göz önünde bulundurmamız gerekir. Örneklem büyüklükleri eşit değilken de daha kompleks görülebilirler (**dengelenmemiş desenler** olarak adlandırılır). Bu bölümde biz dengelenmiş veya yarı dengelenmiş desenlerle uğraşacağız. Daha ileri düzeydeki tartışmalar için okuyuculara Kline (2004) önerilir.

*t* testi ve tek-yönlü varyans analizi yapıldığı durumlarda, etki büyüklüğünü aşağıdaki ifade ile yazarız:

$$\hat{d} = \frac{\hat{\Psi}}{\hat{s}}$$

Burada "şapkalar" kestirimleri örneklem verisinden yaptığımızın göstergesidir. $\Psi$ kestirilirken bir zorlukla karşılaşmayız, çünkü doğrusal bir farklılıktır. Unutma ihtimalinize karşı birazdan karşılaşacağınız örnekte iki grup veya grup kümeleri arasında ciddî anlamda farklılıklar vardır. Diğer taraftan uygun standart sapma kestirimleri değişkenlerimize bağlıdır. Bazı değişkenler, en azından potansiyel olarak, evrende normal olarak değişkenlik gösterir (örneğin, bir insanın bir günde tükettiği kafein miktarı), buna Glass, McGraw ve Smith (1981) çalışmalarında "teorik ilgi değişkeni" demişlerdir. Cinsiyet, dışa dönüklük, metabo-

lik hız ve uyku saatleri diğer örnekler olarak belirtilebilir. Diğer taraftan, pek çok deneysel değişkenler, acı verme gibi uyarıcı sayısı, beyin uyarma alanı, uyarıcı test büyüklüğü ve hatırlama sırasında bir işaret varlığı veya yokluğu gibi değişkenler birçok deneysel değişken evrende normal olarak değişkenlik göstermezler ve daha az kuramsal ilgi alırlar. Ayrımının zor anlaşıldığının farkındayım. Çalıştığımız değişkenler eğer teorik ilgiler değilse, biz onları neden çalışıyoruz?

Eğer probleme biraz daha farklı bakarsak mantıklı gelebilir. Üç farklı psikoterapi arasındaki farkı incelemeye çalıştığımızı varsayalım. Eğer sadece tek-yönlü analiz yaparsam, hata terimleri her türlü şeyden kaynaklanan değişkenliği kapsar, bunlardan birisi kadın ve erkeklerin terapiye verdikleri cevaplar arasındaki farklılık olabilir. Aynı çalışmada cinsiyeti bağımsız değişken olarak ele aldığımızı varsayalım. Cinsiyeti kontrol ettiğim etkide, $KO_{\text{hata}}$ cinsiyet farklılıklarını kapsamaz, çünkü analizde "onlar dışarı çıkarıldı". Böylelikle $KO_{\text{hata}}$ burada tek-yönlü analize göre daha küçük çıkacaktır. Bu güç terimi için iyi bir şeydir, ama etki büyüklüğünü hesaplarken $KO_{\text{hata}}$'nın karekökünü kullanırsam bu iyi olmaz. Eğer böyle bir şey yaparsak tek-yönlü analiz deneyindeki psikoterapiden kaynaklanan etki büyüklükleri faktöriyel deneydekinden daha farklı çıkar. Bu doğru görünmüyor. Terapi etkisi iki durumda da hemen hemen aynı olmalı. Böylece yapmam gereken, etki büyüklüğünü hesaplarken, cinsiyet değişkenliği ile terapi ve cinsiyet etkileşimini yerine koyduktan sonra hata terimlerine geri dönmek olmalıdır.

Fakat üç farklı terapiyi test ederken biraz daha farklı bir çalışma yaptığımızı varsayalım, çalışma aynı zamanda ikinci bir değişken olarak hastanın terapi sırasında soğuk su dolu bir küvete oturup oturmadığını da içersin. Şimdi hastaları soğuk su dolu bir küvete oturtmak normal değil ama, bu, sonuçlardaki değişkenliği etkileyecektir. Burada tek-yönlü analizde değişkenlik olmaz, çünkü biz bazı hastaların kendi su dolu küvetlerini getirip içine oturduklarını hayal bile edemeyiz. Bu hata terimini, bir şekilde yapay olduğu için değişkenliğin içine eklemek istemem. Tek-yönlü analiz veya faktöriyel analiz yaptığımda, terapi türlerinde aynı etki büyüklüğünü elde etmeyi isterim. Başarmak için, ilk örnekteki gibi Cinsiyetten ve Cinsiyet x Terapi etkileşiminden kaynaklanan etkileri de hata terimine eklemek gerekir ve ikinci örnekteki Su ve onun terapi ile etkileşiminin etkileri de eklenmelidir. Bunu yapmak için hangi yollar izlenmelidir. İlgilenen okuyucuların daha ileri düzey okumaları için Glass ve arkadaşları (1981) önerilir.

Eysenck'ın (1974) örneği ile yaptığımız çalışmaya dönelim. Ortalamalar ve varyans analizi özet tablosu aşağıda gösterilmiştir:

| | Sayma | Kafiye | Sıfat | Mecazî | Amaçlı | Ortalama |
|---|---|---|---|---|---|---|
| **Yaşlı** | 7.0 | 6.9 | 11.0 | 13.4 | 12.0 | 10.06 |
| **Genç** | 6.5 | 7.6 | 14.8 | 17.6 | 19.3 | 13.16 |
| **Ortalama** | 6.75 | 7.25 | 12.90 | 15.50 | 15.65 | 11.61 |

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* (Yaş) | 1 | 240.25 | 240.25 | 29.94* |
| *C* (Durum) | 4 | 1514.94 | 378.735 | 47.19* |
| *AC* | 4 | 190.30 | 47.575 | 5.93* |
| Hata | 90 | 722.30 | 8.026 | |
| Toplam | 99 | 2667.79 | | |

*$p < .05$

© Cengage Learning 2013

Beni ilgilendiren sorulardan biri, bu iki düşük düzeyli işlem (Sayma ve Kafiye) ve iki yüksek seviye (Sıfat ve Görüntüler) düzeyleri arasındaki doğrusal karşılaştırmadır. Amaçlı

grup hakkında herhangi bir düşüncem yoktur, bu yüzden göz ardı edeceğiz. Standart doğrusal karşılaştırma için katsayılarımız aşağıdaki gibidir:

| Sayma | Kafiye | Sıfat | Mecazî | Amaçlı |
|---|---|---|---|---|
| $-\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 |

$$\hat{\psi} = \left(-\frac{1}{2}\right)(6.75) + \left(-\frac{1}{2}\right)(7.25) + \left(\frac{1}{2}\right)(12.90) + \left(\frac{1}{2}\right)(15.50) + (0)(11.61) = 7.20$$

Bu karşılaştırmada,

$$t = \frac{\hat{\Psi}}{\sqrt{\frac{(\sum a_i^2)MS_{error}}{n}}} = \frac{7,20}{\sqrt{\frac{(1)(8,026)}{10}}} = \frac{7,20}{0,896} = 8,04$$

olur. Bu $t$ açıkça manidar olup yüksek zihinsel işleyiş gerektiren süreçlerin daha fazla hatırlama düzeylerine yol açtığını göstermektedir. Ancak bu fark için etki büyüklüğü istiyorum.

İki durum kümesi arasındaki fark üzerine olan etki büyüklüğüne bakıyorum. Hata terimlerini göz önünde bulundurmam gerekir. Yaş, dünyamızda normal bir değişkendir ve insanların cevaplarında değişmelere neden olur. (Eğer deneyimi Durumlar için tek-yönlü yapsaydım ve katılımcıların yaşlarını yok saysaydım, yaş değişkenliği $KO_{\text{hata}}$'nın normal bir parçası olacaktı.) Etki büyüklüğünü hesaplarken hatayı etkileyen yaş etkisine sahip olmam gerekir. Böylece $KO_{\text{yaş}}$ ve $KO_{\text{yaş x durum}}$ 'u hataya ekleyeceğim.

$$s_{\text{hata}} = \sqrt{\frac{SS_{\text{hata}} + SS_{\text{Yaş}} + SS_{A\times C}}{df_{\text{hata}} + df_{\text{Yaş}} + df_{A\times C}}} = \sqrt{\frac{722,30 + 240,25 + 190,30}{90 + 1 + 4}}$$

$$= \sqrt{\frac{1152,85}{95}} = \sqrt{12,135} = 3,48$$

Bu etki için hata terimini hesaplarken, aşağıdaki eşitliğe ulaşırız:

$$\hat{d} = \frac{\hat{\psi}}{\hat{s}} = \frac{7,20}{3,48} = 2,07$$

Yüksek düzeyde işlem gerektiren hatırlama ile düşük düzeyde işlem gerektiren hatırlama arasındaki fark yaklaşık iki standart sapmadır, bu kayda değer bir farktır. Çalıştığınız materyalin sizin hatırlamanıza yardım ettiğini düşünün. Kimin aklına gelirdi ki?

Şimdi Yaş etkisine bakmak istediğimizi düşünelim. Çünkü insanların normal olarak belleklerine taşıdıkları görevleri işleme düzeylerinde değişkenlik gösterebileceğini tahmin edebiliriz, o halde etki büyüklüğünü hesaplamak için Durum temel etkisini ve Durum x Yaş etkileşimini de hata terimine eklemeliyiz. Böylece,

$$s_{\text{hata}} = \sqrt{\frac{SS_{\text{error}} + SS_{\text{Condition}} + SS_{A\times C}}{df_{\text{error}} + df_{\text{Condition}} + df_{A\times C}}} = \sqrt{\frac{722,30 + 1514,94 + 190,30}{90 + 4 + 4}}$$

$$= \sqrt{\frac{2427,54}{98}} = \sqrt{24,77} = 4,98$$

Sadece iki yaşa sahip olduğumuzdan, doğrusal farklılık ($\Psi$) sadece iki ortalama arasındadır, bunlar (13,16 - 10,06) = 3,10.

$$\frac{\hat{\psi}}{\hat{s}} = \frac{3,10}{4,98} = 0,62$$

Bu durumda genç katılımcılar yaşlı katılımcılardan standart sapmanın yaklaşık üçte ikisi kadar farklılaşır.

### Basit Etkiler

Basit etkiler için etki büyüklüğü, doğrudan temel etki için hesaplanan yoldan türetilmiştir. Bu hesaplamadaki hata terimi, etki ile ilgili olan hata terimi ile aynıdır. Böylece yüksek düzeyden işlemler (Mecazî) için Yaşın basit etkisi aşağıdaki gibidir:

$$\hat{d} = \frac{\hat{\psi}}{\hat{s}} = \frac{(17.6 - 13.4)}{4.98} = \frac{4.20}{4.98} = 0.84$$

Benzer şekilde genç katılımcılar arasında yüksek düzeyli işlemlere karşı düşük düzeyli işlemler arasındaki doğrusal karşılaştırma için aşağıdaki işlemleri yaparız:

$$\psi = \left(-\frac{1}{2}\right)(6,5) + \left(-\frac{1}{2}\right)(7,6) + \left(\frac{1}{2}\right)(14,8) + \left(\frac{1}{2}\right)(17,6) + (0)(19,3) = 9,15$$

ve etki büyüklüğünü aşağıdaki gibi buluruz.

$$\hat{d} = \frac{\hat{\psi}}{\hat{s}} = \frac{9,15}{3,48} = 2,63$$

Bu da genç katılımcıların, düşük ve yüksek düzeyli işlem gerektiren hatırlamalarda, yaklaşık $2\frac{2}{3}$ standart sapma kadar farklılaştığı anlamına gelir.

## 13.10 Sonuçların Rapor Edilmesi

Çeşitli noktaları açıklamak amacıyla birçok hesaplama yapmış olmamıza rağmen, analizleri rapor ederken bunların bir kısmına yer vermeyeceğiz. Bana göre gereken esas bilgilendirmeyi kısaca gösterelim:

Farklı bilgi işlem düzeylerinde sözlü materyali akılda tutma etkilerinin incelendiği bir araştırmada, katılımcılara sözel bir materyali dört farklı yolun birini kullanarak işlemeleri için eğitim verildi. Bu yollar kelimelerdeki harflerin basit yoldan sayılmasından, kelimenin görsel olarak şekillendirilmesine kadar çeşitlilik göstermektedir. Beşinci durumdaki katılımcılara, kelimeyi hatırlamaları için ne yapmaları gerektiği ile ilgili çalışma dışında herhangi bir eğitim verilmedi. Deneyin ikinci boyutu ise hatırlama düzeyleri için Genç ve Yaşlı katılımcıları karşılaştırmaktı. Böylece deney 2 x 5 faktöriyel desen şeklini almıştır.

Bağımlı değişken, materyalin üç sunumundan sonra hatırlanan öğe sayısı olmuştur. Genç katılımcıların yaşlı katılımcılara göre daha fazla madde hatırladığını gösteren manidar bir Yaş etkisi ($F(1,90) = 29,94$, $p < ,05$, $\omega^2 = 0,87$) ortaya çıkmıştır. Aynı zamanda, Duruma ait manidar bir etki ($F(4,90) = 47,19$, $p < ,05$, $\omega^2 = ,554$) de vardı ve ortalamalar gözle kontrol edildiğinde, daha yüksek işlem düzeylerinde daha fazla hatırlama gerçekleştiği görülmüştür. Son olarak Yaş x Durum etkileşimi ($F4,90) = 5,93$, $p < ,05$, $\omega^2 = ,059$) olacak şekilde genç katılımcılar için daha güçlü Durum etkisi ile manidar bulunmuştur.

Düşük işleme düzeyleri (Sayma ve Kafiye) ile yüksek işleme düzeyleri (Sıfat ve Görüntü) arasındaki doğrusal karşılaştırma, yüksek işlem düzeylerinin lehine manidar bulunmuştur ($t(90) = 8,04$, $p < ,05$). $\hat{d} = 2,07$'ye eşit bir etki büyüklüğüne karşı gelir ve yüksek işlem düzeylerindeki katılımcıların düşük işlem düzeylerindeki katılımcılardan iki standart sapmadan daha üstün olduklarının göstergesidir. Bu etki, sadece genç katılımcılara baktığımızda bile daha büyüktür, $\hat{d} = 2,63$.

## 13.11 Eşit Olmayan Örneklem Büyüklükleri

Birçok deney (kesinlikle hepsi değil) her hücrede eşit sayıda gözlem olması niyetiyle desenlenmesine rağmen, kaderin acımasız eli sıkça dikkatlice kurulan bu planı bozabilir. Katılımcılar teste gelmeyi başaramayabilir, hayvanlar ölebilir, veri kaybolabilir, cihazlar bozulabilir, hasta tedaviden ayrılabilir ve bunun gibi birçok aksilik yaşanabilir. Bu problemler ortaya çıktığında, verinin doğasına bağlı ve verinin kayboluş nedenleriyle ilgili alternatif çözümlerin seçimi ile yüzleşebiliriz.

Tek-yönlü varyans analizi planladığımız zaman çözüm basittir ve bununla nasıl başa çıkacağımızı daha önce gördük. Daha kompleks desenlere sahip olduğumuzda sorunun çözümü basit değildir. Faktöriyel desendeki eşit olmayan örneklem büyüklüklerinde satırlar, sütunlar ve etkileşim etkileri artık bağımsız değildir. Bu bağımsızlık eksikliği yorumlamayı, veri kaybının neden oluştuğuna ve modeli anlamaya yönelik en iyi yaklaşımın ne olacağına karar vermeyi zorlaştırır.

Eşit olmayan örneklemlerle ilgili çok makale yazılmıştır, bunlarda uzun süredir gerçek çözümler göremiyoruz. (Ortaya çıkan karışık sorunların tek bir çözümü olmamasından kaynaklanıyor.) Ancak, genel durum için diğerlerine göre daha akla yakın bazı yaklaşımlar vardır. Ne yazık ki en yaygın ve diğerlerine göre en iyi çözümler standart bilgisayar paketlerindedir. Bölüm 15.1'de buna yönelik bir tartışma vardır. Bununla birlikte analizi düşünürken kâğıt kalem çözümleri ile ilgili bilgiler vereceğim. (Bu yöntemleri hesaplama yaparken kullanmanızı beklemiyorum ama SPSS ve benzeri programların nasıl çalıştığını anlarken size yardımcı olabileceğini umuyorum.) Bu yaklaşım genel olarak **ağırlıklandırılmamış ortalamalar** çözümü veya **eşit ağırlıklı ortalamalar** çözümü olarak anılır. Çünkü hücrelerdeki ortalamaları, hücrelerdeki gözlem sayısını göz ardı ederek ağırlıklandırıyoruz. Bu yaklaşımı tartışırken birincil amacım size kalem ve hesap makinesini bir kenara attırmak değil, ama varsayılan seçenekleri anlamak SPSS ve SAS'ın anlaşılmasında yardımcı olmaktır. Yeni bir örnek göstermeyeceğim ama Alıştırma 13.16'da benzer bir örnek bulunabilir. Eğer zorluk yaşarsanız öğrenci kullanım kitabında çevirim içi olarak bulabilirsiniz (www.uvm.edu/~dhowell/methods8 /StudentManual/StudentManual.html).

**ağırlıklandırılmamış ortalamalar**

**eşit ağırlıklı ortalamalar**

### Problem

Problemin ne olduğunu, daha önceden bildiğimiz 2 x 2'lik basit örneği ele alarak görebiliriz. Katılımcıların alkollü veya alkolsüzken otomobil kullanmaları üzerinde bir dikkat testi yaptığımızı düşünelim. Bu görev katılımcıların otomobil yoldan çıktığında veya karşılarına ansızın bir yaya çıktığındaki tepkilerinin ölçüldüğü bir sürüş simülasyonu içersin. Büyük olasılıkla alkolsüz sürücülerin alkollü sürücülere göre daha az hata yapmasını bekleriz. Biri Arizona'dan diğeri Michigan'dan iki müfettiş ile çalışalım; katılımcıları bu iki müfettişin kendi bölgelerinde ikiye ayıralım. Michigan'daki katılımcıların Arizona'daki katılımcılardan farklı olması için herhangi bir sebebimiz yoktur, Eyalet ve Alkol durumu arasında bir etkileşim olduğunu düşünmemiz için de bir sebep yoktur. Veri kümelerinin grafiğini düşündüğümüzde çizgilerin kesinlikle paralel olmasını bekleriz. Veri kümelerini bu beklentilerle zihnimizde tasarlayalım.

Dört hücrede de eşit olmayan sayıda katılımcının olduğu Tablo 13.13'de gösterilen oldukça sıra dışı veri kümesini elde ettiğimizi varsayalım. Bağımlı değişken her sürücünün yarım saatte yaptığı hata sayısı olarak alınır. Tablodaki hücre ortalamalarından görüleceği gibi, veri beklenendeki sonuçları vermektedir.

**Tablo 13.13** Eşit olmayan örneklem büyüklüklerinin karıştırıcı etkisinin gösterilmesi

| | İçki İçmeyen | İçki İçen | Satır Ortalamaları |
|---|---|---|---|
| Michigan | 13 15 16 12<br>$\overline{X}_{11} = 14$ | 18 20 22 19 21<br>23 17 18 22 20<br>$\overline{X}_{12} = 20$ | $\overline{X}_{1.} = 18{,}30$ |
| Arizona | 9 11 14 10 6<br>8 12 13 11 6 10<br>$\overline{X}_{21} = 10$ | 28 29 21 20 22<br>$\overline{X}_{22} = 24$ | $\overline{X}_{2.} = 14{,}38$ |
| Sütun Ort. | $\overline{X}_{.1} = 11{,}07$ | $\overline{X}_{.2} = 21{,}33$ | |

© Cengage Learning 2013

Michigan ya da Arizona'dan gelip gelmelerine bakılmaksızın alkollü katılımcılar, alkolsüz katılımcılara göre yaklaşık olarak 10 fazla hata yapmıştır. Michigan'da Arizona'dan daha fazla hata yapılmıştır ama bu manidar değildir. Peki sorun nerede?

Sütun ortalamalarına baktığımızda beklenen bir şeyle karşılaşırız, ama satır ortalamalarına baktığımızda Michigan ortalaması 18,3 ve Arizona ortalaması 14,38 olarak görülür. Eyaletler arasında olmaması gereken bir fark vardır. Alkollü olma etkisinin Eyalet etkisi gibi davrandığını düşündünüz mü? Bunun olması normaldir çünkü hücrelerde eşit sayıda katılımcı yoktur. Michigan ortalamasının fazla olmasının sebebi alkollü katılımcıların daha fazla olmasından kaynaklanır, Arizona ortalaması düşük çünkü daha fazla alkolsüz katılımcı vardır. Şimdi caddelerdeki gerçek insanları kullandığımızı ve Michigan'ın daha çok sarhoşa sahip olduğunu düşünürsek, belki Michigan için yüksek ortalama daha mantıklı gelebilir. Ama yaptığımız bu değil ve Alkollü etkiyle kirletilmiş Eyalet etkisi istemiyoruz. Peki ne yapmamız gerekiyor?

Yapılacak en mantıklı şey hücre büyüklüklerini görmezden gelerek satır ve sütun ortalamalarını hesaplamak olur. Her hücrede kaç katılımcı olduğuna dikkat etmeden sadece hücre ortalamasını elde edebiliriz. Bunu yaparsak Michigan ve Arizona'nın ikisi için ortalama (14 + 20)/2 = 17 ve (10 + 24)/2 = 17 olur ve Eyaletten kaynaklanan bir fark ortaya çıkmaz. Daha sonra bu ortalamalar faktöriyel varyans analizi için standart formüllerde yerine yerleştirilebilir, ama örneklem büyüklüğü için ne kullanılacak? İlk düşünceniz örneklem büyüklüklerinin ortalamasını almak olabilir ve bu yakın bir tahmindir. Bunun yerine örneklem büyüklüklerinin harmonik ortalamasını kullanacağız. Harmonik ortalama aşağıdaki gibi tanımlanır:

$$\overline{X}_h = \frac{k}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \cdots + \frac{1}{X_k}}$$

"$h$", ortalamanın "harmonik" olduğunu ve $k$ ortalamalarını aldığımız gözlem sayısını temsil eder. Şimdi Tablo 13.2'de gösterilen formülleri, $n$ yerine $n_k$ kullanarak, satır ve sütunlara ait hücre ortalamaları yardımıyla yeniden hesaplayabilirsiniz. Bu örnek için satır ortalaması 17 ve 17, sütun ortalaması 12 ve 22 olarak bulunacaktır ve genel ortalama hücrelerin ortalaması olacaktır. Tek fark, hata terimi olan $KT_{\text{hata}}$ çıkarma ile elde edilemez; bunun yerine desendeki her hücre için $KT_{\text{hücre içi}}$ hesaplanır ve hatadan kaynaklanan kareler toplamını elde etmek için bunlar toplanır.

Problemi dengelenmemiş desenlerde de bu yolla çözmenizi önermiyorum. Fisher'ın $F$ dağılımı gibi dağılmadığı $F$ değerleri ile 2 x 2'den büyük desenlerde de Bölüm 15'te önereceğim çözümün cevabı buna benzer olacaktır. Bu yaklaşımı burada açıkladım çünkü, eşit büyüklükte olmayan örnekleminiz varsa ve varsayılan seçenekleri seçmişseniz (Tip III kare toplamları gibi) software programlarını anlamanıza yardımcı olacaktır. Aynı zamanda ağırlıklandırılmamış sütun ortalamaları farklı değilken satırlar üzerindeki sütun etkisi nasıl oluşuyor sorusunun cevabını anlamanıza da yardımcı olacağını umuyorum.

## 13.12 Yüksek Mertebeden Faktöriyel Desenler

İki-yönlü faktöriyel desene ilişkin bütün ilkeler üç yönlü veya yüksek mertebeden desenlere aynı şekilde uygulanabilir. Ek bir bilgi olarak, varyans analizi hayal edilebilecek bütün faktöriyel desenlere uygulanabilir, sadece, değişken eklendikçe aritmetik daha karışık hale gelir. Biz en kolayı olduğu için üç yönlü faktöriyel desen ile çalışacağız.

Üç yönlü analizin iki-yönlü analizden en ciddî farkı birden fazla etkileşim teriminin bulunmasıdır. Bunu görmek için üç değişkenli faktöriyel desene ait yapısal modele bakalım:

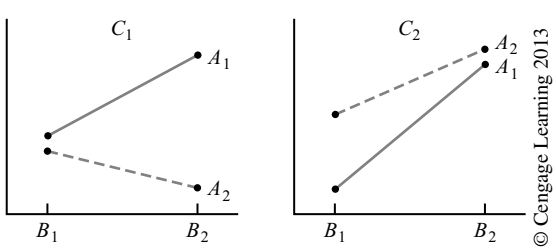
$$X_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + e_{ijkl}$$

**birinci dereceden etkileşimler**

**ikinci dereceden etkileşim**

Bu modelde $\alpha_i$, $\beta_j$ ve $\gamma_k$ gibi ana etkiler olduğu gibi iki çeşit de etkileşim terimi vardır. İki-değişkenli veya **birinci dereceden etkileşimler** olan $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$ ve $\beta\gamma_{jik}$ *A* ve *B*, *A* ve *C*, *B* ve *C* değişkenlerinin ortak etkileşimlerini gösterir. Bununla birlikte **ikinci dereceden etkileşim** olarak adlandırılan bir diğer etkileşim terimimiz olan $\alpha\beta\gamma_{ijk}$, üç değişkenin birden ortak etkileşimini gösterir. Birinci dereceden etkileşimleri iki-yönlü analizde tartışmış bulunuyoruz. İkinci dereceden etkileşimler çeşitli açılardan görülebilir. *ABC* etkileşimini görmenin en kolay yolu *AB* etkileşimimin C ile etkileşimine bakmak olabilir. Her iki değişken için iki düzeyimizin olduğunu ve *AB* etkileşiminin grafiğini *C*'nin her düzeyi için ayrı ayrı çizdiğimizi düşünelim. Şekil 13.3 ile gösterilen grafiği elde ederiz. Dikkat ederseniz $C_1$ ve $C_2$ için farklı *AB* etkileşimlerimizin olduğunu görürsünüz. Böylece, *AB* etkileşimi *C* ile etkileşmekte ve *ABC* ortak etkileşimini üretmektedir. Benzer bir akıl yürütme ile *B*'nin farklı düzeylerinde *AC* etkileşimi kullanılarak veya *BC* etkileşimi farklı A düzeyleri için elde edilerek yapılabilir. Sonuç aynı olacaktır.

Daha önce söylediğim gibi, üç-yönlü faktöriyel desen sadece iki-yönlü desenin küçük bir bükülme ile uzantısıdır. Bükülme etkileşim için olan kareler toplamının elde edilişinden gelir. İki-yönlü varyans analizinde, *A x B* tablosunun hücre ortalamalarını aldık, $KT_{\text{hücreler}}$'i hesapladık, ana etkileri çıkardık, $KT_{\text{AB}}$'yi elde ettik. Üç-yönlüde, çeşitli etkileşimlerimiz vardır ama bunları daha önce uyguladığımız tekniklere benzer olarak hesaplarız. Şöyle ki; $KT_{BC}$'yi elde etmek için bir *B* x *C* hücre ortalamaları tablosunu alırız (*A* üzerinden ortalanmış), $KT_{\text{hücreler BC}}$ elde ederiz, *B*'nin ve *C*'nin temel etkileri çıkarılır ve $KT_{BC}$ elde edilerek sonuçlandırılır. Aynı işlemler $KT_{AB}$ ve $KT_{AC}$ için de uygulanır. Aynı yöntemi $KT_{AB}$'yi elde etmek için de kullanabiliriz, fakat burada *A* x *B* x *C* ile başlamamız, $KT_{\text{hücreler}ABC}$'yi elde etmemiz, ardından ana etkileri ve düşük mertebeden etkileşimleri çıkararak $KT_{ABC}$'ye ulaşmamız gerekir. Başka bir deyişle her ayrı etkileşim için o an ilgilenmediğimiz değişkenler üzerine kurulu olan farklı bir ortalama tablosuna ihtiyaç vardır. Daha sonra da o tablo için $KT_{\text{hücreler}}$ elde edilir ve etkileşimdeki terimleri içeren temel etkiler ve düşük mertebeden etkileşimler bundan çıkarılır.




**Şekil 13.3** İkinci dereceden etkileşim grafiği

## Sürüş Performansını Etkileyen Değişkenler

Örneğin; deneyimli ($A_1$) ve deneyimsiz ($A_2$) sürücü tiplerinin sürüş yeteneklerini içeren hipotetik bir deney düşününüz. Bu insanlar birinci sınıf ($B_1$), ikinci sınıf ($B_2$), kötü ($B_3$) olmak üzere üç farklı yolda ve gündüz ($C_1$), gece ($C_2$) olmak üzere iki farklı koşulda sürüş yapıyor olsunlar. Bu durumda 2 x 3 x 2 faktörlü bir desenimiz olur. Deney her koşul için dört katılımcı (toplamda 48) kapsasın ve bağımlı değişken de bir millik bir yoldaki direksiyon düzeltmeleri sayısı olsun. Ham veri Tablo 13.14a'da yer almaktadır.

**Tablo 13.14** 2 x 3 x 2 faktöriyel desene ilişkin hesaplama örneği

**(a) Veri**

| | $C_1$ | | | $C_2$ | | |
|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | $B_1$ | $B_2$ | $B_3$ |
| $A_1$ | 4 | 23 | 16 | 21 | 25 | 32 |
| | 18 | 15 | 27 | 14 | 33 | 42 |
| | 8 | 21 | 23 | 19 | 30 | 46 |
| | 10 | 13 | 14 | 26 | 20 | 40 |
| $A_2$ | 6 | 2 | 20 | 11 | 23 | 17 |
| | 4 | 6 | 15 | 7 | 14 | 16 |
| | 13 | 8 | 8 | 6 | 13 | 25 |
| | 7 | 12 | 17 | 16 | 12 | 12 |

**Hücre Ortalamaları**

| | $C_1$ | | | $C_2$ | | | |
|---|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | $B_1$ | $B_2$ | $B_3$ | Ort. |
| $A_1$ | 10,000 | 18,000 | 20,000 | 20,000 | 27,000 | 40,000 | 22,500 |
| $A_2$ | 7,500 | 7,000 | 15,000 | 10,000 | 15,500 | 17,500 | 12,083 |
| Ortalama | 8,750 | 12,500 | 17,500 | 15,000 | 21,250 | 28,750 | 17,292 |

**Daha Fazla Hücre Ortalaması**

| | *AB* Hücreler | | | | *AC* Hücreler | | | |
|---|---|---|---|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | Ort. | | $C_1$ | $C_2$ | Ort. |
| $A_1$ | 15,000 | 22,500 | 30,000 | 22,500 | $A_1$ | 16,000 | 29,000 | 22,500 |
| $A_2$ | 8,750 | 11,250 | 16,250 | 12,083 | $A_2$ | 9,833 | 14,333 | 12,083 |
| Ort. | 11,875 | 16,875 | 23,125 | 17,292 | Ort. | 12,917 | 21,667 | 17,292 |

| | *BC* Hücreler | | | |
|---|---|---|---|---|
| | $B_1$ | $B_2$ | $B_3$ | Ortalama |
| $C_1$ | 8,750 | 12,500 | 17,500 | 12,917 |
| $C_2$ | 15,000 | 21,250 | 28,750 | 21,667 |
| Ortalama | 11,875 | 16,875 | 23,125 | 17,292 |

© Cengage Learning 2013

**Tablo 13.14** *(Devam ediyor.)*

**(b) Hesaplamalar**

$$KT_{\text{toplam}} = \sum (X - \overline{X}...)^2 = (4 - 17,292)^2 + ... + (12 - 17,292)^2 = 4727,92$$

$$KT_A = nbc \sum (\overline{X}_{i..} - \overline{X}...)^2 = 4 \times 3 \times 2\left[(22,50 - 17,292)^2 + (12,083 - 17,292)^2\right]$$
$$= 1302,08$$

$$KT_B = nac \sum (\overline{X}_{.j.} - \overline{X}...)^2 = 4 \times 2 \times 2\left[(11,875 - 17,292)^2 + ...\right.$$
$$= \left. + (23,125 - 17,292)^2\right] = 1016,67$$

$$KT_C = nab \sum (\overline{X}_{..k} - \overline{X}...)^2 = 4 \times 2 \times 3\left[(12,917 - 17,292)^2 + (21,667 - 17,292)^2\right]$$
$$= 918,75$$

$$KT_{\text{HücreAB}} = nc \sum (\overline{X}_{ij.} - \overline{X}...)^2 = 4 \times 2\left[(15,00 - 17,292)^2 + ... + (16,25 - 17,292)^2\right]$$
$$= 2435,42$$

$$KT_{AB} = KT_{\text{HücreAB}} - KT_A - KT_B = 2435,42 - 1302,08 - 1016,67$$
$$= 116,67$$

$$KT_{\text{HücreAC}} = nb \sum (\overline{X}_{i.k} - \overline{X}...)^2 = 4 \times 3\left[(16,00 - 17,292)^2 + ... + (14,333 - 17,292)^2\right]$$
$$= 2437,58$$

$$KT_{AC} = KT_{\text{HücreAC}} - KT_A - KT_C = 2437,58 - 1302,08 - 918,75$$
$$= 216,75$$

$$KT_{\text{HücreBC}} = na \sum (\overline{X}_{.jk} - \overline{X}...)^2 = 4 \times 2\left[(8,75 - 17,292)^2 + ... + (28,75 - 17,292)^2\right]$$
$$= 1985,42$$

$$KT_{BC} = KT_{\text{HücreBC}} - KT_B - KT_C = 1985,42 - 1016,67 - 918,87$$
$$= 50,00$$

$$KT_{\text{HücreABC}} = n \sum (\overline{X}_{ijk} - \overline{X}...)^2 = 4\left[(10,00 - 17,292)^2 + ... + (17,50 - 17,292)^2\right]$$
$$= 3766,92$$

$$KT_{ABC} = KT_{\text{HücreABC}} - KT_A - KT_B - KT_C - KT_{AB} - KT_{AC} - KT_{BC}$$
$$= 3766,82 - 1302,08 - 1016,67 - 918,75 - 116,67 - 216,75 - 50,00$$
$$= 146,00$$

$$KT_{\text{hata}} = KT_{\text{toplam}} = KT_{\text{HücreABC}} = 4727,92 - 3766,92 = 961,00$$

**(c) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* (Deneyim) | 1 | 1302,08 | 1302,08 | 48,78* |
| *B* (Yol) | 2 | 1016,67 | 508,33 | 19,04* |
| *C* (Durum) | 1 | 918,75 | 918,75 | 34,42* |
| *AB* | 2 | 116,67 | 58,33 | 2,19 |
| *AC* | 1 | 216,75 | 216,75 | 8,12* |
| *BC* | 2 | 50,00 | 25,00 | <1 |
| *ABC* | 2 | 146,00 | 73,00 | 2,73 |
| Hata | 36 | 961,00 | 26,69 | |
| Toplam | 47 | 4727,92 | | |

*$p < ,05$

Tablo 13.14a'nın alt kısmı etkileşim kareleri toplamını elde etmek için gerekli bütün hücre ortalamaları matrislerine sahiptir. Bu matrisler ilgilenilmeyen değişkenin düzeyleri boyunca ortalama alınarak elde edilmiştir. Böylece, *AB* özet tablosunun sol üst tarafı, *C*'nin düzeyleri önemsenmeksizin, $AB_{11}$ kombinasyonundan elde edilen bütün puanların toplamını içerir (örneğin; $ABC_{111} + ABC_{112}$). (*Not: Tablolardaki bütün değerleri iki ondalığa yuvarladım ama hesaplamalar daha fazla ondalıkla yapıldı. Yuvarlama hatalarına dikkat ediniz.*[6])

Table 13.14b kareler toplamının hesaplanmasını gösterir. Ana etkiler için kareler toplamı tam olarak tek-yönlü analizde olduğu gibi elde edilir. Birinci dereceden etkileşimlere ait hesaplamalar, iki yönlü analizde olduğu gibi aynı anda iki değişkenin ele alınmasıyla yapılır. Yeni olan sadece ikinci dereceden olan etkileşimin hesaplanmasıdır ve bu fark oldukça küçüktür. Burada $KT_{\text{hücreler}}$ üç boyutlu matristen elde edilecektir. Bu kareler toplamı tam-faktöriyel desendeki hücre ortalamaları arasındaki bütün değişkenliği yansıtır. Bizim bundan, bütün *ana* etkilerden ve birinci derecenden etkileşimlerden kaynaklı değişkenliği çıkarmamız gerekir. Geriye kalan üç değişkenin ortak etkileşiminden kalan ve $KT_{\text{ABC}}$ olarak adlandıracağımız değişkenliktir.

Son kalan kareler toplamı $KT_{\text{hata}}$'dır. Bunu elde etmenin en kolay yolu $KT_{\text{hücreler ABC}}$'yi $KT_{\text{toplam}}$'dan çıkarmaktır. $KT_{\text{hücreler ABC}}$ hücreler arasındaki farklılığa atfedilebilecek bütün değişkenliği temsil ettiği için ($KT_{\text{hücreler ABC}} = KT_{\text{A}} + KT_{\text{B}} + KT_{\text{C}} + KT_{\text{AB}} + KT_{\text{AC}} + KT_{\text{BC}} + KT_{\text{ABC}}$), bunun $KT_{\text{toplam}}$'dan farkı hücrelerin kendi içlerindeki değişkenliği verecektir.

Varyans analizinin özet tablosu Tablo 13.14c'te verilmiştir. Buradan üç temel etkinin, ve *A* x *C* ortak etkisinin manidar olduğunu görebiliriz. Diğerleri manidar değildir[7].

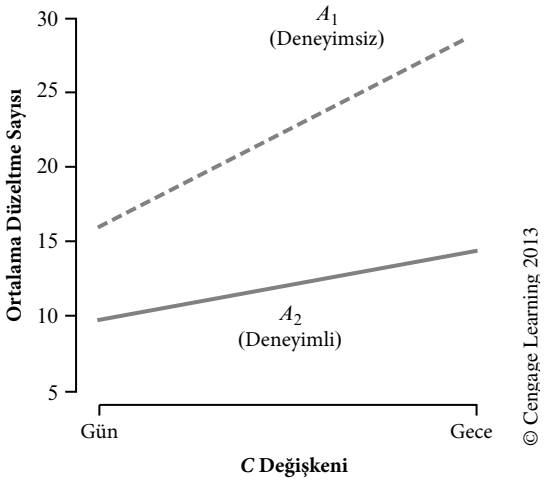
## Basit Etkiler

Manidar bir ortak etkimiz olduğundan *A* ve *C*'nin ana etkileri dikkatle yorumlanmalıdır. Burada Şekil 13.4'te yer alan *AC* etkileşim grafiğine bakalım. Veri göstermektedir ki deneyimsiz sürücülerin gece koşullarında gündüz koşullarına göre daha çok direksiyon düzeltmesi ihtiyacı olurken, deneyimli sürücünün her iki koşuldaki ihtiyaç duyduğu direksiyon düzeltmesi arasındaki fark izafî olarak azdır. *A*'nın manidar etkisine (deneyimli ve deneyimsiz sürücüler arasındaki fark) dair veri bize bir miktar güven verse de *C*'ye bağlı farklılıklar için bir miktar şüphe bırakmaktadır. Hızlıca bakarsak, *C*'nin deneyimsiz sürücüler için manidar bir etkisi olurken, deneyimli sürücüler için olmayabilir. Bu soruyu daha yakından sınamak için $A_1$ ve $A_2$ altında *C*'nin basit etkisini düşünmemiz gerekir. Bu analiz Tablo 13.15'te verilmiştir; buna göre gece ve gündüz arasında sadece deneyimsiz sürücüler için değil, deneyimli sürücüler için de fark vardır. (Dikkat ediniz, hesaplamalarımızın uygunluğunu tekrar kontrol edebiliriz; basit etki $KT_C + KT_{AC}$ olacaktır.)

---

6 Mevcut olan yuvarlama hatalarının birikimliliğinin en büyük nedeni ortalamalarla çalıştığınız zaman hesap makinelerinin kullandığı formüllerin toplamlarla çalışmasıdır. Ben bu bölümde tanımsal formülleri kullanıyorum çünkü onlar oldukça açık ama yerine koymamız gereken ortalamalar yuvarlama hatalarıyla gelmektedir. İyi hesaplama, yazılımları gelişmiş algoritmalar kullanarak yuvarlama hatalarını en aza indirmektedir.

7 Dikkat etmeniz gerekir ki bu varyans analizi yedi *F* değeri içermektedir, bu da yedi hipotez testi demektir. Çok sayıda hipotez testiyle deneysel hata oranı oldukça yüksek olacaktır. (Bu Tukey'nin "ailesel" kavramına kaymasının nedenlerinden biri olabilir çünkü, bir etkideki doğrusal farklılık kümesinin her biri aile olarak düşünülebilir. Birçok insan problemi görmezden gelmekte ve basitçe her karşılaştırmadaki her *F* değerini $\alpha = ,05$ hata oranıyla test etmektedir. Hata oranları hakkında endişeli iseniz Bonferroni veya çok basamaklı Bonferroni *t* yönteminden birinin dengini kullanmanız uygun olur. Bu genel olarak, her *F* değerine ilişkin olasılık değerine sahip olduğunuzda, bu değeri Bonferroni (veya çok basamaklı Bonferroni) istatistiğinin gerektirdiği olasılık değeri ile karşılaştırırsanız oldukça pratik olur. Buna benzer bir yaklaşıma ilişkin ilgi çekici bir örnek Rosenthal ve Rubin'de (1984) bulunabilir. Bir çok insanın *F*'yi kendi kendine değerlendirdiğinden ve aile-boyu hata oranı hakkında endişelenmediğinden şüpheleniyorum.




**Şekil 13.4** Tablo 13.14 için AC etkileşimi

Bu hipotetik deneyden, üç yol tipi arasında, deneyimli ve deneyimsiz sürücüler arasında manidar farklılıklar olduğuna karar veririz. Aynı zamanda hem deneyimli hem de deneyimsiz sürücüler için gündüz ve gece koşulları arasında manidar fark olduğu sonucuna da ulaşabiliriz.

**Tablo 13.15** Tablo 13.14'teki veri kümesi için basit etkiler

**(a) Veri**

| | $C_1$ | $C_2$ | Ortalama |
|---|---|---|---|
| $A_1$ | 16,000 | 29,000 | 22,500 |
| $A_2$ | 9,833 | 14,333 | 12,083 |

**(b) Hesaplamalar**

$$KT_{A_1 \text{ de } C} = nb\sum(\overline{X}_{1.k} - \overline{X}_{1..})^2$$
$$= 4 \times 3\left[(16,000 - 22,500)^2 + (29,000 - 22,500)^2\right] = 1014,00$$
$$KT_{A_2 \text{ de } C} = nb\sum(\overline{X}_{2.k} - \overline{X}_{2..})^2$$
$$= 4 \times 3\left[(9,833 - 12,083)^2 + (14,333 - 12,083)^2\right] = 121,50$$

**(c) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| $A_1$'de $C$ | 1 | 1014,00 | 1014,00 | 37,99* |
| $A_2$'de $C$ | 1 | 121,50 | 121,50 | 4,55* |
| Hata | 36 | 961,00 | 26,69 | |

*$p < ,05$

**(d) Kareler Toplamlarının Ayrıştırılması**

$$KT_{A_1 \text{ de } C} + KT_{A_2 \text{ de } C} = KT_C + KT_{AC}$$
$$1014,00 + 121,50 = 918,75 + 216,75$$
$$1135,50 = 1135,50$$

© Cengage Learning 2013

## Basit Etkileşim Etkileri

**basit ana etkiler**

**basit etkileşim etkisi**

Yüksek mertebeden faktöriyellerle, başka bir değişkenin düzeylerinde bir değişkenin etkisine bakabileceğimiz gibi (bizim basit etki olarak adlandırdığımız ama **basit ana etki** olarak adlandırmanın daha uygun olacağı), iki değişkenin etkileşimine üçüncü bir değişkenin düzeylerinde de bakabiliriz. Buna **basit etkileşim etkisi** diyeceğiz.

Temel olarak basit etkileşim etkileri, basit ortak etkileri elde ettiğimiz gibi elde edilir. Sadece değişkenin bir düzeyi için veriyi kullanırız. Eğer örneğimizdeki basit *AB* etkileşimlerine bakmak istersek, veriyi $C_1$ ve $C_2$ düzeyleri için alıp bunlara iki-yönlü analizmiş gibi davranabiliriz. Bir örnek üzerinde çalışmayacağım çünkü ne yapılacağı açıktır.

Yüksek mertebeden manidar etkilerin olmaması durumunda, basit etkileşim etkilerini sınamaktan kimse alıkonulamaz, fakat bunun mantıklı olabileceği çok ender durumlar vardır. Eğer deneycinin bunu yapması için özel bir nedeni varsa, örneğin *AB* etkileşimini *C*'nin düzeyleri için incelemek gibi, bunu yapmakta da serbesttir. Diğer taraftan, yüksek-mertebeden etkileşim manidar ise, deneyci düşük merteben etkileşimlere dikkatlice göz gezdirmelidir ve önemli temel etkileri test etmeyi düşünmelidir. Winer (1971, s. 442) şöyle der: "İstatistiksel olarak elde edilen iyi sonuçlar, bilimsel olarak manidar çıkarımlar yapabileceğiniz anlamına gelmez." Sağduyu en az istatistiksel manipülasyonlar kadar önemlidir.

# 13.13 Bir Bilgisayar Örneği

Örneğimiz farklı hücrelerdeki eşit olmayan katılımcılarla üç-yönlü faktör analizine örnek teşkil etmektedir. Seligman, Nolen-Hoeksema, Thornton ve Thornton (1990) çalışmasına dayanmasına rağmen veri sunidir ve bağımsız değişkenlerden birisi olay (Event) hayalîdir. Örneğin ana kararları rapor edilenle aynı doğrultıdadır. Bu örneğimizde, Bölüm 15'e gelene kadar SPSS veya diğer programların eşit olmayan örneklem büyüklüklerini nasıl ele aldığı ile ilgilenmeyeceğiz.

Çalışma kolej yüzme takımları ile ilgilidir. Bir takım antrenmanında, bütün katılımcılardan yüzebildiği kadar hızlı yüzmeleri istenmiş, fakat her seferinde rapor edilen zaman üzerinde beklenenden daha düşük performans sergilendiğini gösterecek şekilde oynama yapılmıştır. Böylelikle her yüzücü kötü sonuçları aldığında hayal kırıklığına uğramıştır. Yarım saat sonra, her yüzücüden aynı performansı göstermesi istenmiş ve zamanlar tekrar kaydedilmiştir. Yazarlar karamsar yüzücülerin ikinci denemede birinciden daha kötü yaptıklarını, iyimserlerin ise daha iyi yaptıklarını tahmin etmişlerdir.

Katılımcılar açıklayıcılık Stiline göre (iyimsere karşı kötümser), Cinsiyet ve tercih edilen Olay için sınıflandırılmıştır. Bağımlı değişken olarak $\text{Zaman}_1/\text{Zaman}_2$ kullanılmıştır. 1,00'dan daha büyük değerler yüzücünün ikinci denemede daha iyi yaptığı anlamına gelmektedir. Veri ve sonuçlar Tablo 13.16'da verilmiştir. Sonuçlar SPSS kullanılarak elde edilmiştir. Sonuçları incelerken SPSS'in ürettiği çıktının bir kısmının çok ender ihtiyaç duyulacak parçalardan oluştuğunu ve atlanabileceğini unutmayınız.

SPSS çıktısından stile göre manidar farklılıklar olduğu, başarısızlık geri bildirimi aldıktan sonra İyimserlerin biraz daha iyi performans sergilerken, kötümserlerin ikincide daha kötü performans sergiledikleri görülebilir. Ortalamadaki farklılıklar küçük görülebilir ama bu tip yarışların birbirine ne kadar yakın olduğunu düşünün, küçük bir farklılık bile önemlidir. Ayrıca İyimser x Cinsiyet etkileşiminin de olduğunu görebilirsiniz. Ortalamalara baktığımızda İyimser baylar ve bayanlar arasında neredeyse fark olmadığını görürüz ama bu kötümserler için doğru değildir. Bu verideki kötümser erkeklerin kayıp algısından bayanlara göre daha çok etkilendiği görülmektedir. Bu İyimserlik x Cinsiyet etkileşimi, izleyen özet tabloda bir bar grafiği olarak gösterilmiştir. Bu grafik Olay değişkeni üzerine konuşlandırılmıştır, çünkü bu değişkenin etkisi yoktur.[8]

8 Seligman ve arkadaşlarına (1990) adil olmak için bunun onların bulduğu sonuç olmadığını söylemeliyim.

**Tablo 13.17** Tablo 13.14'teki veri kümesi için basit etkiler

**(a) Veri**

| | İyimserler | | | | | | Kötümserler | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Erkek | | | Kadın | | | Erkek | | | Kadın | | |
| | Serbest | Göğüs | Sırt | Serbest | Göğüs | Sırt | Serbest | Göğüs | Sırt | Serbest | Göğüs | Sırt |
| | 0,986 | 1,026 | 1,009 | 1,108 | 1,048 | 1,004 | 0,983 | 0,962 | 0,936 | 0,997 | 1,045 | 1,045 |
| | 1,108 | 1,045 | 1,065 | 0,985 | 1,027 | 0,936 | 0,947 | 0,944 | 0,995 | 0,983 | 1,095 | 0,864 |
| | 1,080 | 0,996 | 1,053 | 1,001 | | 1,040 | 0,932 | 0,941 | 0,872 | 1,105 | 0,944 | 0,982 |
| | 0,952 | 0,923 | | 0,924 | | | 1,078 | 0,831 | | 1,116 | 1,039 | 0,915 |
| | 0,998 | 1,000 | | 0,968 | | | 0,914 | | | 0,997 | 0,927 | 1,047 |
| | 1,017 | 1,003 | | | | | 0,955 | | | 0,960 | 0,988 | |
| | 1,080 | 0,934 | | | | | | | | | 1,015 | |
| $\overline{X}$ | 1,032 | 0,990 | 1,042 | 0,997 | 1,038 | 0,993 | 0,968 | 0,920 | 0,934 | 1,026 | 1,008 | 0,971 |

**(b) SPSS'den Özet Tablo**

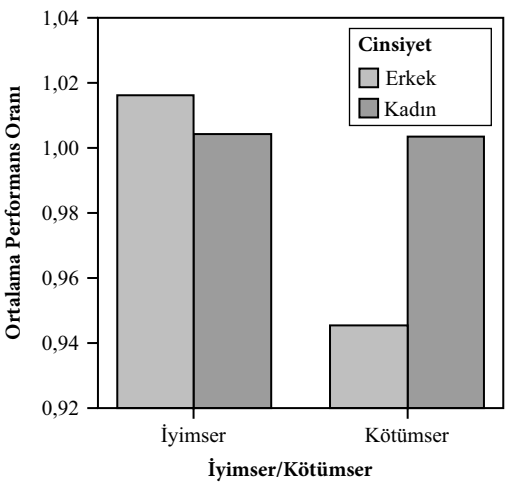
**Katılımcılar Arasındaki Etkilerin Testleri**

Bağımlı Değişken: ANS

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | F | Mnd. |
|---|---|---|---|---|---|
| Düzeltilmiş Model | 6,804E-02[a] | 11 | 6,186E-03 | 1,742 | ,094 |
| Düzeltilmiş Model | 48,779 | 1 | 48,779 | 13738,573 | ,000 |
| İYİMSER | 2,412E-02 | 1 | 2,412E-02 | 6,793 | ,012 |
| CİNS | 7,427E-03 | 1 | 7,427E-03 | 2,092 | ,155 |
| DARBE | 4,697E-03 | 2 | 2,348E-03 | ,661 | ,521 |
| İYİMSER * CİNS | 1,631E-02 | 1 | 1,631E-02 | 4,594 | ,037 |
| İYİMSER * DARBE | 5,612E-03 | 2 | 2,806E-03 | ,790 | ,460 |
| CİNS * DARBE | 1,142E-02 | 2 | 5,708E-03 | 1,608 | ,211 |
| İYİMSER * CİNS * DARBE | 1,716E-03 | 2 | 8,578E-04 | ,242 | ,786 |
| Hata | ,163 | 46 | 3,550E-03 | | |
| Toplam | 57,573 | 58 | | | |
| Düzeltilmiş Toplam | ,231 | 57 | | | |

[a] R Kare = ,294 (Düzeltilmiş R Kare = ,125)

**(c) Cinsiyet × İyimserlik Etkileşim Grafiği**




SPSS, Inc. çıktısından adapte edilmiştir.

## Anahtar Terimler

Faktörler (Giriş)
İki-yönlü faktöriyel desen (Giriş)
Faktöriyel desen (Giriş)
Tekrarlı ölçüm desenleri (Giriş)
Etkileşim (Giriş)
2 x 5 faktöriyel desen (Giriş)
Hücre (Giriş)
Ana Etki (13.1)
Basit Etki (13.1)
Koşullu Etki (13.1)
$KT_{\text{hücreler}}$ (13.1)
Çaprazlanmış (13.8)
Tesadüfî faktör (13.8)
İç içe (Yuvalanmış desen) (13.8)
Hiyerarşik modeller (13.8)
Karma modeller (13.8)
Çaprazlanmış deneysel desen (13.8)
Beklenen ortalama kareler (13.8)
Kısmî $\eta^2$ (13.9)
Kısmî etki (13.9)
Dengelenmemiş desen 3.9)
Ağırlıklandırılmamış ortalamalar (13.11)
Eşit ağırlandırılmış ortalamalar (13.11)
Birinci dereceden etkileşimler (13.12)
İkinci dereceden etkileşimler (13.12)
Basit ana etkiler (13.12)
Basit etkileşim etkileri (13.12)

## Alıştırmalar

Aşağıdaki problemlerin hepsi el ile çözülebilir, ama herhangi bir bilgisayar yazılım paketi de aynı sonuçları üretecektir.

13.1 Bir anne-bebek etkileşimi çalışmasında, anneler eğitilmiş gözlemciler tarafından bebekleriyle olan etkileşimlerinin kalitesine göre derecelendirilmiştir. Anneler ilk çocuk veya çok çocuk ve Düşük Doğum Ağırlıklı (DDA) ve Normal Doğum Ağırlıklı (NDA) olarak sınıflandırılmışlardır. Düşük doğum ağırlıklı bebeklerin anneleri daha sonra 18 yaşından küçük olup olmamalarına göre gruplandırılmıştır. Veri 12 puanlı ölçek üzerindendir ve yüksek puan daha iyi bebek anne etkileşimini göstermektedir. Uygun varyans analizini uygulayıp sonuçları yorumlayınız.

| İlk Doğum | | | Çok Doğum | | | İlk Doğum | | | Çok Doğum | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DDA <18 | DDA >18 | NDA | DDA <18 | DDA >18 | NDA | DDA <18 | DDA >18 | NDA | DDA <18 | DDA >18 | NDA |
| 4 | 6 | 8 | 3 | 7 | 9 | 7 | 6 | 2 | 7 | 2 | 10 |
| 6 | 5 | 7 | 4 | 8 | 8 | 4 | 2 | 5 | 1 | 1 | 9 |
| 5 | 5 | 7 | 3 | 8 | 9 | 5 | 6 | 8 | 4 | 9 | 8 |
| 3 | 4 | 6 | 3 | 9 | 9 | 4 | 5 | 7 | 4 | 9 | 7 |
| 3 | 9 | 7 | 6 | 8 | 3 | 4 | 5 | 7 | 4 | 8 | 10 |

© Cengage Learning 2013

13.2 Alıştırma 13.1 pratik bakış açısından büyük bir zayıflığa sahip olabilir. 18 yaşından küçük çok doğum anneleri grubuna dikkat ediniz. Veriden bağımsız olarak, bu grubun diğerleriyle aynı süreklilik üzerinde yer alacağına inanıyor musunuz?

13.3 Alıştırma 13.1'e atfen. Örneklem büyüklüklerinin yaş ve doğurganlık özelliklerinin evrendeki bağıl frekanslarını yansıtmadığı açık olarak görünmektedir. Bu, hangi koşullar altında ve hangi durumlar için uygun bir düşünce olur, hangi durumlar için olmaz?

13.4 Alıştırma 13.1'de yer alan çok doğum yapan annelerin üç grubunu basit etkileri kullanarak karşılaştırınız.

13.5 Hayvanlar, hafıza süreçleriyle ilgili bir çalışmada, bir-denemede kaçınmayı öğrenme-görevi açısından test edilmişlerdir. Hayvanlara bir test çizgisinin arkasına adım atmaları halinde korku üretecek bir uyarıcı verilmiştir. Bağımlı değişken çizginin arkasına geri adım atmaları için geçen zamandır. Hayvanlar elektrotun kortekslerinde bağlandığı bölgeye göre (Doğal taraf, A Bölgesi ve B Bölgesi) üçe ayrılmışlardır.

Daha sonra hayvanlar tekrar bölünmüş ve çizgiyi geçtikten 50, 100 ve 150 milisaniye sonra elektiriksel uyarıcı verilerek korku içeren uyarıcıya maruz bırakılmıştır.

Bunu bir noktayı örneklendirmek için ben yapılandırdım.

Eğer beyinde uyarılan alan hafızada yer ederse, uyarıcının hafıza pekişmesini engellemesi kaçınma tepkisinin öğrenilmesini geciktirmesi ve hayvanın çizgiyi tekrar geçmekte çekinmemesi beklenir. Çizgiyi geçmedeki gecikmeye ilişkin veri aşağıdaki gibidir:

Simülasyon Alanı

| Doğal Site | | | A Alanı | | | B Alanı | | |
|---|---|---|---|---|---|---|---|---|
| 50 | 100 | 150 | 50 | 100 | 150 | 50 | 100 | 150 |
| 25 | 30 | 28 | 11 | 31 | 23 | 23 | 18 | 28 |
| 30 | 25 | 31 | 18 | 20 | 28 | 30 | 24 | 21 |
| 28 | 27 | 26 | 26 | 22 | 35 | 18 | 9 | 30 |
| 40 | 35 | 20 | 15 | 23 | 27 | 28 | 16 | 30 |
| 20 | 23 | 35 | 14 | 19 | 21 | 23 | 13 | 23 |

© Cengage Learning 2013

Varyans analizi yapınız.

13.6 Alıştırma 13.5'te yer alan hücre ortalamalarının grafiğini çiziniz.

13.7 Alıştırma 13.5'te yer alan çalışma için (eğer A Alanı temsil etmek için kullanılsaydı) $\alpha_1$ neyi temsil ederdi?

13.8 Alıştırma 13.5'te yer alan Alan faktörünün sonuçlarını açık hale getirmek için basit etkileri kullanınız. Bu basit etkilerin doğru şekli oluşturduğunu gösteriniz.

13.9 Alıştırma 13.5'te yer alan doğal taraf ile diğer iki tarafı, uyarıcı uzunluğunu ihmal ederek ve Bonferroni testi kullanarak karşılaştırınız. (İpucu: Bölüm 11'de ve Bölüm 12'de ana hatları çizilmiş olan prosedürü izleyiniz. *n*'yi kullanırken ortalamanın elde edildiği puan sayısı olmasına dikkat ediniz.)

13.10 Alıştırma 13.5'te yer alan veri üzerinde basit etkileri kullanarak A alanına uyarıcıyı ertelemenin etkilerini test ediniz.

13.11 Bölüm 11'de yer alan Alıştırma 11.3a'yı referans alınız. Bir 2 x 2 faktöriyel form göreceksiniz. Faktöriyel varyans analizi yapınız ve sonuçları yorumlayınız.

13.12 Alıştırma 11.3'te Grup 1 ve Grup 3 birleşimini Grup 2 ve Grup 4 birleşimiyle karşılaştıracak bir test yapmıştınız. Alıştırma 13.11'deki işlemenin düzeylerinin temel etkisini test etmeyle o nasıl karşılaştırılır? Bir farklılık var mıdır?

13.13 İki ana etkisi olacak fakat etkileşimi olmayacak 2 x 2'lik bir desen için veri oluşturunuz.

13.14 Etkileşim etkisi olacak fakat ana etkisi olmayacak 2 x 2'lik bir desen için veri oluşturunuz.

13.15 Öncelikli ilginin etkileşim etkisinde olacağı uygun bir deney tanımlayınız.

13.16 Lemchuck, Bond ve Howell (1990) küçük ve büyük çocuklarda rol alma yeteneğini çocukların kreşe gidip gitmemesine göre incelemişlerdir. Bağımlı değişken ölçeklenmiş görev alma puanıdır. Örneklem büyüklükleri eşit değildir. Veri aşağıdaki gibidir:

| | Genç | | | | | Yaşlı | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gün. Bakımı Var | –0,139 | –2,002 | –1,631 | –2,173 | 0,179 | –0,167 | –0,285 | 0,851 | –0,397 |
| | –0,829 | –1,503 | 0,009 | –1,934 | –1,448 | 0,351 | –0,240 | 0,160 | –0,535 |
| | –1,470 | –1,545 | –0,137 | –2,302 | | –0,102 | 0,273 | 0,277 | 0,714 |
| Gün. Bakımı Yok | –1,412 | –0,681 | 0,638 | –0,222 | 0,668 | 0,859 | 0,782 | 0,851 | –0,158 |
| | –0,896 | –0,464 | –1,659 | –2,096 | 0,493 | | | | |

© Cengage Learning 201

SPSS kullanarak varyans analizi yapınız, uygun çıkarımları yapınız.

13.17 Alıştırma 13.1 için $\eta^2$ ve $\hat{\omega}^2$'yi hesaplayınız.

13.18 Alıştırma 13.1 için $\hat{d}$'yı hesaplayınız.

13.19 Alıştırma 13.5 için $\eta^2$ ve $\hat{\omega}^2$'yi hesaplayınız.

13.20 Alıştırma 13.5 için $\hat{d}$'yı hesaplayınız.

13.21 Koşullanma üzerine olan önceki deneyde, bir deneyci dört fare grubu oluşturmuştur. Bunlar (1) herhangi özel bir uyarıcı olmadığı durum, (2) tonlu uyarıcı (3) titreşimli uyarıcı ve (4) hem titreşimli hem tonlu uyarıcı verilen gruplar olarak belirlenmiştir. Daha sonra fareler ton veya titreşim koşullu uyarıcı, üç ayak şoku düzeyinden birisi de koşulsuz uyarıcı olacak şekilde klâsik koşullamayla koşullanmıştır. Bu 4 x 2 x 3 olacak şekilde faktöriyel bir desendir. Ham veri yerine hücre ortalamaları sunulmuştur. $KT_{total}$= 41.151,00 ve $n_{ijk}$ = 5'tir. Bağımlı değişken, önceden belirlenmiş bir ölçüte göre denemelerin sayısıdır.

| | Şartlı Uyarıcı | | | | | |
|---|---|---|---|---|---|---|
| | Ton | | | Titreşim | | |
| | Yüksek | Orta | Düşük | Yüksek | Orta | Düşük |
| Kontrol | 11 | 16 | 21 | 19 | 24 | 29 |
| Ton | 25 | 28 | 34 | 21 | 26 | 31 |
| Titreşim | 6 | 13 | 20 | 40 | 41 | 52 |
| Ton ve Titreşim | 22 | 30 | 30 | 35 | 38 | 48 |

© Cengage Learning 2013

Veriyi analiz ediniz ve sonuçları yorumlayınız.

13.22 İkinci Bölüm'de, bir test uyarıcısının öncül uyarıcı gösteriminin bir parçası olup olmadığını rapor edeceğimizde, Sternbergs'in deneyini göz önünde bulundurmuştuk. Bağımlı değişkenler gösterimdeki uyarıcı sayısı (1, 3 veya 5) ve test uyarıcısının gösterime dahil edilip edilmemesidir (Yes/No). Veriye RxTime.dat dosyasının yer aldığı (www.uvm.edu/~dhowell/methods8/DataFiles/Tab13-22.dat ) Web sitesinden ulaşılabilir. Bu iki-yönlü bir varyans analizidir. Analizi etki büyüklüğünü belirterek ve yorumlayarak yapınız, sonuçları yorumlayınız.

13.23 Herhangi bir istatistiksel paket programı kullanarak Web sitesinde yer alan Epineq.dat dosyası için Aralık ve Dozajı kapsayacak iki-yönlü varyans analizi yapınız. Buradan elde ettiğiniz sonuçları Bölüm 11'de Alıştırma 11.28-11.30'da elde ettiğiniz sonuçlarla karşılaştırınız.

13.24 Alıştırma 11.30'da dokuz hücreye ilişkin varyansları hesaplamıştınız. Bu cevapları Alıştırma 13.23'te yer alan $KO_{error}$ ile karşılaştırdığınızda ne dersiniz?

13.25 Alıştırma 13.23'teki varyans analizindeki Dozaj için Tukey testi yapınız. Sonuçları yorumlayınız.

13.26 Tablo 13.14'teki üç yönlü varyans analizinde kullanılan veri Web'de yer almaktadır. Adı Tab13-14.dat olarak geçmektedir. Her kayıttaki ilk üç giriş A (Deneyim), B (Yol) ve C (Koşullar) içindir. Dördüncü giriş ise bağımlı değişkendir. Tablo 13.14c'de yer alan tabloyu yeniden üretecek şekilde herhangi bir varyans analizi paketini kullanarak analiz yapınız.

13.27 Alıştırma 13.26'daki veri kümesini kullanarak Tablo 13.14'te yer alan basit etkileri yeniden elde ediniz.

13.28 Estetik ile ilgilenen bir psikolog klâsik ve romantik döneme ait bestecileri karşılaştırmak istemiştir. Her dönemden tesadüfî olarak dört besteci seçmiştir, çalışmalarından birini çalmış ve 5 denekten çalınanları derecelendirmesini istemiştir. Her besteci için farklı denekler kullanılmıştır. Veri aşağıdaki gibidir: (Bunun bir iç-içe desen olduğuna dikkat ediniz.) Uygun varyans analizini yapınız.

| | Klasik Dönem | | | | Romantik Dönem | | | |
|---|---|---|---|---|---|---|---|---|
| Besteci | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
| | 12 | 10 | 15 | 21 | 10 | 9 | 8 | 12 |
| | 14 | 9 | 18 | 17 | 11 | 12 | 7 | 14 |
| | 15 | 10 | 16 | 16 | 9 | 7 | 11 | 9 |
| | 11 | 12 | 18 | 18 | 8 | 15 | 12 | 7 |
| | 16 | 13 | 17 | 17 | 13 | 8 | 8 | 8 |

© Cengage Learning 2013

13.29 Bir eğitim araştırmacısı katı kıyafet kuralları uygulayan okulların yüksek akademik başarılı öğrenciler yetiştirdiği hipotezini test etmek istemektedir. Bunun için eyalette kıyafet kuralı uygulayan 7 okulu ve kıyafet kuralı uygulamayan 7 okulu tesadüfî olarak seçmiştir. Araştırmacı daha sonra her okuldan da tesadüfî 10 öğrenci seçmiş ve bu öğrencilerin standartlaştırılmış bir testte gösterdikleri başarıları kaydetmişlerdir. Sonuçlar aşağıdaki gibidir:

| | Kıyafet Kuralı Var | | | | | | | Kıyafet Kuralı Yok | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Okul | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| | 91 | 75 | 80 | 84 | 59 | 62 | 87 | 69 | 72 | 78 | 66 | 67 | 52 | 63 |
| | 78 | 73 | 77 | 92 | 67 | 93 | 78 | 74 | 56 | 77 | 55 | 82 | 71 | 65 |
| | 86 | 65 | 70 | 78 | 68 | 83 | 83 | 67 | 71 | 75 | 58 | 76 | 73 | 75 |
| | 70 | 68 | 68 | 78 | 64 | 78 | 79 | 64 | 92 | 56 | 73 | 78 | 68 | 82 |
| | 78 | 70 | 70 | 77 | 75 | 65 | 53 | 61 | 88 | 84 | 55 | 87 | 65 | 77 |
| | 48 | 60 | 69 | 76 | 74 | 71 | 66 | 76 | 64 | 83 | 70 | 87 | 69 | 81 |
| | 89 | 72 | 64 | 74 | 67 | 65 | 76 | 74 | 79 | 67 | 64 | 63 | 79 | 67 |
| | 90 | 77 | 73 | 81 | 56 | 85 | 67 | 71 | 73 | 70 | 52 | 68 | 67 | 73 |
| | 85 | 75 | 70 | 75 | 61 | 74 | 74 | 62 | 72 | 31 | 64 | 86 | 66 | 72 |
| | 82 | 80 | 74 | 81 | 67 | 83 | 72 | 67 | 70 | 70 | 79 | 84 | 64 | 56 |

© Cengage Learning 2013

Uygun varyans analizini yapınız ve sonucu açıklayınız. (Ç. N.)

13.30 Alıştırma 13.29'da yer alan analizi tekrar yapınız, fakat bütün değişkenlere sabitlenmiş ve çaprazlanmış olacak şekilde davranınız. Alıştırma 13-31'deki $KT_{\text{okul(kural)}}$'ın bu analizdeki $KT_{\text{okul}}$ ve $KT_{\text{okul*kural}}$ toplamından oluştuğunu gösteriniz. (*İpucu:* Eğer bunu SPSS kullanarak yaparsanız 1'den 7'ye kadar numaralanmış okul numaralarına sahip olmanız gerekir. )

13.31 Garlett ve Boss (2010) aynı cinsiyetten (lezbiyen) ve karşı-cinsiyetten ebeveynler tarafından yetiştirilen erkek ve kız çocuklarına ilişkin veriyi sunmuşlardır. Uzun süreli bir çalışmada aynı cinsiyetteki ebeveynlere ait 78 çocuk 17 yıl boyunca izlenmiş ve çocuklar 17 yaşındayken Achenbach'ın Çocuk Davranışı Kontrol Listesi kullanılarak veri toplanmıştır. Anchenbach'ın normatif çalışmasındaki tesadüfî örnekleme eşit örneklem kullanılarak ham veri elde edilmiştir. Daha rahat hareket edebilmek için her hücrede 43 çocuğun yer alacağı sayıltısıyla çalışılacaktır. Veri aşağıdaki gibidir.

| Grup | Aynı Cinsiyet Erkek | Aynı Cinsiyet Kadın | Karşıt Cinsiyet Erkek | Karşıt Cinsiyet Kadın |
|---|---|---|---|---|
| Ortalama | 25,8 | 26,3 | 23,0 | 20,3 |
| *sd* | 3,6 | 5,0 | 4,0 | 4,5 |
| *n* | 43 | 43 | 43 | 43 |

© Cengage Learning 2013

Bu veri için bir varyans analizi yapınız ve sonuçları yorumlayınız. (Yüksek puanlar daha çok yetkinlik göstergesidir.) (*F* değerleri araştırmayı yapan kişinin değerlerinden farklılaşabilir, çünkü araştırmayı yapan çok değişkenli varyans analizi kullanmıştır; fakat ortalamalar, araştırmacının ortalamalarıyla uyumludur. Verinin bu ortalamalarla ve varyanslarla olan bir SPSS versiyonu Web'de Ex13-13.sav dosyasında yer almaktadır.)

## Tartışma Soruları

13.32 Seligman ve arkadaşlarının (1990) analizindeki Tablo 13.15'te yer alan açıklayıcı stil üzerine olan veriyi ele alalım. İyimser grupta erkek sayısının kadın sayısından fazla olduğu, Kötümser grupta ise kadın sayısının erkek sayısından daha fazla olduğunu görebilirsiniz. Hangi koşullar altında bu eşit olmayan örneklem büyüklükleriyle başa çıkmak istersiniz ve bunu ne zaman görmezden gelebilirsiniz?

13.33 Araştırma literatüründe etkileşimlerden en az birinin manidar ve anlamlı olduğu üç yönlü faktöriyel bir desen örneği bulunuz. Daha sonra bu sonuçları yansıtacak bir veri oluşturunuz.



Bölüm 14

# Tekrarlı-Ölçümler Desenleri

## Hedefler

Aynı deneğin bir veya daha fazla değişkenin tüm düzeylerinde ölçüldüğü deneysel desenlerde varyans analizinin tartışılması.

## Konular


14.1 Yapısal Model
14.2 *F* Oranları
14.3 Kovaryans Matrisi
14.4 Varyans Analizinin Rahatlama Terapisine Uygulanması
14.5 Tekrarlı-Ölçümler Desenlerinde Karşılaştırmalar ve Etki Büyüklüğü
14.6 Sonuçların Raporlanması
14.7 Bir Denekler-Arası Değişken ve Bir Denekler-İçi Değişken
14.8 İki Denekler-Arası Değişken ve Bir Denekler-İçi Değişken
14.9 İki Denekler-İçi Değişken ve Bir Denekler-Arası Değişken
14.10 Sınıflar İçi Korelasyon
14.11 Tekrarlı Ölçüm Analizlerine Dair Diğer Düşünceler
14.12 Tekrarlı Ölçüm Desenleri İçin Karma Modeller

VARYANS ANALİZİ İLE İLGİLİ TARTIŞMAMIZDA, farklı hücrelerde farklı deneklerin olduğu, daha doğrusu hücrelerin bağımsız veya ilişkisiz olduğu desenlerle ilgilendik. (Varyans analizi varsayımlarında *bağımsız* ve *ilişkisiz* aynı anlamda kullanılmaktadır.) Bu ünitede hücrelerin bazılarının veya tümünün bağımsız olmadığı problemler ile ilgileneceğiz. Bu gibi desenler biraz daha zordur ve formülleri daha karmaşıktır. Okuyucuların büyük bir kısmı veya tümü SPSS veya SAS gibi programları kullanarak bu problemlere yaklaşacaktır. Ancak gördüklerinizi anlamak için, bilgisayar programı kullanmadan (el ile) bu problemlere nasıl yaklaşabileceğiniz hakkında birşeyler bilmeye ihtiyacınız vardır; bu durum pek çok formülü kullanmanıza yol açar. Bu formüllere kaygı duymadan yaklaşmanızı ve bunların hiçbirini ezberlemek zorunluluğu hissetmemenizi öneririm. Bu bölümün eksiksiz tamamlanması gerekir; bunun anlamı analizlere belli bir derinlikte girmemiz gereğidir, fakat her ihtiyacınızda formüllere geri dönebileceğinizi unutmayın ve ihtiyacınız oluncaya kadar hesaplamalar konusunda çok fazla endişelenmeyin.

Farklı uygulamalar altında farklı deneklerin olduğu tipik bir tek-yönlü varyans analizi düşündüğünüzde, işlem 1 ve 2, 1 ve 3, 2 ve 3 arasındaki ilişkinin sıfır olması beklentisini muhtemelen kabul etmek istersiniz.

| İşlem 1 | İşlem 2 | İşlem 3 |
|---|---|---|
| $X_{11}$ | $X_{21}$ | $X_{31}$ |
| $X_{12}$ | $X_{22}$ | $X_{32}$ |
| ... | ... | ... |
| $X_{1n}$ | $X_{2n}$ | $X_{3n}$ |

Ancak, burada yukarıdaki desende aynı denekleri her üç işlemde de kullandığımızı varsayın. Dolayısıyla bir defa ölçülen $3n$ denek yerine 3 defa ölçülen $n$ deneğimiz olur. Bu durumda üç işlem arasındaki ilişkinin sıfır olması beklentisine inanmamız oldukça zordur. Aksine işlem 1'de başarılı olan denekler muhtemelen işlem 2 ve 3'te de başarılı olacak veya işlem 1'de başarısız olanlar diğer işlemlerde de başarısız olacaktır, ki bu durum bizi işlemler arasında manidar bir ilişkinin olduğu sonucuna ulaştırır.

**ayırma**

Eğer ayrıştırma veya **ayırma** yapamasaydık ve aynı deneklerle yapılan tekrarlı ölçümlerden kaynaklanan bağımlılığı ortadan kaldırmasaydık, işlemlerin birbirinden bağımsız olmaması, ciddî problemlere yol açardı. (Böylelikle bağımlılığa neden olan etkileri *kontrol altına* almış oluruz - bu terim Ünite 15'te daha çok kullanılacaktır.) Aslında **tekrarlı ölçüm desenleri**nin avantajlarından biri, tüm işlemlerde aynı deneklerin kullanılması sonucunda değişkenliği azaltması ve aynı zamanda hata terimimizden denek farklılıklarını ortadan kaldırmamıza izin vermesidir. Böylece hata bileşeninin işlemden işleme veya hücreden hücreye bağımsız olmamasını sağlar.

**tekrarlı ölçüm desenleri**

Bir örnek olarak, dört denek üzerinde yapılmış üç işlemden elde edilen Tablo 14.1'deki oldukça abartılmış veri kümesini göz önüne alalım. Burada bağımlı değişken, deneğe verilmiş olan bir görevde belli bir ölçüte göre yapılan deneme sayısıdır. Öncelikle işlem ortalamalarına baktığınızda küçük farklar göreceksiniz, fakat bunlar çok fazla heyecanlandırıcı değildir. Her bir işlem içinde o kadar çok değişkenlik vardır ki ilk bakışta ortalamalar arasındaki farkın kaynağı yalnızca şansa bağlıymış gibi görünür. Fakat deneklerin ortalamaları incelediğinde birinci deneğin bütün koşullar altında çabuk öğrendiği, üçüncü ve dördüncü deneklerin ise oldukça yavaş öğrendikleri görülmektedir. Denekler arasındaki bu farklılıklar, işlemler *içindeki* farklılıkların çoğunu oluşturur ve işlem etkisi ile ilgisi yoktur. Eğer bu denek etkilerini ortadan kaldırabilirsek daha iyi (daha küçük) hata kestirimleri elde edebiliriz. Aynı zamanda, işlemler arası korelasyonların yüksek ve pozitif çıkmasının

**Tablo 14.1** Basit tekrarlı ölçüm deseni için kurgusal veri

| | İşlem | | | |
|---|---|---|---|---|
| Denek | 1 | 2 | 3 | Ort. |
| 1 | 2 | 4 | 7 | 4,33 |
| 2 | 10 | 12 | 13 | 11,67 |
| 3 | 22 | 29 | 30 | 27,00 |
| 4 | 30 | 31 | 34 | 31,67 |
| **Ortalama** | 16 | 19 | 21 | 18,67 |

© Cengage Learning 2013

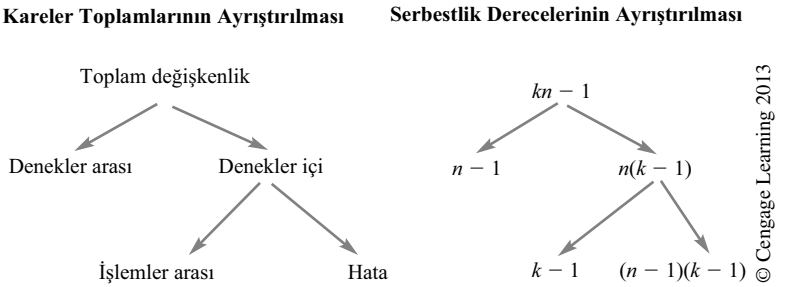
nedeni de denekler arasındaki farklardır. Denekler için ayrı bir terim oluşturarak bunları da kontrol altına alabiliriz.

Bunu yapmanın zahmetli bir yolu, bir deneğe ait puanların arasındaki ilişkiyi bozmadan denek ortalamalarını eşitleyerek, bütün deneklerin katkılarını ortak bir tabanda toplamaktır. Böylece, *i*'yinci deneğin ortalaması $\overline{X}_i$ olmak üzere $X'_{ij} = X_{ij} - \overline{X}_i$ ifadesiyle tanımlayabiliriz. Böylece tüm deneklerin ortalaması aynı olur $(\overline{X}'_i = 0)$ ve puanlar arasında gözlenen farklar sadece hataya ya da uygulamaya dayandırılabilir. Bu yaklaşım işe yaramakla birlikte pratik değildir. Alternatif ve daha kolay bir yöntem ise başlangıçta denekler arası kareler toplamını ($\boldsymbol{KT_{\text{denekler arası}}}$ veya $KT_{\text{D}}$ olarak gösterilir) hesaplayıp, $KT_{\text{toplam}}$'dan çıkarmaktır. Kitapta kullanılacak yöntem temelde budur, bunun ilk yöntem ile cebirsel olarak denk olduğu gösterilebilir.

$KT_{\text{denekler arası}}$

Çözüm, Şekil 14.1'de diyagramla temsil edilmiştir. Burada toplam varyans denekler arası ve denekler içi varyans olarak ikiye ayrılmıştır. Aynı ayrışım serbestlik dereceleri için de yapılır. Deneğin kendi içindeki varyansının bir kısmı farklı işlemlerden aldığı puanlardan, bir kısmı da hatadan kaynaklanır. Varyanstaki bu ikinci ayrıştırma şeklin üçüncü satırında gösterilmiştir. Bir tekrarlı ölçüm analizi her zaman birinci ayrıştırma, yani $KT_{\text{toplam}}$'ın $KT_{\text{denekler arası}}$ ve $KT_{\text{denekler içi}}$ olarak ayrılması olarak düşünülecektir. Desenin karmaşıklığına bağlı olarak, bu ayrıştırmalardan biri veya her ikisi tekrar ayrılabilir.

Bu kısımda tartışılan tekrarlı ölçüm desenleri, alanın keşfinin sadece başlangıcını kapsar. Tarihî nedenlerden dolayı, istatistik literatürü bu desenlerin önemi üzerinde gerektiği kadar durmamıştır. Bu sebeple çoğunlukla davranış bilimciler, özellikle de psikologlar tarafından geliştirilmişlerdir. Bu desenlerin en kapsamlı bir şekilde ele alındığı çalışma Winer, Brown ve Michels'a (1991) aittir. Bu çalışmada tekrarlı ölçme desenlerine ayrıntılı ve mükemmel bir şekilde yaklaşılmıştır; bu bölümün büyük bir kısmında da Winer'ın çalışmalarının etkisi vardır.




**Şekil 14.1** Kareler toplamlarının ve serbestlik derecelerinin ayrıştırılması

## 14.1 Yapısal Model

Bazı modeller beni mutlu eder. Tablo 14.1'de gösterilen analizin temelini iki yapısal model oluşturur. Bu modellerden basit olanı

$$X_{ij} = \mu + \pi_i + \tau_j + e_{ij}$$

ifadesiyle yazılır. Bu eşitlikte;

$\mu$ = genel ortalama,

$\pi_i$ = $i$'yinci kişi veya denekle ilgili sabit sayı, bu sayı bu kişinin ortalama bir kişiden ne kadar farklı olduğunu gösterir.

$\tau_j$ = $j$'yinci işlemle ilgili sabit bir sayı; bu sayı, ilgili işlemin ortalamasının, tüm işlemlerin ortalamasından ne kadar farklı olduğunu gösterir.

$e_{ij}$ = $i$'yinci deneğe ait $j$'yinci işlemle ilgili deneysel hatadır.

$\pi_i$ ve $e_{ij}$ değişkenlerinin işlemlerin her birinde sıfır etrafında birbirinden bağımsız ve normal dağıldığı ve bunların varyanslarının, $\sigma_\pi^2$ ve $\sigma_e^2$, işlemlerde homojen olduğu varsayılır. Bu varsayımlar ile Tablo 14.2 Model I'de gösterilen beklenen kareler ortalamaları elde edilebilir. (Beklenen kareler ortalamalarını göstermek için önceki bölümlerde geliştirilen simgeyi kullanıyorum. Hata terimi ve denek faktörü tesadüfî olarak alındı ve bu varyanslar $\sigma_\pi^2$ ve $\sigma_e^2$ olarak gösterildi. [Deneklerin her zaman tesadüfî olarak seçildiği düşünülür.] Ancak, işlem faktörü genellikle sabit faktördür ve onun değişkenliği $\theta_\tau^2$ ile gösterilir.)

Diğer bir alternatif model ve belki de daha gerçekçi olanı aşağıdaki ifadeyle yazılır:

$$X_{ij} = \mu + \pi_i + \tau_j + \pi\tau_{ij} + e_{ij}$$

Burada modele farklı deneklerin uygulamalar boyunca farklı biçimde değişmelerine imkân sağlayan Denek x İşlem etkileşimi terimini ekledik. İlk model için geçerli olan varsayımların bu model için de geçerli olmasıyla birlikte, burada ayrıca, $\pi\tau_{ij}$ değişkeninin sıfırın etrafında modelin diğer elemanlarından bağımsız olarak dağıldığı varsayılır. Bu model Tablo 14.2 Model II'de gösterilen beklenen kareler ortalamalarını ortaya çıkarır.

İstatistiksel teori çerçevesinde bir grup ortalamanın karşılaştırılması gibi pratik bir sorunun çözümü yerine, bu iki model ve onların beklenen kareler ortalamalarıyla ilgili tartışma yapılması asıl sorudan uzaklaştığımız izlenimini verebilir. Ancak, testlerin nereden geldiğini ve analizin nasıl gerçekleşeceğini tanımlamak çok önemlidir. Bir süre daha bana karşı sabırlı olunuz.

## 14.2 *F* Oranları

Tablo 14.2'deki beklenen kareler ortalamaları, ele aldığımız modeldeki $F$ oranlarını etkiler. Eğer Denek x İşlem etkileşimin olmadığı varsayılırsa aşağıdaki oranlar oluşturulur:

$$\frac{E(KO_{\text{denekler arası}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + k\sigma_\pi^2}{\sigma_e^2}$$

**Tablo 14.2** Basit tekrarlı ölçümler modeli için beklenen kareler ortalamaları

| Model I | | Model II | |
|---|---|---|---|
| $X_{ij} = \mu + \pi_i + \tau_j + e_{ij}$ | | $X_{ij} = \mu + \pi_i + \tau_j + \pi\tau_{ij} + e_{ij}$ | |
| Kaynak | E(KO) | Kaynak | E(KO) |
| Denekler | $\sigma_e^2 + k\sigma_\pi^2$ | Denekler | $\sigma_e^2 + k\sigma_\pi^2$ |
| İşlemler | $\sigma_e^2 + n\theta_\tau^2$ | İşlemler | $\sigma_e^2 + \sigma_{\pi\tau}^2 + n\theta_\tau^2$ |
| Hata | $\sigma_e^2$ | Hata | $\sigma_e^2 + \sigma_{\pi\tau}^2$ |

SPSS çıktısından adapte edilmiştir.

ve

$$\frac{E(KO_{\text{islem}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + n\sigma_\pi^2}{\sigma_e^2}$$

Küresellik ile ilgili ek bir varsayım verildiğinde, ki bu varsayım bir sonraki kesimde tartışılacaktır, bütün bunlar yokluk hipotezinin test edilmesinde işe koşulacak güvenilir $F$ oranlarının bulunmasına yol açar.

Fakat genellikle, Denek x İşlem etkileşiminin olmadığını varsaymak konusunda temkinli davranırız. Araştırmalarımızın çoğunda, farklı deneklerin farklı işlemlere özellikle bu "işlemler" sürmekte olan deneyin evrelerine karşılık geliyorsa, farklı cevaplar verdiğini varsaymak daha akla yakındır. Sonuç olarak, genellikle daha eksiksiz bir model ile çalışmayı tercih ederiz.

Etkileşim terimini de içeren kapsamlı model aşağıdaki oranları içerir:

$$\frac{E(KO_{\text{denekler arası}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + k\sigma_\pi^2}{\sigma_e^2 + \sigma_{\pi\tau}^2}$$

ve

$$\frac{E(KO_{\text{islem}})}{E(KO_{\text{hata}})} = \frac{\sigma_e^2 + \sigma_{\pi\tau}^2 + n\theta_\tau^2}{\sigma_e^2 + \sigma_{\pi\tau}^2}$$

İşlemler için ilk oluşturulan $F$ oranı uygun olmasına rağmen (pay, paydada olmayan sadece bir terim içerir), denekler için oluşturulan $F$ yanlıdır. Bir sonraki oranı oluşturmasaydık ve manidar bir $F$ elde etmeseydik, denek farklılıklarının gerçekten var olduğuna tamamıyla inanabilirdik. Ancak, $F$ manidar olmasaydı, yorumumuz çok belirsiz olacaktı. Manidar olmayan $F$, $k\sigma_\pi^2 = 0$ veya $k\sigma_\pi^2 > 0$ olması anlamına gelir, fakat $\leq \sigma_{\pi\tau}^2$. Genellikle ikinci model tercih edildiğinden ve belirsizlikten hoşlanmadığımızdan, Deneklerden kaynaklanan etkileri nadiren test ederiz. Ancak Denek etkisini test ettiğimizde olan kazancımız çok az olduğundan bu çok büyük bir kayıp değildir. İlk aşamada $KT_{\text{denekler arası}}$ elde etmemizin başlıca nedeni, işlemler arasındaki korelasyonun etkisini azaltmak ve böylece hata teriminden denek farklılıklarını çıkarmaktır. Denek etkisi ile ilgili test yaptığımızda, testin manidar çıkması sadece bu bireylerin farklı olduğunu gösterir ve bu önemli bir bulgu değildir. Her iki modelde de vurgulanan ve önemli olan işlemlerin etkisini test etmek için payda olarak $KO_{\text{hata}}$'yı kullanabilmemizdir.

## 14.3 Kovaryans Matrisi

Tekrarlı-ölçümler desenlerindeki herhangi bir $F$ oranının merkezî $F$ olarak dağılması için gerekli olan önemli bir varsayım, kovaryans matrisinin birleşik simetriye[1] sahip olmasıdır. Bunun ne anlama geldiğini anlamak için Tablo 14.1'de gösterilen veri için, üç işlem arasındaki kovaryans matrisini temsil eden matrisi ($\hat{\Sigma}$) dikkate almak gerekir.

$$\hat{\Sigma} = \begin{array}{c|ccc} & A_1 & A_2 & A_3 \\ \hline A_1 & 154{,}67 & 160{,}00 & 160{,}00 \\ A_2 & 160{,}00 & 176{,}67 & 170{,}67 \\ A_3 & 160{,}00 & 170{,}67 & 170{,}00 \end{array}$$

[1] Bu varsayım oldukça katıdır ve bir miktar esneyecektir. Yine de yeterli bir varsayımdır ve sıklıkla yapılır.

**esas köşegen**

**köşegen-dışı elemanlar**

**birleşik simetri**

**küresellik**

Matrisin **esas köşegen**inde her bir işlem için varyanslar vardır ($\hat{\theta}_A^2$). Bunlar birbirlerine oldukça yakın değerlerdir ve varyansların homojen olduğu varsayımının karşılandığını gösterir. **Köşegen-dışı eleman**lar işlemler arasındaki kovaryansları gösterir ($kov_{12}$, $kov_{13}$ ve $kov_{23}$). Bunlar da oldukça birbirine yakın değerlerdir. (Varyanslar ilgisiz olduğundan, aynı büyüklükte olmaları, sadece işlemler arasındaki oldukça yüksek korelasyonun olduğuna işaret eder.) Köşegen üzerindeki sabit varyansların ve köşegen-dışı sabit kovaryansların örüntüsü *evrendeki* **birleşik simetri**yi gösterir. (Yine varyanslar ve kovaryanslar arasında ilişki yoktur.) Kovaryans matrisinin ($\sum$) birleşik simetri varsayımı ki burada $\hat{\sum}$ simgesi birleşik simetrinin değerini gösterir, tekrarlı ölçümler varyans analizindeki yeterli şartı temsil eder. En genel durum **küresellik** olarak bilinir ve bu kapsamlı varsayım kaynaklarda sıklıkla görülür. Eğer birleşik simetri varsa küresellik varsayımı ile karşılaşılır, ancak pratikte olmasa da birleşik simetri olmadan küreselliğin olması muhtemel bir durumdur. Eski kitaplar çok katı bir varsayım olmasına rağmen, genellikle birleşik simetriyi kullanırlar. Son yıllarda "küreselliği" kullanma eğilimi vardır ve bu bölümün sonlarına doğru karmaşık modellerle ilgilenildiğinde, birleşik simetriyi kullanmamıza rağmen biz de genellikle "küreselliği" kullanacağız. Küresellik varsayımı olmadan $F$ oranları, $F$ tablosundaki dağılıma sahip olamaz. Bu varsayım her türlü varyans analizi desenine uygulanabilir olsa da, hücreler bağımsız olduğunda kovaryans daima sıfırdır ve sorun yoktur - sadece varyans homojenliği varsayımına ihtiyacımız olur. Ancak tekrarlı ölçümler desenlerinde kovaryanslar sıfıra eşit olmaz ve tümünün eşit olduğunu varsaymamız gerekir. Bu durum bazı kişilerin (ör. Hays, 1981) tekrarlı ölçümler desenlerinde önemli hususları dahil etmemesine sebep olmuştur. Ancak küreselliği ele aldığımızda, $F$ değerleri geçerlidir; almadığımızda ise ya yakınlaştırma süreçleri (bu bölümde daha sonra tartışacağız) veya $\sum$ ile ilgili varsayımlara dayalı olmayan alternatif yöntemler kullanılır. Kovaryans matrisiyle ilgili varsayım gerektirmeyen bir alternatif yöntem, **çok değişkenli varyans analizi**dir **(MANOVA).** Çoklu varyans analizi **çok değişkenli yöntem**dir. Bu yöntem çoklu bağımlı değişkenleri aynı anda ele alır. Bu yöntemde veri kümesinin tamamı gereklidir ve Kesim 14.12'de sözü edilen karmaşık modeller analizleri yerine kullanılır.

**çok değişkenli varyans analizi (MANOVA)**

**çok değişkenli yöntem**

Basit sezgiyle anlaşılabilen bir anlamı olmadığından, çoğu kişi, kovaryans yönünden düşünürken sorun yaşar. Korelasyon ile ilgili olduğunu düşünsek çok şey kaybetmiş olmayız. Eğer gerçekten varyans homojenliği varsa, birleşik simetri, denemeler arasındaki sabit korelasyonları düşürür.

## 14.4 Varyans Analizinin Rahatlama Terapisine Uygulanması

Tekrarlı ölçüm desenlerine basit bir örnek verebilmek için rahatlama tekniklerinin migrene bağlı baş ağrısını kontrol altına almadaki etkisini inceleyen bir çalışmayı ele alalım. Burada tanımlanan veri kümesi hayalîdir, ancak benzer fakat daha karmaşık bir çalışma yapan Blanchard, Theobald, Williamson, Silver ve Brown'ın (1978) verisi ile uyumludur.

Bu araştırmada migren ağrısı çeken 9 kişiden migren ağrılarının sıklığını ve süresini raporlamaları istenmiştir. Herhangi bir eğitimin verilmediği başlangıç düzeyinin ölçüldüğü 4 haftalık bir süre sonunda, 6-haftalık rahatlama eğitimi verilmiştir. (Her bir denek programa değişik zamanlarda katılmıştır, böylece hava değişimi veya tatil gibi durumlar veriyi etkilememiştir.) Örneğimizde başlangıç düzeyinin son iki haftalık süresini ve eğitimin son 3 haftasını analiz edeceğiz. Bağımlı değişken, bu beş haftalık dönemdeki baş ağrısı süresidir (saat/hafta). Veri ve hesaplamalar Tablo 14.3'te görülmektedir.[2] Ortalamaları,

[2] Ortalamaları virgülden sonra üç basamak olacak şekilde yuvarladığımdan, cevaplarda yuvarlama hatası vardır. Buradaki cevaplar daha çok ondalık basamağı olan sayılar içermektedir.

**Tablo 14.3** Migren baş ağrılarıyla ilgili veri analizi

**(a) Data**

| | Başlangıç Düzeyi | | Eğitim | | | Denek |
|---|---|---|---|---|---|---|
| Denek | 1. Hafta | 2. Hafta | 3. Hafta | 4. Hafta | 5. Hafta | Ort. |
| 1 | 21 | 22 | 8 | 6 | 6 | 12,6 |
| 2 | 20 | 19 | 10 | 4 | 4 | 11,4 |
| 3 | 17 | 15 | 5 | 4 | 5 | 9,2 |
| 4 | 25 | 30 | 13 | 12 | 17 | 19,4 |
| 5 | 30 | 27 | 13 | 8 | 6 | 16,8 |
| 6 | 19 | 27 | 8 | 7 | 4 | 13,0 |
| 7 | 26 | 16 | 5 | 2 | 5 | 10,8 |
| 8 | 17 | 18 | 8 | 1 | 5 | 9,8 |
| 9 | 26 | 24 | 14 | 8 | 9 | 16,2 |
| Hafta Ortalamaları | 22,333 | 22,000 | 9,333 | 5,778 | 6,778 | 13,244 |

**(b) Hesaplamalar**

$$KT_{\text{toplam}} = \sum (\text{X} - \overline{\text{X}}..)^2 = (21 - 13{,}244)^2 + ... + (9 - 13{,}244)^2 = 3166{,}31$$

$$KT_{\text{denekler}} = \text{H}\sum (\overline{\text{X}}_{\text{D}} - \overline{\text{X}}..)^2 = 5\left[(12{,}6 - 13{,}244)^2 + ... + (6{,}778 - 13{,}244)^2\right] = 486{,}71$$

$$KT_{\text{haftalar}} = \text{n}\sum (\overline{\text{X}}_{\text{H}} - \overline{\text{X}}..)^2 = 9\left[(22{,}333 - 13{,}244)^2 + ... + (6{,}778 - 13{,}244)^2\right] = 2449{,}20$$

$$KT_{\text{hata}} = KT_{\text{toplam}} - KT_{\text{denekler}} - KT_{\text{haftalar}} = 3166{,}31 - 486{,}71 - 2449{,}20 = 230{,}40$$

**(c) Özet tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 8 | 486,71 | | |
| Denekler içi | 36 | 2679,60 | | |
| Haftalar | 4 | 2449,20 | 612,30 | 85,04* |
| Hata | 32 | 230,40 | 7,20 | |
| Toplam | 44 | 3166,31 | | |

* $p < ,05$

© Cengage Learning 2013

değişkenlerin isimlerini alt simge ile göstererek tanımladığımı unutmayın. Böylece standart "nokta işareti" yerine (örneğin, hafta ortalamalarını belirtmek için $\overline{X}_i$ kullanılması), harfler kullanarak değişken isimlerini alt simge ile belirttim (örneğin; Hafta ortalamaları $\overline{X}_H$). Deneklerin ortalamaları $\overline{X}_D$ ve genel ortalama alışılmış olduğu şekliyle $\overline{X}..$ sembolüyle gösterildi, bu sembollerdeki $X$ bireysel gözlemleri temsil eder.

İlk olarak Tablo 14.3a'da bulunan veriye bakınız. Dikkat edilirse, oldukça fazla bir değişkenlik var, ancak bu değişkenliğin çoğu bireylerin bir kısmının diğerlerine göre daha fazla veya daha az süre baş ağrısı çekmesinden kaynaklanıyor ki bu durum tedavi programı ile oldukça az ilişkilidir. Belirttiğim gibi, tekrarlı ölçümler desenlerinde yapabildiğimiz, fakat denekler-arası desenlerde yapamadığımız şey, bu değişkenliği $KT_{\text{hata}}$'dan çıkararak daha küçük bir $KO_{\text{hata}}$ elde edilmesidir.

Tablo 14.3b'den de görebileceğiniz gibi, $KT_{\text{toplam}}$ her zamanki gibi hesaplanmıştır. Benzer şekilde, $KT_{\text{denekler}}$ ve $KT_{\text{haftalar}}$ da her zaman ana etkilerde olduğu gibi hesaplanmıştır (genel ortalamadan sapmaların kareleri toplamının uygun bir sabit sayı ile çarpımı [örneğin her bir ortalamayı oluşturan gözlem sayısı]). Son olarak hata terimi $KT_{\text{toplam}}$'dan $KT_{\text{denekler}}$ ve $KT_{\text{haftalar}}$'ın çıkarılmasıyla bulunur.

Özet tablo Tablo 14.3c'de gösterilmiştir, ancak bu tablo daha önce gördüklerinizden biraz farklıdır. Bu tabloda, Denkler-Arası ve Denekler-İçi faktörler planlı bir şekilde ayrılmıştır. Haftalar ve hata için terimler Denekler-İçi teriminin bir parçasıdır ve gerçekten de altında yer alır. (Bu desende Denekler-Arası faktörün altında hiçbir şey olmamasının nedeni daha fazla parçalara ayrılamamasıdır. Fakat birkaç sayfa bekleyin ve bunun da olacağını göreceksiniz.) Daha önce belirtilen nedenler yüzünden denekler için değil haftalar için bir $F$ değeri hesapladım. Haftalar için hesaplanan $F$ değeri, 4 ve 32 serbeslik derecelerine dayalı olarak hesaplanmış ve $F_{,05}(4, 32) = 2{,}68$ bulunmuştur. Sonuçta $H_0$: $\mu_1 = \mu_2 = \ldots\ldots = \mu_5$ hipotezi ret edilmiş ve "rahatlama programı, denekler tarafından raporlanmış her bir haftaya düşen ağrı süresinde bir azalmaya sebep olmuştur" kararına ulaşılmıştır. Tablo 14.3'teki ortalamalar incelendiğinde, son üç haftalık tedavi süresindeki her bir haftaya düşen baş ağrısı süresi miktarının başlangıç düzeyine göre neredeyse üçte-biri olduğu sonucuna ulaşılır.

Özet tabloda Denek × Haftalar etkileşiminin olmadığını fark etmişsinizdir. Her hücrede tek puan olmak üzere, etkileşim terimi hata terimidir ve hatta bazı kişiler onu hata yerine Denek x Haftalar olarak tanımlamayı tercih eder. Daha farklı bir biçimde ifade etmek gerekirse, burada tartışılan desende, tamamiyle karışmış olduğundan, hatayı herhangi Denek x Haftalar etkileşiminden ayırmak imkânsızdır. Yapısal modeller ile ilgili konumuzda gördüğümüz gibi, bu etkilerin her ikisi, varsa, hatanın beklenen kareler ortalaması içinde birleşmiş bir halde bulunur.

Küresellikten veya birleşik simetri varsayımından daha önce söz ettim. Örnekteki veri için, varyans-kovaryans matrisi aşağıda yer almaktadır ve $\hat{\sum}$ ile gösterilir. Buradaki (^) sembolü, bu değerin evrene ait varyans-kovaryans matrisinin $\sum$ kestirimi olduğunu belirtir.

$$\hat{\sum} = \begin{matrix} 21{,}000 & 11{,}750 & 9{,}250 & 7{,}833 & 7{,}333 \\ 11{,}750 & 28{,}500 & 13{,}750 & 16{,}375 & 13{,}375 \\ 9{,}250 & 13{,}750 & 11{,}500 & 8{,}583 & 8{,}208 \\ 7{,}833 & 16{,}375 & 8{,}583 & 11{,}694 & 10{,}819 \\ 7{,}333 & 13{,}375 & 8{,}208 & 10{,}819 & 16{,}945 \end{matrix}$$

Bu matrisin görsel olarak incelenmesi sonucunda, küresellik varsayımının mantıklı olduğu sonucuna ulaşılır. Esas köşegende bulunan varyanslar 11,5 ile 28,5 ve esas köşegenin dışındaki kovaryanslar 7,333 ile 16,375 arasında değişmektedir. Sadece 9 deneğimiz olduğunu düşünürsek bu değerler kabul edilebilir sabitlik düzeyini gösterir. (Varyansların kovaryanslara eşit olmak zorunda olmadığını unutmayınız; aslında çok ender eşit olurlar.) Küreselliğin bu varsayımı ile ilgili istatistiksel bir test Mauchly (1940) tarafından geliştirilmiş ve Winer'ın çalışmasında (1971, s. 596) verilmiştir. Bu test küresellik hipotezini reddetmek için nedenimiz olmadığını gösterir. Popüler bir test olmamasına rağmen –çoğu istatistikçi bu testin normallik varsayımına dayandığına ve gücünün düşük olduğuna işaret eder– çoğu istatistiksel yazılım bu testin sonucunu düzenli olarak verir. Mauchly'nin testini göz ardı etmeyi ve Bölüm14.7'de tartışılan alternatifleri tercih etmeyi öneririm.

Önceden de belirttiğim gibi, tekrarlı ölçümler desenlerinin önemli avantajlarından biri tüm uygulamalarda aynı deneklerin kullanılmasıyla hata terimini azaltmayı sağlamasıdır. Bir an için Tablo 14.3'te gösterilen verinin beş farklı denek grubundan üretildiğini varsayalım. Böyle bir analizde $KT_{\text{hata}}$ 717,11'e eşit olur. Bu durumda, $KT_{\text{denekler arası}}$ ile $KT_{\text{toplam}}$ eş anlamlı olduğundan (Denek toplamı ve bireysel puan aynıdır.) denekler için bir terim çıkarmak mümkün olmaz. Sonuç olarak, denekler arasında farklılıklar hatadan ayrılamaz ve hatta $KT_{\text{hata}}$, tekrarlı-ölçümler deseni için, $KT_{\text{hata}}$ ve $KT_{\text{denekler arası}}$'nın toplamından (=230,4 + 486,71 = 717,71; 32 + 8 = 40 serbestlik derecesinde) oluşur. Bu durum $F$'in

$$F = \frac{KO_{haftalar}}{KO_{hata}} = \frac{612,30}{17,93}) = 34,15$$

olmasına neden olur. Bu değer hâlâ manidar olmasına rağmen Tablo 14.3'te bulunan değerin yarısından daha azdır.

Kısaca söylemek gerekirse denekler farklılık göstermektedir. Denekler sadece bir kez gözlendiğinde, deneklerdeki bu farklılık hata terimine katkıda bulunur. Denekler tekrarlı bir biçimde gözlendiğinde, denek farklılıklarının derecesine ilişkin bir tahmin elde edilir ve kısmen bu farklılıklar hata teriminin dışında kalır. Genel olarak, denekler arasındaki farklılıklar arttıkça, uygulama çiftleri arasındaki korelasyon artar. Uygulamalar arasındaki korelasyon arttıkça, tekrarlı ölçüm desenlerinin bağıl gücü artar.

## 14.5 Tekrarlı-Ölçümler Desenlerinde Karşılaştırmalar ve Etki Büyüklüğü

Tek-yönlü ve faktöriyel-desenlerde yaptığımız gibi, tekrarlı ölçümlere ait değişkenlerin ortalamalarını nasıl karşılaştıracağımızı düşünmeliyiz. Neyse ki burada çok fazla yeni bir şey yapmayacağız. Burada yine bir durumun veya durum kümelerinin ortalamalarını başka bir durumun veya durum kümelerinin ortalamalarıyla karşılaştıracağız ve daha önce kullandığımız katsayıları kullanacağız.

Örneğimizde ilk iki hafta Başlangıç düzeyi ve son üç hafta ise Eğitim sonu ölçümleriydi. Çok amaçlı $F$ değerimiz, bu beş Hafta arasında istatistiksel olarak manidar bir fark olduğunu söyler, fakat farklılığın nereden geldiğini söylemez. Şimdi başlangıç haftaları ortalamasıyla Eğitim haftaları ortalamasını karşılaştırmak istiyorum. Bu karşılaştırmayı gerçekleştirecek katsayılar aşağıda ortalamalar ile birlikte verilmiştir.

| | 1. Hafta | 2. Hafta | 3. Hafta | 4. Hafta | 5. Hafta |
|---|---|---|---|---|---|
| Katsayı | 1/2 | 1/2 | −1/3 | −1/3 | −1.3 |
| Ortalama | 22,333 | 22,000 | 9,333 | 5,778 | 6,778 |

© Cengage Learning 2013

Yapmakta olduğumuz gibi karşılaştırmaları aşağıdaki gibi tanımlayacağız:

$$\begin{aligned}\hat{\psi} &= \sum a_i X_i \\ &= \left(\frac{1}{2}\right)(22,333) + \left(\frac{1}{2}\right)(22,000) + \left(-\frac{1}{3}\right)(9,333) + \left(-\frac{1}{3}\right)(5,778) + \left(-\frac{1}{3}\right)(6,778) \\ &= \frac{22,333 + 22,000}{2} - \frac{9,333 + 5,778 + 6,778}{3} = \frac{44,333}{2} - \frac{21,889}{3} \\ &= 22,166 - 7,296 \\ &= 14,870\end{aligned}$$

Bu karşılaştırmayı $t$ veya $F$ ile yapabiliriz. Ancak burada $t$'yi kullanacağım. ($F$, $t$'nin karesidir.)

$$t = \frac{\hat{\psi}}{\sqrt{\frac{(\sum a_i^2) KT_{\text{hata}}}{n}}} = \frac{14,870}{\sqrt{\frac{0,833(7,20)}{9}}} = \frac{14,870}{\sqrt{0,667}} = \frac{14,870}{0,816} = 18,21$$

Burada $sd_{\text{hata}}$= 32 $sd$'dir ve $t$ çok açık bir biçimde manidardır.

$t$'yi hesaplarken tüm analizdeki $KO_{\text{hata}}$'yı kullandığıma dikkat ediniz. Bu değer Haftaların etkisini test etmek için kullandığım hata terimi ile aynıdır. Bu eşitliğe burada dikkat

çektim, çünkü daha karmaşık analizlerde daha fazla hata terimlerimiz olacaktır ve belirli bir karşılaştırmayı yapmak için kullanılan hata terimi, bağımsız değişkenin ana etkisini test etmek için kullanılan hata terimidir.

### Etki Büyüklükleri

Ortalamalar arasında karşılaştırma yaparken tek-yönlü desenlerden tekrarlı ölçüler desenlerine doğrudan aktarım olmasına rağmen, etki büyüklüğü konusuna geldiğimizde bu durum biraz daha karmaşıktır. Etki büyüklüğü aşağıdaki gibi tanımlanır:

$$\hat{d} = \frac{\hat{\psi}}{s_{\text{hata}}}$$

$\hat{\psi}$ ile bir problemimiz olmamalı, çünkü yukarıda hesapladığımız karşılaştırma ile aynıdır - başlangıç haftaları ortalaması ile eğitim haftaları ortalaması arasındaki fark. Fakat, $s_{\text{hata}}$'nın hesaplanmasında birkaç seçenek vardır. Kline (2004) paydanın hesaplanmasında üç muhtemel seçenek sunmuş, ancak bunlardan ikisinin haftalar arasındaki korelasyonu göz ardı ettiğinden veya $\hat{\psi}$'yı özel bir anlamı olmayan standart sapma ile standardartlaştırdığından tatmin edici olmayabileceğine dikkat çekmiştir.

Aslında yapacağımız şey belirli bir karşılaştırma için benzersiz bir hata terimi oluşturmaktır. İki başlangıç düzeyi ölçümü ve üç işlem ölçümü yerine her biri için birer hata terimine sahip olduğunuzu düşünün. Burada uygun olan test tekrarlı-ölçümler için *t* testidir. Ve etki büyüklüğünün paydası büyük bir olasılıkla başlangıç düzeyi puanlarının standart sapmasıdır. ANOVA'mızla bu karşılaştırma için akla yakın standartlaştırma ölçüsü, başlangıç haftaları puanlarının ortalamalarına ait standart sapma olabilir. Örneğimiz için bu ortalamalar;

21,5 19,5 16 27,5 28,5 23 21 17,5 25

olacaktır. Bunların standart sapması 4,272 ve bu durumda anlamlı etki büyüklüğü de;

$$\hat{d} = \frac{\hat{\psi}}{s_{\text{hata}}} = \frac{14,87}{4,272} = 4,95$$

olur.

Bu bilgi, işlem sırasında yaşanan baş ağrılarının, başlangıç düzeyi baş ağrılarına göre 5 standart sapma daha az olduğunu gösterir. Bu oldukça önemli bir azalmadır. Etki büyüklüğü hesaplamasıyla ilgili farklı yolları görmek için Kline (2004) ile Olejnik ve Algina'nın (2000) çalışmalarına bakınız.

## 14.6 Sonuçların Yazılması

Bu deneyin sonuçlarını yazmak için basitçe aşağıdakileri söyleyebiliriz:

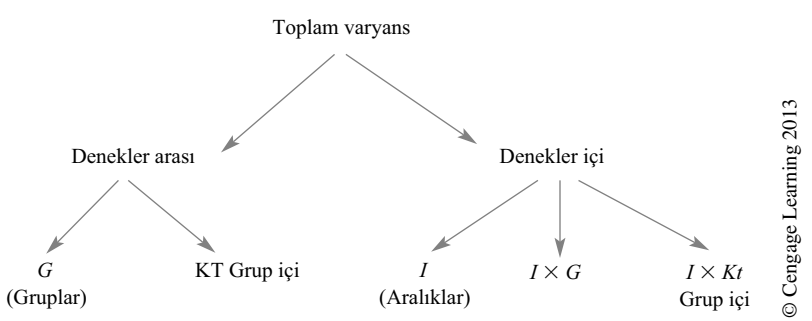
> Rahatlama tekniklerinin migren ağrısının şiddeti üzerindeki etkisini araştırmak için, dokuz katılımcı rahatlama terapisi almadan önce, iki hafta ve üç hafta süren terapiyi alırken migren ağrılarının şiddetini derecelendirmiştir. Tekrarlı ölçümler için varyans analizi sonuçları haftalar arasında manidar bir fark olduğunu göstermiştir ($F(4,32) = 85,04$, $p < ,05$). Başlangıç düzeyi haftalarının ağrı şiddeti ortalaması 22,166 iken terapi haftaları ortalaması 7,296 olmuştur ve aralarındaki fark 14,87'dir. Bu fark manidardır ($t(32) = 18,21$, $p < ,05$). Standartlaştırılmış ölçümü elde etmek için başlangıç düzeyi ölçümlerinin standart sapması kullanılarak elde edilen etki büyüklüğü ölçümü 4,95 bulunmuştur. Bu, rahatlama terapisinin migren baş ağrısını tedavi etmedeki önemini gösterir.

## 14.7 Bir Denekler-Arası Değişken ve Bir Denekler-İçi Değişken

İçinde bir bağımsız değişkenin (denekler dışında) olduğu ve her bir deneği bu değişkenin her bir düzeyinde test ettiğimiz basit bir durumla ilgilendik. Aslında, tekrarlı ölçümleri kullanarak bir çalışmayı desenleyeceğimiz çok farklı yollar vardır. Örneğin, iki bağımsız değişken kullanarak bir deney oluşturabilir ve her bir deneği her iki değişkenin birleşiminde test edebiliriz. Ya da bunun yerine, her bir denek, değişkenlerden birinin sadece bir düzeyinde, ancak diğerinin ise tüm düzeylerinde bulunabilir. Değişken sayısı üç olduğunda ise muhtemel durumlar daha da artar. Bu kesimde muhtemel desenlerden çok az bir kısmını tartışacağız. Eğer burada tartışılan desenleri anlarsanız, daha karmaşık problemlere genelleme yaparken zorluk yaşamazsınız.

Tablo 14.4'te gösterilen veri kümesini düşünün. Bu veri kümesi King'in (1986) bir çalışmasından alınan gerçek veridir. Bu çalışma bazı açılardan Bölüm 12'de incelediğimiz Siegel'ın (1975) morfin toleransı örneğine benzemektedir. King, farelerin Midazolam ilâcı enjekte edildikten sonra motor etkinliklerini incelemiştir. Bu ilâç ilk kez enjekte edildiğinde motor etkinliklerde belirgin bir azalmaya neden olur. Morfin gibidir, ancak midazolamın toleransı çok hızlı bir biçimde gelişmektedir. King, kazanılan bu toleransın Siegel'ın çalışmasındaki gibi, ilâcın verildiği fiziksel şartlara bağlı *şartlı* tolerans olup olmadığını bilmek istemiştir. Tablo 14.4'te görülen çok önemli veri kümesini toplarken üç grup kullanmıştır. Veriyi test günü olan son gün toplamıştır. Öntest sırasında, iki gruptaki farelere birçok gün tekrarlı olarak midazolam ilâcı, Kontrol gruptakilere ise fizyolojik serum suyu enjekte edilmiştir. Test günü, bir gruptaki farelere -"Aynı" grup olarak adlandırılmıştır- daha önce enjekte edilenle aynı ortamda midazolam enjekte edilmiştir. "Farklı" gruba ise midazolam *farklı* bir ortamda enjekte edilmiştir. Son olarak "Kontrol grubuna" da ilk defa midazolam verilmiştir. Böylece Kontrol grubu ilâca karşı ilk tipik tepkilerini (azaltılmış ambulatuvar davranış), Aynı grup da normal tolerans tepkisini göstermelidir - yani en son denemede, ilâca tepkilerinde, etkinliklerinde ya çok az veya hiç azalma olmamalıdır. Eğer King haklıysa, Farklı grup Kontrol grup ile benzer tepkiler vermelidir, çünkü daha önce birkaç kez ilâca maruz kalmalarına rağmen, ilâcı, yabancı farklı bir ortamda almaktadırlar ve şartlara bağlı geliştirebilecekleri tolerans böyle bir çıkarım için gerekli belirtiler olmayabilir. Tablo 14.4'teki bağımlı değişken tesadüfî birimlerdeki motor davranışların ölçümleridir. Yine değişken isminin ilk harfi alt simge olarak kullanılarak hangi ortalamaya karşılık geldiği belirtilmiştir.

İlâcın yaklaşık olarak 1 saat sonrasında metabolize edildiği bilindiği için, King kayıtlarını 5 dakikalık aralıklarla kaydetmiştir. İlâcın etkisinin ilk aralıklarda artacağını ve daha sonra yavaşça giderek azalarak yok olacağını bekleriz. Analizlerde ilk 6 aralıktaki ölçümler kullanılmıştır. Bu çalışmanın deseni diyagram ile aşağıdaki gibi gösterilebilir;




© Cengage Learning 2013

**Tablo 14.4** Gruplara ve Denemelere göre ayakta tedavi davranışı

**(a) Veri**

| | | Aralık | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Ortalama |
| Kontrol | | 150 | 44 | 71 | 59 | 132 | 74 | 88,333 |
| | | 335 | 270 | 156 | 160 | 118 | 230 | 211,500 |
| | | 149 | 52 | 91 | 115 | 43 | 154 | 100,667 |
| | | 159 | 31 | 127 | 212 | 71 | 224 | 137,333 |
| | | 159 | 0 | 35 | 75 | 71 | 34 | 62,333 |
| | | 292 | 125 | 184 | 246 | 225 | 170 | 207,000 |
| | | 297 | 187 | 66 | 96 | 209 | 74 | 154,833 |
| | | 170 | 37 | 42 | 66 | 114 | 81 | 85,000 |
| | Ortalama | 213,875 | 93,250 | 96,500 | 128,625 | 122,875 | 130,125 | 130,875 |
| Aynı | | 346 | 175 | 177 | 192 | 239 | 140 | 211,500 |
| | | 426 | 329 | 236 | 76 | 102 | 232 | 233,500 |
| | | 359 | 238 | 183 | 123 | 183 | 30 | 186,000 |
| | | 272 | 60 | 82 | 85 | 101 | 98 | 116,333 |
| | | 200 | 271 | 263 | 216 | 241 | 227 | 236,333 |
| | | 366 | 291 | 263 | 144 | 220 | 180 | 244,000 |
| | | 371 | 364 | 270 | 308 | 219 | 267 | 299,833 |
| | | 497 | 402 | 294 | 216 | 284 | 255 | 324,667 |
| | Ortalama | 354,625 | 266,250 | 221,000 | 170,000 | 198,625 | 178,625 | 231.521 |
| Farklı | | 282 | 186 | 225 | 134 | 189 | 169 | 197,500 |
| | | 317 | 31 | 85 | 120 | 131 | 205 | 148,167 |
| | | 362 | 104 | 144 | 114 | 115 | 127 | 161,000 |
| | | 338 | 132 | 91 | 77 | 108 | 169 | 152,500 |
| | | 263 | 94 | 141 | 142 | 120 | 195 | 159,167 |
| | | 138 | 38 | 16 | 95 | 39 | 55 | 63,500 |
| | | 329 | 62 | 62 | 6 | 93 | 67 | 103,167 |
| | | 292 | 139 | 104 | 184 | 193 | 122 | 172,333 |
| | Ortalama | 290,125 | 98,250 | 108,500 | 109,000 | 123,500 | 138,625 | 144,667 |
| | Aralık ortalama | 286,208 | 152,583 | 142,000 | 135,875 | 148,333 | 149,125 | 169,021 |

**(b) Hesaplamalar**

$$KT_{\text{toplam}} = \sum (\text{X} - \overline{\text{X}}...)^2 = (150 - 169{,}021)^2 + ... + (122 - 169{,}021)^2 = 1.432.292{,}9$$

$$KT_{\text{denek}} = \text{e}\sum (\overline{\text{X}}_{\text{S}} - \overline{\text{X}}...)^2 = 6\left[(88{,}333 - 169{,}021)^2 + ... + (172{,}333 - 169{,}021)^2\right] = 670.537{,}1$$

$$KT_{\text{gruplar}} = \text{na}\sum (\overline{\text{X}}_{\text{G}} - \overline{\text{X}}...)^2 = 8 \times 6\left[(130{,}875 - 169{,}021)^2 + ... + (144{,}667 - 169{,}021)^2\right] = 285.815{,}0$$

$$KT_{\text{aralıklar}} = \text{ng}\sum (\overline{\text{X}}_{\text{I}} - \overline{\text{X}}...)^2 = 8 \times 3\left[(286{,}208 - 169{,}021)^2 + ...\text{"}(149{,}125 - 169{,}021)^2\right] = 399.736{,}5$$

$$KT_{\text{hücreler}} = \text{n}\sum (\overline{\text{X}}_{\text{GI}} - \overline{\text{X}}...)^2 = 8\left[(213{,}875 - 169{,}021)^2 + ... + (138{,}625 - 169{,}021)^2\right] = 766.371{,}5$$

$$SS_{\text{aralıklar} \times \text{G}} = KT_{\text{hücreler}} - KT_{\text{aralıklar}} - KT_{\text{gruplar}} = 766.371{,}5 - 285.371{,}5 - 285.815{,}0 - 399.736{,}5 = 80.820{,}0$$

© Cengage Learning 2013

*(devam ediyor)*

**Tablo 14.4** *(Devamı)*

**(c) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 23 | 670.537,1 | | |
| Gruplar | 2 | 285.815,0 | 142.907,5 | 7,80* |
| *Kt* Gruplar içi** | 21 | 384.722,0 | 18.320,1 | |
| Denekler içi** | 120 | 761.755,8 | | |
| Aralıklar | 5 | 399.736,5 | 79.947,3 | 29,85* |
| $A \times G$ | 10 | 80.820,0 | 8.082,0 | 3,02* |
| $A \times Kt$ gruplar içi** | 105 | 281.199,3 | 2.678,1 | |
| Toplam | 143 | 1.432.292,9 | | |

* $p < .05$;

** Çıkarma işlemiyle hesaplandı.

Burada denekler içi farklılıkları temsil eden etkilerden, denekler arasındaki farklılıkları temsil eden etkileri ayırdık. Denekler arası terimi denek grupları arasındaki farklılıklar, (*G*) ve aynı grup içindeki (*Kt* grup içi) denekler arası farklılıklar olarak ayrıştırılabilir. Denekler içi terimi aynı şekilde aralıkların temel etkisi (tekrarlı ölçümler) ve bu aralıkların denekler arası varyansın iki ayrışımı ile ayrı ayrı etkileşimi ile birlikte 3 bileşene ayrılabilir. Zamanı geldiğinde bu ayrışımı özet tabloda göreceksiniz.

## Denekler-Arası Etkilerin Ayrıştırılması

Öncelikle denekler-arası terimini daha ayrıntılı bir biçimde inceleyelim. Deneyin deseninden, bu terimin iki kısma ayrıştırılacağını biliyoruz. Bu kısımlardan biri uygulamalar (Kontrol, Aynı ve Farklı) farklı denek gruplarını içerdiğinden Grupların (*G*) ana etkisidir. Ancak bu, denekler arası farklılıkların tek kaynağı değildir. Kontrol grubunda 8 farklı denek vardır ve bunlar arasındaki farklılık elbette denekler-arası farklılıktır. Aynı durum diğer gruplardaki denekler için de geçerlidir. Şu anda gruplar içinden, yani aynı grup içinde olan denekler arası farklılıklardan bahsediyoruz - bu, *Kt* grup içidir.

Eğer geçici olarak aralıklar tamamıyla görmezden gelinirse (örneğin, basit bir biçimde 5-dakikalık aralıklara bölmek yerine veriyi oturumun tamamından toplarız), aşağıdaki veri kümesinin oluşturulduğu bir çalışma düşünebiliriz.

| Kontrol | Aynı | Farklı |
|---|---|---|
| 88,333 | 211,500 | 197,500 |
| 211,500 | 233,500 | 148,167 |
| 100,667 | 186,000 | 161,000 |
| 137,333 | 116,333 | 152,500 |
| 62,333 | 236,333 | 159,167 |
| 207,000 | 244,000 | 63,500 |
| 154,833 | 299,833 | 103,167 |
| 85,000 | 324,667 | 172,333 |
| 130,875 | 231,521 | 144,667 |

© Cengage Learning 2013

Bu tablodaki "ham puanlar" Tablo 14.4'teki denek ortalamalarıdır. Her bir denek bu toplamlar içinde sadece bir kez gösterildiğinden, bağımsız gruplarda tek-yönlü varyans

analizinde kullanılan yöntemler uygulanır. Aslında her bir deneğin puan sayısını (son kısımda sadeleştirilen) temsil eden sabit sayı dışında, bu verinin tek-yönlü varyans analizi için kareler toplamı, asıl analizlerle aynıdır. Eğer bu analiz denek toplamları üzerinde tek yönlü olsaydı grupların temel etkisini test eden *F* testi, analizin tümünden elde edilen ile muhtemelen aynı olacaktı. Böylece, toplam varyansın denekler-arası kısmı ayrı bir denekler-arası değişkenlik olarak görülebilir, ki bu esas olarak kendisine ait hata terimi ile (bazen **hata$_{\text{arasında}}$** olarak adlandırılır) denekler-içi etkilerden bağımsız ayrı bir varyans analizi olur.

hata$_{\text{arasında}}$

## Denekler-İçi Etkilerin Ayrıştırılması

Bir sonraki aşama $KT_{\text{toplam}}$ ayrışımlarının denekler-içi ögelerinin dikkate alınmasıdır. Daha önce gördüğümüz gibi, kendisi üç terime ayrıştırılmıştır. Altı aralığın karşılaştırılması aynı denekten gelen puanların karşılaştırılmasını içerir ve böylece Aralıklar denekler içi terimidir yani her bir deneğin içindeki farklılıklara dayanır. Aralıklar denekler içi terimleri olduğundan, Aralıkların Gruplar ile etkileşimi de denekler içi etkidir. Üçüncü terim, denekler içi etkilerin hata terimi olduğundan (Aralıklar × *Kt* gruplar içi) bazen **hata$_{\text{gruplar içi}}$** olarak adlandırılır. $KT_{\text{Aralıklarx}Kt\text{gruplariçi}}$ terimi her bir grup için hesaplanan A × D etkileşimi için kareler toplamının toplamıdır. Böylece, bir önceki desende kullanılan hata terimine mantıksal olarak eşit olduğu görülebilir.

hata$_{\text{gruplar içi}}$

## Analizler

Analizleri ayrıntılı olarak incelemeden önce, sonuçların genel örüntüsüne bakmak yol gösterici olur. Dağılımları ciddî bir biçimde incelemek için, her bir hücrede yeterli sayıda gözlem olmasa da, gruplar içindeki aralıkların her birinde azımsanmayacak bir değişkenliğin olduğu çok açıktır. Örneğin, kontrol grubundaki ikinci aralık için puan aralığı 0 ile 270 arasındadır. Aşırı aykırı değerlerin olmadığı görülmekle birlikte hücreler içindeki değişkenlik fazla olmasına rağmen, aşağı yukarı eşit olması bu gibi araştırmalarda sıklıkla görülür. Aynı zamanda hayvanlardan bazılarının nispeten daha az ve bazılarının ise oldukça çok motor davranışları gösterdiğini yani oldukça fazla bireysel farklılıkların olduğunu görebilirsiniz. Bu farklılıklar, analizlerimizde ayrıştırdığımız türde farklılıklardır. Aralıkların ortalamalarına bakınız. Beklendiği gibi, davranışların ilk beş dakikadan sonra azaldığını ve diğer aralıklarda çok az miktarda arttığını göreceksiniz. Sonuç olarak Kontrol ve Aynı gruplarının ortalamaları arasındaki farka bakarak beklenen tolerans etkisini göreceksiniz ve Farklı grubuna bakıldığında, bu grubun Aynı grubuna değil daha çok Kontrol grubuna benzediğini fark edeceksiniz. Bu durum King'in daha önceden kestirdiği sonuçtur.

Her zaman yapılan asıl ve etkileşim etkileri hesaplamalarından çok farklı olmadığından Tablo 14.4b'deki hesaplamalar hakkında söylenecek çok az şey vardır. Bir faktörün denekler-arası veya denekler-içi olması özet tablodaki yerleşiminde ve nihaî *F*'in hesaplanmasında katkısı olsa da kareler toplamının hesaplanmasında yer almamaktadır.

Tablo 14.4c'de bulunan özet tabloda Kaynak sütunu, ilk olarak $KT_{\text{denekleraras\i}}$ ve $KT_{\text{denekleriçi}}$ olarak ayrıştırılan $KT_{\text{toplam}}$ ile deneyin desenini göstermektedir. Bu kareler toplamları daha sonra tekrar alt kısımlara ayrılacaktır. Üç terimin yanındaki çift yıldız imi (**) bunların çıkarma ile hesaplandığını belirtir. ($KT_{\text{denekleraras\i}}$, $KT_{Kt\text{denekleriçi}}$ ve $KT_{\text{A}x{Kt}\text{denekleriçi}}$), kareler toplamı toplanarak bulunur ve bütün parçalarının toplamına eşittir. Bu durum işimizi kolaylaştırır. Böylece aşağıdaki eşitlikler elde edilir:

$$KT_{\text{denekleriçi}} = KT_{\text{toplam}} - KT_{\text{deneklerarası}}$$
$$KT_{Kt\text{ gruplar içi}} = KT_{\text{deneklerarası}} - KT_{\text{gruplar}}$$
$$KT_{AxKt\text{ gruplar içi}} = KT_{\text{denekleriçi}} - KT_{\text{aralıklar}} - \text{KT}_{AG}$$

Son iki terim analizin hata terimi olacaktır.

Serbestlik derecesi nispeten daha kolay elde edilir. Asıl etkilerin her biri için serbestlik derecesi değişken düzeyinden 1 çıkarılarak bulunur. Böylece Denekler için *sd*, 24 – 1 = 23; Gruplar için *sd,* 3 - 1 = 2; ve Aralıklar için *sd,* 6 - 1 = 5 olur. Tüm etkileşim A × G için *sd* ise bileşenlerin serbestlik dereceleri çarpılarak elde edilir. Böylece, $sd_{AG} = (6 - 1)(3 - 1) = 10$'dur. Kalan serbestlik derecelerini hesaplamak için en kolay yol kareler toplamında yaptığımız gibi çıkarma yapmaktır.

$$sd_{\text{denekler içi}} = sd_{\text{toplam}} - sd_{\text{deneklerarası}}$$
$$sd_{Kt\text{ gruplar içi}} = sd_{\text{deneklerarası}} - sd_{\text{gruplar}}$$
$$sd_{AxKt\text{ gruplar içi}} = sd_{\text{denekleriçi}} - sd_{\text{aralıklar}} - sd_{AG}$$

Bu serbestlik dereceleri, neyi temsil ettikleri göz önünde bulundurularak da doğrudan bulunabilir. Her bir denekte, 6 – 1 = 5 *sd* vardır. 24 denek ile bu (5)(24) = 120 $sd_{\text{denekleriçi}}$ olarak hesaplanır. Grup faktöründeki her bir düzey içinde, 8 – 1 = 7 *sd* gruplar arası ve 3 grup ile (7)(3) = 21 $sd_{\text{gruplariçi}}$ olur. $A\times Kt_{\text{gruplariçi}}$ gerçekten bir etkileşim terimidir ve bunun serbestlik derecesi $sd_{\text{A}}$ ile $sd_{Kt\text{ denekleriçi}}$'nin çarpımıdır yani (5)(21) = 105'tir.

Kareler toplamının serbestlik derecesine bölünmesiyle elde edilen kareler ortalamalarını atladığımızda *F*'e ulaşırız. *F* sütununda beklenildiği gibi Gruplar ve Aralıklar manidardır. Etkileşim de manidardır. Farklı grup, başlangıçta, Aynı ve Kontrol grupları arasında orta düzeyde olmasına rağmen, ikinci beş dakikalık aralıkla birlikte aşağı doğru düşerek Kontrol grubuna eşit olur. Bu bulgu şarta bağlı tolerans teorisi ile açıklanabilir. En azından son aralıklarda, sadece bir hayvana daha önce aldığından, farklı bir ortamda ilâç enjekte etmenin geliştirmiş olduğu toleransını düşürmesi için yeterli olması gerçekten ilginç bulgudur. Bu hayvanlar neredeyse hiç midazolam almamış hayvanlar gibi ilâca tepki vermişlerdir. Aralıklarda Grupları tek tek karşılaştırmaya daha sonra geri döneceğiz.

## Varsayımlar

*F* oranlarını daha doğrusu *F* dağılımlarını anlayabilmek için, her zaman yaptığımız normallik varsayımlarına, varyans homejenliğine ve $\widetilde{\sum}$'nın küreselliğine bakmamız gerekir. Bu durumda *Denekler-arası* terim(leri) için, bir Grup düzeyindeki denek ortalamalarının varyansının diğer bütün Grup düzeylerindeki denek ortalamalarının varyansıyla aynı olduğu varsayılmalıdır. Gerekirse bu varsayım Ünite 7'de verilen Levene (1960), Brown ve Forsythe (1974), O'Brien (1981) tarafından önerilen testler ile test edilebilir. Ancak pratikte varyans analizi bu varsayımların ihlâline karşı kısmen dayanıklıdır (bakınız Collier, Baker ve Mandeville, 1967; Collier, Baker, Mandeville ve Hayes, 1967). Gruplar bağımsız olduğundan, varyans homojenliğimiz varsa, köşegen dışındaki girdiler sıfır olacağından kovaryans matrisinin birleşik simetrisi ve böylece de küreselliği garantidir.

*Denekler-içi* terimler için küresellik varsayımları ile birlikte varyans homojenliği ve normallik varsayımlarını dikkate almak zorundayız. Bu durumda varyans homojenliği varsayımı $A \times D$ etkileşiminin Gruplar için sabit olmasıdır.

Varyans homojenliği varsayımına ek olarak kovaryans matrisi veya matrisler üzerinde iki varsayım vardır. Yine $\sum$ Aralık ($A$) düzeyleri arasındaki varyans ve kovaryansların matrisini temsil eder. Böylece 6 aralık ile $\hat{\sum}$ aşağıdaki gibidir:

$$\hat{\sum} = \begin{array}{c|cccccc} & A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \\ \hline & \hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \hat{\sigma}_{14} & \hat{\sigma}_{15} & \hat{\sigma}_{16} \\ & \hat{\sigma}_{21} & \hat{\sigma}_{22} & \hat{\sigma}_{23} & \hat{\sigma}_{24} & \hat{\sigma}_{25} & \hat{\sigma}_{26} \\ & \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_{33} & \hat{\sigma}_{34} & \hat{\sigma}_{35} & \hat{\sigma}_{36} \\ & \hat{\sigma}_{41} & \hat{\sigma}_{42} & \hat{\sigma}_{43} & \hat{\sigma}_{44} & \hat{\sigma}_{45} & \hat{\sigma}_{46} \\ & \hat{\sigma}_{51} & \hat{\sigma}_{52} & \hat{\sigma}_{53} & \hat{\sigma}_{54} & \hat{\sigma}_{55} & \hat{\sigma}_{56} \\ & \hat{\sigma}_{61} & \hat{\sigma}_{62} & \hat{\sigma}_{63} & \hat{\sigma}_{64} & \hat{\sigma}_{65} & \hat{\sigma}_{66} \end{array}$$

© Cengage Learning 2013

Her bir grup için ayrı evren varyans-kovaryans matrisimiz, $X_{Gi}$ olabilir. ($\sum$ ve $\sum_{Gi}$ sırasıyla $\hat{\sum}$ ve $\hat{\sum}_{Gi}$ ile kestirilir.) $KO_{AxKt\ \text{gruplariçi}}$'nin uygun bir hata terimi olabilmesi için, öncelikle bireysel varyans-kovaryans matrislerinin ($X_{Gi}$), $G$'nin tüm düzeyleri için aynı olduğu varsayılır. Bunun yaygın olarak kullanılan varyans homojenliği varsayımının genişletilmiş (kovaryanslara) bir durumu olduğu düşünülebilir.

Kovaryanslarla ilgili ikinci varsayım, genel $\sum$ matrisinin (ki burada $\sum$ matrisi $\sum_{Gi}$ matrisinin birleştirilmiş ortalamasıdır) tümüyle ilgilenir. (Örneğin, her gruba ait eşit örneklem büyüklükleri için $\sum$, karşılık gelen bireysel $\sum_{Gi}$ matrisindeki girdilerin ortalaması olmalıdır.) Gerekli olmayan ancak yaygın ve yeterli bir varsayım ise, matrisin birleşik simetri göstermesidir. Birleşik simetri göstermesinin anlamı, daha önce de söylediğim gibi, esas köşegendeki tüm varyansların eşit olması ve aynı şekilde esas köşegen dışındaki tüm kovaryasların eşit olmasıdır. Yine varyanslar ve kovaryansları eşit olmak zorunda değildir ve genellikle de olmaz. Bu varsayım gereğinden fazla katıdır. Gerçekten ihtiyacımız olan, Aralık ortalamaları çiftlerinin farkının standart hatasının sabit olması varsayımıdır. Diğer bir deyişle, $\sigma^2_{\overline{A}_i - \overline{A}_j}$ bütün $i$ ve $j$ için sabit bir sayıdır ($i \neq j$). Bu küresellik zorunluluğu, $\Sigma$'nın birleşik simetri göstermesi durumunda otomatik olarak karşılanır, ancak $\Sigma$'nın diğer örüntülerinin de aynı özelliği göstermesi gerekir. Küresellik Mauchley testi kullanılarak test edilir. Bu test yukarıda tartışılmıştır. Ancak Mauchley testini kullanmak yerine küreselliği endişe etmeden aşağıda gösterilen Greenhouse and Geisser (1959) veya Huynh ve Feldt (1976) tarafından geliştirilen küreselliğin olmadığı durumlar için kullanılabilen düzeltmeleri kullanarak daha ileriye gidilebilir.

Kovaryans varsayımları ile ilgili daha geniş bilgi için Huynh ve Feldt (1970) ve Hunh ve Mandeville'in (1979) çalışmalarına ve daha da ötesi için Edwards'a (1985, s. 327–329, 336–339) bakılabilir .

## Serbestlik Derecelerinin Düzeltilmesi

Greenhouse ve Geisser (1959) $\sum$ üzerindeki küresellik varsayımının kullanılmadığı durumda oluşabilecek etkiler ile ilgilenmişlerdir. $\sum$'nın yapısına bakmadan, varyans analizinin denekler-içi kısmındaki $F$ oranını incelemiş A x G etkileşimi için

$(i - 1)\ \varepsilon,\ g\ (n - 1)(i - 1)\varepsilon$

ve

$(g - 1)(i - 1)\ \varepsilon,\ g\ (n - 1)(i - 1)\varepsilon$

*sd*'de yaklaşık olarak $F$ dağılımı gösterdiği ve $i$ = aralık sayısı olmak üzere $\varepsilon$'un aşağıdaki formül yardımıyla elde edilen $\hat{\omega}$ ile kestirilebileceği sonucuna ulaşmışlardır.

$$\hat{\varepsilon} = \frac{i^2(\bar{s}_{jj} - \bar{s})^2}{(i-1)(\sum s_{jk}^2 - 2i\sum \bar{s}_j^2 + i^2\bar{s}^2)}$$

Burada,

$\bar{s}_{jj}$ = $\hat{\sum}$'nın esas köşegenindeki elemanlarun ortalamasını

$\bar{s}$ = $\hat{\sum}$'nın tüm elemanlarının ortalamasını

$s_{jk}$ = $\hat{\sum}$'daki *jk*'yıncı elemanı

$\bar{s}_j$ = $\hat{\sum}$'nın *j*'yinci satırındaki elemanların ortalamasını

gösterir.

$\hat{\omega}$ kullanılmasının etkisi $sd_{\text{etki}}$ ve $sd_{\text{hata}}$'nın normal olarak olması gerekenden daha az olmasını sağlamaktır. Böylece $\hat{\omega}$ düzeltmeyi yaptığımız basit bir orandır. Greenhouse ve Geisser $\hat{\omega}$ kullanarak serbestlik derecesini düzeltmeyi önermişlerdir. Daha sonra küresellik varsayımının karşılanması durumunda $\hat{\omega}$ = 1 olduğunu ve küresellikten ayrıldıkça da $\varepsilon$'nun en küçük değeri olarak 1/(*i* - 1)'e yaklaştığını göstermişlerdir.

$\varepsilon$'nun büyük değerleri için $\hat{\varepsilon}$'nın kullanılmasının serbetlik derecesinin düzeltilmesi için tutucu bir test olduğuna ilişkin bazı düşünceler vardır. Huynh ve Feldt (1976) bu düzeltmeyi incelemiş ve $\varepsilon$'nun gerçek değerinin 0,75 civarında veya daha büyük olduğuna inanmaya dair bir neden olduğunda $\hat{\varepsilon}$ için bir değişiklik önermişlerdir. Huynh ve Feldt daha sonra Lecoutre'yi (1991)[3] düzelterek N = *n* x *g* olmak üzere aşağıdaki eşitliği oluşturmuşlardır.

$$\tilde{\varepsilon} = \frac{(N - g + 1)(i-1)\hat{\varepsilon} - 2}{(i-1)[N - g - (i-1)\hat{\varepsilon}]}$$

Burada $\omega$'nun asıl değerini tahmin etmek için $\hat{\omega}$ *veya* $\overline{\omega}$ kullanırız. (Bazı durumlarda $\overline{\omega}$, 1'i geçer. Bu durumda değer 1'e sabitlenir.)

İhtiyacımız olduğunda, Mauchley'ninki gibi küresellik testi kullandığımızda, bu testler ile ilgili muhtemelen çok ciddî problemler olabilmesi sebebiyle, daima $\hat{\omega}$ veya $\tilde{\omega}$'dan hangisi uygunsa serbestlik derecesinde düzeltme yaparız veya uygun olursa daha sonra tartışacağımız çoklu bir süreç izleriz. (Hayes ve Cai'nin (2007) çalışmalarının sonuçlarına dayalı olarak bağımsız örneklemler *t* testi için benzer bir strateji öneririm. Hayes ve Cai, bütün durumlarda, varyans homojenliği testine aldırmadan Welch-Satterthwaite yaklaşımını önermişlerdir.)

$\hat{\omega}$ *ve* $\overline{\omega}$'yi hesaplayarak ve uygun *sd*'ni kullanarak *F* değerini değerlendirebiliriz. Birleşik varyans-kovaryans matrisi (farklı matrislerin ortalaması alınarak bulunmuştur) Tablo 14.5'de görülmektedir. (Kabaca eşit olduklarından ve matristeki her bir elemanın sadece 8 gözleme dayalı olması nedeniyle birkaç grup için varyans-kovaryans matrisini göstermedim.)

Tablo 14.5'ten $\hat{\omega}$ *ve* $\overline{\omega}$ değerlerinin sırasıyla 0,6569 ve 0,8674 olduğu görülür. Bu değerler 0,75'in komşuluğunda olduğundan Huynh ve Feldt'in önerisi dikkate alınarak $\overline{\omega}$ değeri kullanılır. Bu durumda, etkileşim için serbestlik derecesi

$(g - 1)(i - 2)(0{,}7508) = 7{,}508)$

ve

$g(n - 2)(i - 1)(0{,}7508) = 78{,}834)$

bulunur. $F_{0,05}$ (7,508;78,834)'in kritik değeri 2,09'dur. Bu değer etkileşim için sıfır hipotezinin reddedileceği anlamına gelir. Böylece küresellikle ilgili herhangi bir probleme aldırış edilmeksizin, bu analizdeki tüm etkiler manidardır. ($\hat{\varepsilon}$ yerine $\overline{\varepsilon}$ kullanılması durumunda da muhtemelen manidar olacaktır.)

---

3 Chen ve Dunlop (1994), Huynh ve Feldt'in orijinal Lecoutre'ın formülünde yaptıkları düzeltmeyi doğrulamışlardır. Ancak hem SPSS hem de SAS Huynh ve Feldt epsilonunu yanlış hesaplamaya devam etmektedir.

**Tablo 14.5** Varyans-kovaryans matrisi ve $\hat{\varepsilon}$ ve $\tilde{\varepsilon}$'un hesaplanması

| Aralık | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | Ortalama |
| 6388,173 | 4696,226 | 2240,143 | 681,649 | 2017,726 | 1924,066 | 2991,330 |
| 4696,226 | 7863,644 | 4181,476 | 2461,702 | 2891,524 | 3531,869 | 4271,074 |
| 2240,143 | 4181,476 | 3912,380 | 2696,690 | 2161,690 | 3297,762 | 3081,690 |
| 681,649 | 2461,702 | 2696,690 | 4601,327 | 2248,600 | 3084,589 | 2629,093 |
| 2017,726 | 2891,524 | 2161,690 | 2248,600 | 3717,369 | 989,310 | 2337,703 |
| 1924,066 | 3531,869 | 3297,762 | 3084,589 | 989,310 | 5227,649 | 3009,208 |

$$\bar{s}_{ij} = \frac{6388,173 + 7863,644 + ... + 5227,649}{6} = 5284,090$$

$$\bar{s} = \frac{6388,173 + 4696,226 + ... + 989,310 + 5227,649}{36} = 3053,350$$

$$\sum s_{jk}^2 = 6388,173^2 + 4696,226^2 + ... + 5227,649^2 = 416.392,330$$

$$\sum \bar{s}_{jk}^2 = 2991,330^2 + ... + 3009,208^2 = 58.119,260$$

$$\hat{\varepsilon} = \frac{i^2(\bar{s}_{ij} - \bar{s})^2}{(i-1)\left(\sum s_{jk}^2 - 2i\sum \bar{s}_{jk}^2 + i^2\bar{s}^2\right)}$$

$$= \frac{36(5285,090 - 3053,350)^2}{(6-1)\left[416.392,330 - (2)(6)(58.119,260) + (36)(3053,350^2)\right]}$$

$$= \frac{179.303,883}{5\left[416.392,330 - 697.431,120 + 335.626,064\right]} = 0,6569$$

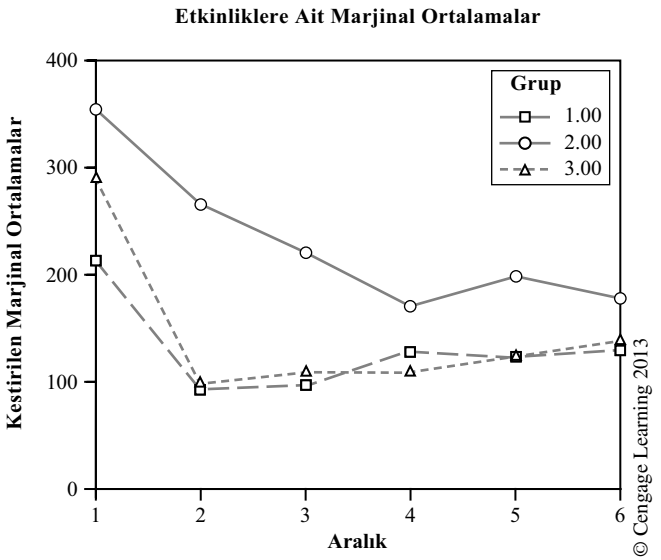
$$\tilde{\varepsilon} = \frac{(N - g + 1)(i-1)\hat{\varepsilon} - 2}{(i-1)[N - g - (i-1)\hat{\varepsilon}]}$$

$$= \frac{(24 - 3 + 1)(5)(0,6569) - 2}{5[25 - 3 - 5(0,6569)]} = \frac{70,259}{5[22 - 5(0,6569)]} = 0,7508$$

© Cengage Learning 2013

## Basit Etkiler

Şekil 14.2'de Aralık x Grup etkileşiminin grafiği çizilmiştir. Verinin yorumlanması nispeten açıktır. Aynı grubu, tutarlı bir biçimde diğer iki grubun düzeylerinin üstünde performans göstermiştir. Yani midazolama karşı şartlara dayalı tolerans bu gruptaki hareketliliğin diğer gruplara göre daha fazla olmasına sebep olmuştur. İlk 5 dakikalık aralık sonunda hareketler gözle görülebilir düzeyde azalmıştır. İlk aralıkta Farklı grubu diğer iki grubun arasında yer alsa da, daha sonra Kontrol grubundan ayırt edilebilmesi neredeyse imkânsız hale gelmiştir. Buna ek olarak en azından dördüncü aralığa kadar Aynı grubu azalmaya devam etmiş, öte taraftan diğer iki grup hızlı bir biçimde düşmüş ve sonra aynı düzeye gelmiştir. Bu sonuçları yorumlarken, özellikle ilk ve son aralıklardaki grup farklılıkları incelenerek basit etkilerin faydalı olduğu ispatlanacaktır. Kontrol grubu içindeki aralıklar arasındaki farklılıkları test etmek için sadece göstermek amacıyla yine basit etkiler kullanılacaktır. Her bir grup içerisinde aralık farklılıklarının olduğu belirgindir.




**Şekil 14.2** Tablo 14.4'teki veri için Aralık x Grup etkkileşimi

Daha önce de tavsiye ettiğim gibi, etkileşimin asıl etkisini test ederken Greenhouse ile Geisser, Huynh ve Feldt'in yaptığı serbestlik dereceleri düzeltmeleri küresellik varsayımı ile ilgili sorunları düzeltmek için yeterli bir çalışmadır. Ancak denekler içi basit etkiler düşünüldüğünde, düzeltmelerin yeterliliği ile ilgili ciddî bir problem ortaya çıkar (Boik, 1981; Harris, 1985). Basit etkileri test etmek için kullanılan geleneksel yöntem, bireysel denekler-içi karşılaştırmayı birleştirilmiş hata terimleriyle ($KO_{AxKt\ \text{gruplariçi}}$) karşılaştırarak test etmeyi içerir. Bu varsayım ile ilgili problem olduğunda, bu hata terimi bazen daha azını bazen de daha çoğunu tahmin ederek $F$ için uygun paydanın ne olacağını gösterir. Bu da I. Tip hata yapma olasılığı ile karışıklıklar olmasına neden olur. Bu sebeple farklı ve bazı açılardan daha basit bir yaklaşımı kabul edeceğiz.

Bu yaklaşım test edilen her bir etki için ayrı hata terimi elde eden Boik'in tavsiyelerini dikkate alacaktır. Böylece, Kontrol durumu için aralıkların basit etkilerine baktığımızda, örneğin hata terimi özellikle bu etkiyi belirtecek ve diğer basit hata terimlerini biraraya getirmeyecektir. Başka bir deyişle, sadece Kontrol grubuna bağlı olacaktır. Basit Aralık etkisini her bir grup için ayrı ayrı tekrarlı-ölçümlerle ilgili varyans analizi yaparak oldukça basit bir biçimde test edebileceğiz. Örneğin, Kontrol grubundaki aralıklar için Kesim 14.4'te tartışılan tek yönlü varyans analizi yapılabilir. Daha sonra geriye dönerek Aynı ve Farklı gruplarındaki aralıklar için ayrı ayrı benzer analiz uygulanabilir. Bu sonuçlar Tablo 14.6'da gösterilmiştir. Her birinde aralık farkları manidardır, $\hat{\omega}$ veya $\tilde{\omega}$'den hangisi uygunsa kullanılarak düzeltilen serbestlik derecesi sonrasında bile aralıklar arasındaki karşılaştırmalar manidardır.

Tablo 14.6'daki denekler-içi analize bakıldığında, $KO_{\text{hata}}$ ortalaması (ki bu ortalama 469. sayfada tüm analiz için hesaplanan $KO_{AxKt\ \text{gruplar içi}}$'dir) (2685,669 + 3477,571 + 1871,026)/3 = 2678,089 olacaktır. $F$ oranlarındaki bu paydalar geleneksel bir yöntem olan birleştirilmiş terimlerle işlem yaptığımızda, hesapladıklarımızdan gözle görülebilir şekilde farklıdır. Her bir analiz için $KO_{\text{Aralık}}$ terimlerinin etkisi için yaptığımız klâsik hesaplarla aynı olduğunu küçük bir çalışma ile gösterebilirsiniz .

Denekler arası etkiler için (örneğin Aralık 1'deki gruplar) süreç daha karmaşıktır. Denekler-içi örneğini izleyebilsek veya her bir Aralıktaki ayrı analizleri uygulayabilsek, ge-

**Tablo 14.6** King'den (1986) alınan veri kümesi için denekler-içi basit etkilerin hesaplanması

**(a) Kontrol Grubunda Aralık**

| Kaynak | sd | KT | KO | F |
|---|---|---|---|---|
| Denekler arası | 7 | 134,615.58 | | |
| Aralık | 5 | 76,447.25 | 15,289.45 | 5,69* |
| Hata | 35 | 93,998.42 | 2685.67 | |
| Toplam | 47 | 305,061.25 | | |

*$p < ,05$; $\hat{\varepsilon} = ,404$; $\tilde{\varepsilon} = ,570$

**(b) Aynı Grubunda Aralık**

| Kaynak | sd | KT | KO | F |
|---|---|---|---|---|
| Denekler arası | 7 | 175,600.15 | | |
| Aralık | 5 | 193,090.85 | 38,618.17 | 11,10* |
| Hata | 35 | 121,714.98 | 3477.57 | |
| Toplam | 47 | 490,405.98 | | |

*$p < ,05$; $\hat{\varepsilon} = ,578$; $\tilde{\varepsilon} = 1,00$

**(c) Farklı Grubunda Aralık**

| Kaynak | sd | KT | KO | F |
|---|---|---|---|---|
| Denekler arası | 7 | 74,506.33 | | |
| Aralık | 5 | 211,018.42 | 42,203.68 | 22,56* |
| Hata | 35 | 65,485.92 | 1871,03 | |
| Toplam | 47 | 351,010.67 | | |

*$p < ,05$; $\hat{\varepsilon} = ,598$; $\tilde{\varepsilon} = 1,00$

© Cengage Learning 2013

reksiz olarak serbestlik derecesinde kayıplara neden olur. Burada hata terimlerini birleştirmek geçerli bir yaklaşımdır ve genellikle de akıllıcadır.

Bu örnek için Aralık 1 ve Aralık 6'daki Grup etkisini inceleyeceğiz. Orijinal veri kümesi sayfa 468'de Tablo 14.4'de bulunmaktadır. Bu etkiler için kareler toplamı aşağıdaki gibidir:

$$KT_{\text{1. aralıkta gruplar}} = 8[(213,875 - 286,208)^2 + (354,625 - 286,208)^2 + (290,125 - 286,208)^2]$$
$$= 79.426,33$$

$$KT_{\text{6. aralıkta gruplar}} = 8[(130,125 - 149,125)^2 + (178,625 - 149,125)^2 + (138,625 - 149,125)^2]$$
$$= 10.732,00$$

Denekler-arası terimlerinin basit etkisini test etmek bir miktar kurnazlık gerektirir. Bir an için Aralık 1'deki grup etkisini düşünün. Grup ortalamaları, tekrarlı ölçümler yerine denekler üzerindeki gözlemlerin tek bir ortalamasını temsil ettiğinden bu durum aslında tekrarlı olmayan ölçümler için tek yönlü varyans analizidir. Böylece, denek farklılıkları deneysel hata ile karışmıştır. Bu durumda uygun kareler toplamı hatası $KT_{\text{hücreler içi}}$,Tablo 14.4'ten aşağıdaki gibi hesaplanır.

$$KT_{\text{hücreler içi}} = KT_{Kt \text{ grup içi}} + KT_{\text{Ax}Kt \text{ gruplar içi}}$$

$$KT_{\text{hücreler içi}} = 384.722,03 + 281.199,34 = 665.921,37$$

ve

$$KO_{\text{hücre içi}} = \frac{KT_{\text{hücre içi}}}{sd_{\text{grup içi kareler toplamı}} + sd_{\text{Axhücreler içi kareler toplamı}}}$$
$$= \frac{665.921,37}{21 + 105} = 5285,09$$

Eğer gerçekten neyi temsil ettiği hakkında düşünülürse bu özel $KO_{\text{hücre içi}}$ terimine niçin ihtiyaç duyulduğu kolaylıkla anlaşılır. Eğer size, Tablo 14.4'te bulunan veriden sadece 1. Aralıktaki veri verilseydi, bu üç grup arasındaki farklılıkları test etmek isteyecek, standart tek yönlü varyans analizi kullanacaktınız ve $KO_{\text{hata}}$ da gruplar içindeki varyansların ortalaması olacaktı. Benzer bir biçimde, eğer sadece 2., 3. ve böylece diğer aralıklardaki veri kümeleri alınmış olsaydı, yine bu üç grup içindeki varyansların ortalaması alınırdı. Hesaplamasını yeni yaptığımız $KO_{\text{hücre içi}}$ bu altı farklı veri kümesinin (Aralıklar) hata terimlerinin ortalamasıdır. Aslında bu değer, 18 hücrenin varyans ortalamasıdır. Artık $F$ oranları oluşturulabilir.

$$F_{1.\text{ aralıkta G}} = \frac{KO_{1.\text{ aralıkta G}}}{KO_{\text{hücre içi}}} = \frac{79.426,33/2}{5285,09} = 7,51$$

$$F_{6.\text{ aralık G}} = \frac{KO_{6.\text{ aralıkta grup}}}{KO_{\text{hücre içi}}} = \frac{10,732/2}{5285,09} = 1,02$$

Diğer bir zorluk $F$ değerinin değerlendirilmesinde ortaya çıkar. $KO_{\text{hücre içi}}$ değeri aynı zamanda iki *heterojen* hata kaynağının toplamını temsil ettiğinden, $KO_{\text{grup içi}}$ ve A x $\text{Kt}_{\text{gruplariçi}}$ için $E(KO)$ incelendiğinde görülebileceği gibi $F$ değeri 2 ve 126 *sd*'de dağılmayacaktır. Bu zorluktan, Bölüm 7'deki $t$ testine ilişkin karşılaştığımız benzer bir problemin çözümünde kullandığımız yolu kullanarak kurtulabiliriz. Basit olarak $F$ değerini hesaplayan uygun *sd* değerini hesaplarız. Daha açık bir ifade ile $f'$ sembolü ile gösterilen bir istatistik hesaplayarak $F_{,05}(\text{a-1}, f')$ değeri aşağıdaki eşitlikle hesaplanır:

$$f' = \frac{(u + \nu)^2}{\frac{u^2}{sd_u} + \frac{\nu^2}{sd_\nu}}$$

Bu eşitlikte,

$u = KT_{\text{Kt gruplar içi}}$

$\nu = KT_{AxKt\text{ gruplar içi}}$

olup $sd_u$ ve $sd_v$ bu veri kümelerine karşılık gelen serbestlik dereceleridir. Bizim örneğimizde;

$u = 384.722,03 \quad sd_u = 21$

$\nu = 281.199,34 \quad sd_\nu = 105$

$$f' = \frac{(384.722,03 + 281.199,34)^2}{\frac{384.722,03^2}{21} + \frac{281.199,34^2}{105}} = 56,84$$

olur. Bu sayı en yakın tam sayıya yuvarlandığında $f' = 57$ olur. Böylece $F$, $H_0$ altında, $(g - 1) = (2,57)$ *sd* üzerine dağılmıştır. 2 ve 57 *sd* için $F_{,05} = 3,16$'dır. Sadece 1. Aralıktaki farklılık manidardır. 30 dakikanın sonunda, üç grup eşit düzeyde performans gösterir. Birinci ve altıncı aralıklar arasında üç grubun farkının manidar olmadığını ve birçok kişinin her bir aralık için bu sonucu test edeceği sonucuna ulaşmak mantığa uygundur. Ancak bu

uygulamanın genel bir kural olmasına şiddetle karşıyım. Zaten birçok manidar test yapıldı ve daha fazla test yapılması sadece hata oranının artmasına neden olur. Grup farklılıklarının hangi noktada manidar olmamaya başladığına ilişkin önemli bir kuramsal sebep yoksa -ve ben bu duruma nadiren rastlandığını düşünüyorum- her bir aralığı test ederek hiçbir şey elde edilmez. Testler, gerçeklere dayanmayarak sadece merak için veya "tamamına" bakmak için değil, önemli sorulara cevap vermek amacıyla kullanılır.

### Çoklu Karşılaştırmalar

Denekler-içi değişkenlerin ortalamaları arasındaki farkı test etmek için kullanılan çoklu-karşılaştırma yöntemlerinin dayanıklılığı birçok çalışma ile araştırılmıştır. Maxwell (1980) denekler-arası birleşeni olmaksızın basit bir tekrarlı-ölçüm deseni üzerinde çalışmış ve birleştirilmiş hata teriminin kullanılmadığı çoklu-karşılaştırma yöntemlerinin benimsenmesini önermiştir. Buna benzer bir yöntemi (Games-Howell yöntemi) Bölüm 12'de tartıştık. (Migren ile ilgili çalışmanın analizinde birleştirilmiş hata terimi kullandım. Ancak varyans homojenliğinin olduğunu varsaymak akla yakındı ve bütün haftaları kullanıyordum. Eğer sadece üç haftayı içeren bir araştırma yapsaydım, sadece bu haftalara dayalı bir hata terimi hesaplamayı düşünebilirdim.)

Keselman ve Keselman (1988), Maxwell'in denekler-arası birleşenine sahip deseni ile ilgili çalışmasını genişletmiş ve benzer bir öneri getirmiştir. Aslında Gruplar farklı büyüklükte olduklarında ve küresellik ihlâl edildiğinde aileboyu hata oranları çok kötü biçimde bozulur. Biraz önce tartıştığımız tek-yönlü basit etkiler yöntemlerinde, her bir grup için tekrarlı ölçümler analizi yapılmasını öneririm. Daha sonraki çoklu karşılaştırma yöntemleri için özellikle eşit olmayan örneklem büyüklüklerinde bu basit etkileri araştırarak bu ayrı kovaryans matrislerini kullanan Games-Howel sürecini uygulamak muhtemelen daha akıllıca olur. Başka bir deyişle, Kontrol grupları için 3. ve 4. aralıkları karşılaştırmak, sadece Kontrol grubunda 3. Aralık ve 4. Aralık için hata terimleri oluşturulabilir.

Myers (1979) eşleştirilmiş örneklemler *t* testi ve Bonferroni düzeltmesini kullanarak Çoklu-karşılaştırma testi (Post hoc test) yapılmasını önermiştir. (Bu Migren örneğinde yaptığım ile aynıdır. Sadece bir karşılaştırma yaptığımdan Bonferroni düzeltmesini kullanmadım). Maxwell (1980) bu yaklaşımın aile-boyu hata oranını kontrol etmek için uygun, Baker ve Lew (1987) de güç bakımından Tukey'nin testine karşı genel olarak daha iyi olduğunu göstermiştir. Baker, gücü biraz daha fazla olan basit bir Bonferroni düzeltmesi önermiştir.

## 14.8 İki Denekler-Arası Değişken ve Bir Denekler-İçi Değişken

Önceki desenlerde tekrarlı-ölçümler için temel varyans analizi teorisi tartışıldı. Ancak araştırmacılar deneyleri bazıları veya tümü aynı denekler üzerinde tekrarlı ölçümleri temsil eden 3 veya daha çok değişken ile planlarlar, bir kısmı veya tamamı tekrarlı ölçümleri aynı denekler üzerinde tekrarlarlar. Bu desenlerin analizlerini kısaca tartışacağız. Ana etkiler ve etkileşimler için kareler toplamları her zaman başvurduğumuz yollardan ve hata terimleri çıkarma işlemi ile elde edileceğinden hesaplamalar kolay olacaktır.

Bu desenlerin ardındaki teorilere uzun uzadıya bakmayacağız. Öncelikli olarak iki değişkenli durumların genişletilmesinden daha farklı bir durum olmayacak. Bu desenlerle ilgili mükemmel bir tartışma için Maxwell ve Delaney'nin (2004) çalışmasına bakınız.

Örnek olarak St. Lawrence, Brasfield, Shirley, Jefferson, Alleyne ve O'Bannon (1995) tarafından Afrikalı-Amerikan yetişkinlerde HIV bulaşmasını önlemek için geliştirilen bir müdahale programı ile ilgili çalışmayı alıyorum. Bu çalışma, biri Eğitsel Kontrol

(Eğt. Kon.) durumu için 2 saatlik standart eğitim programı ve diğeri 8 haftalık Davranış Becerileri Eğitim programı (DBE) olmak üzer iki yaklaşımın karşılaştırmasını içermektedir. Denekler yetişkin erkek ve kadınlardan oluşmaktadır; ölçümler Öntest, Sontest ve 6. Ay, 12. Ay sonunda izleme testleri (İT6 ve İT12) yapılarak alınmıştır. Bu çalışmada çoklu bağımlı değişken vardır ancak bizim ilgileneceğimiz log(frek+1)'dir. Burada frek prezervatif kullanarak cinsel ilişkide bulunma sayısıdır.[4] Bu müdahale ve cinsiyet denekler arası faktör ile Zaman denekler içi faktör olmak üzere 2 x 2 x 4 tekrarlı-ölçümler desenidir. Bu desen aşağıdaki tablo ile belirtilebilir. Burada $G_i$, $i$'yinci denek grubunu temsil eder.

| | Davranış Becerileri Öğretimi | | | | Eğitsel Kontrol | | | |
|---|---|---|---|---|---|---|---|---|
| | Ön test | Son test | İT6 | İT12 | Ön test | Son test | İT6 | İT12 |
| Erkek | $G_1$ | $G_1$ | $G_1$ | $G_1$ | $G_1$ | $G_2$ | $G_2$ | $G_2$ |
| Kadın | $G_3$ | $G_3$ | $G_3$ | $G_3$ | $G_4$ | $G_4$ | $G_4$ | $G_4$ |

© Cengage Learning 2013

Ham veri ve hücrelerin toplamlarıyla ilgili gerekli özet bilgi Tablo 14.7a'da gösterilmiştir. (Bu veri kümesi St. Lawrence ve ark. tarafından raporlanan veriden oluşturulmuştur. Orijinalinde denek sayısı daha fazladır. Ondalık kısımlar dahil edilmemiştir.) Tablo 14.7b'de ana etkiler ve etkileşimler ile ilgili hesaplamalar bulunmaktadır. Burada, ana etki ve etkileşim ile ilgili hesaplamalar daha önce yapılanlarla tamamen aynıdır. Şimdi ihtiyacım oluncaya kadar hesaplamaları atlamak ve onun yerine desene ve sonuçların ana hatlarına odaklanma zamanıdır.

**Tablo 14.7** Lawrence ve arkadaşlarının çalışmasına (1995) ait veri ve analizi
**a) Veri**

| | Erkek | | | | Kadın | | | |
|---|---|---|---|---|---|---|---|---|
| | Ön test | Son test | İT6 | İT12 | Ön test | Son test | İT6 | İT12 |
| **Davranış Becerisi Eğitimi (DBE)** | 7 | 22 | 13 | 14 | 0 | 6 | 22 | 26 |
| | 25 | 10 | 17 | 24 | 0 | 16 | 12 | 15 |
| | 50 | 36 | 49 | 23 | 0 | 8 | 0 | 0 |
| | 16 | 38 | 34 | 24 | 15 | 14 | 22 | 8 |
| | 33 | 25 | 24 | 25 | 27 | 18 | 24 | 37 |
| | 10 | 7 | 23 | 26 | 0 | 0 | 0 | 0 |
| | 13 | 33 | 27 | 24 | 4 | 27 | 21 | 3 |
| | 22 | 20 | 21 | 11 | 26 | 9 | 9 | 12 |
| | 4 | 0 | 12 | 0 | 0 | 0 | 14 | 1 |
| | 17 | 16 | 20 | 10 | 0 | 0 | 12 | 0 |
| **Eğitimsel Kontrol (Eğt. Kon.)** | 0 | 0 | 0 | 0 | 15 | 28 | 26 | 15 |
| | 69 | 56 | 14 | 36 | 0 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 5 | 6 | 0 | 23 | 0 |
| | 4 | 24 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 35 | 8 | 0 | 0 | 25 | 28 | 0 | 16 |
| | 7 | 0 | 9 | 37 | 36 | 22 | 14 | 48 |
| | 51 | 53 | 8 | 26 | 19 | 22 | 29 | 2 |
| | 25 | 0 | 0 | 15 | 0 | 0 | 5 | 14 |
| | 59 | 45 | 11 | 16 | 0 | 0 | 0 | 0 |
| | 40 | 2 | 33 | 16 | 0 | 0 | 0 | 0 |

© Cengage Learning 2013

*(devam ediyor)*

4 Orijinal verinin pozitif çarpık bir dağılım göstermesi nedeniyle yazarlar burada logaritmik dönüşüm kullanmışlardır. Log(0) tanımlı olmadığından log(x+1) kullanmışlardır.

**Tablo 14.7** *(Devamı)*

Grup X Cinsiyet X Zaman Ortalamaları

| | | Ön test | Son test | İT6 | İT12 | Ortalama |
|---|---|---|---|---|---|---|
| BST | Erkek | 19.7 | 20.7 | 24.0 | 18.1 | 20.625 |
| BST | Kadın | 7.2 | 9.8 | 13.6 | 10.2 | 10.200 |
| EC | Erkek | 29.5 | 18.8 | 7.5 | 15.1 | 17.725 |
| EC | Kadın | 10.1 | 10.0 | 9.7 | 9.5 | 9.825 |
| Ortalama | | 16.625 | 14.825 | 13.700 | 13.225 | 14.594 |

Grup X Cinsiyet Ortalamaları

| | Erkek | Kadın | Ortalama |
|---|---|---|---|
| BST | 20.625 | 10.200 | 15.412 |
| EC | 17.725 | 9.825 | 13.775 |
| Ortalama | 19.175 | 10.012 | 14.594 |

**(b) Hesaplamalar**

$$KT_{\text{toplam}} = \sum (X - \overline{X})^2 = (7 - 14,594)^2 + \ldots + (0 - 14,594)^2 = 35404,594$$

$$KT_{\text{denek}} = z\sum (\overline{X}_{\text{denek}} - \overline{X})^2 = 4\left[(14 - 14,594)^2 + \ldots + (0 - 14,594)^2\right] = 21490,344$$

$$KT_{\text{grup}} = nzc\sum (\overline{X}_{\text{G}} - \overline{X})^2 = 10 \times 4 \times 2\left[(15,412 - 14,594)^2 + (13,775 - 14,594)^2\right] = 107,256$$

$$KT_{\text{cinsiyet}} = nzg\sum (\overline{X}_{\text{cinsiyet}} - \overline{X})^2 = 10 \times 4 \times 2\left[(19,175 - 14,594)^2 + (10,012 - 14,594)^2\right] = 3358,056$$

$$KT_{\text{GC hücreler}} = nz\sum (\overline{X}_{\text{GC hücreler}} - \overline{X})^2 = 10 \times 4\left[(20,625 - 14,594)^2 + \ldots + (9,825 - 14,594)^2\right] = 3529,069$$

$$KT_{\text{GC}} = KT_{\text{GC hücreler}} - KT_{\text{G}} - KT_{\text{C}} = 3529,069 - 107,256 - 3358,056 = 63,757$$

$$KT_{\text{zaman}} = ngc\sum (\overline{X}_{\text{Z}} - \overline{X})^2 = 10 \times 2 \times 2\left[(16,625 - 14,594)^2 + \ldots + (13,325 - 14,594)^2\right] = 274,069$$

$$KT_{\text{ZG hücreler}} = nc\sum (\overline{X}_{\text{ZC hücreler}} - \overline{X})^2 = 10 \times 2\left[(13,45 - 14,594)^2 + \ldots + (12,300 - 14,594)^2\right] = 1759,144$$

$$KT_{\text{ZG}} = KT_{\text{ZG hücreler}} - KT_{\text{Z}} - KT_{\text{G}} = 1759,144 - 274,069 - 107,256 = 1377,819$$

$$KT_{\text{ZC hücreler}} = ngKT_{\text{ZG hücreler}} = ng\sum (\overline{X}_{\text{ZC hücreler}} - \overline{X})^2 = 10 \times 2\left[(24,60 - 14,594)^2 + \ldots + (9,85 - 14,594)^2\right]$$
$$= 4412,044$$

$$KT_{\text{ZC}} = KT_{\text{ZC hücreler}} - KT_{\text{Z}} - KT_{\text{C}} = 4412,044 - 274,069 - 3358,056 = 779,919$$

$$KT_{\text{GZC hücreler}} = n\sum (\overline{X}_{\text{GZC hücreler}} - \overline{X})^2 = 10\left[19,7 - 14,594)^2 + \ldots + (9,50 - 14,594)^2\right] = 6437,294$$

$$KT_{\text{GZC}} = KT_{\text{GZC hücreler}} - KT_{\text{G}} - KT_{\text{Z}} - KT_{\text{C}} - KT_{\text{GZ}} - KT_{\text{GC}} - KT_{\text{ZC}}$$
$$= 6437,294 - 107,256 - 274,069 - 3358,056 - 1377,819 - 63,757 - 779,919 = 476,419$$

**(c) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 39 | 21.490,344 | | |
| Grup | 1 | 107,256 | 107,256 | 0,21 |
| Cinsiyet | 1 | 3358,056 | 3358,056 | 6,73* |
| $G \times C$ | 1 | 63,757 | 63,757 | 0,13 |
| Gruplar içi kareler toplamı** | 36 | 17.961,275 | 498,924 | |
| Denekler içi** | 120 | 13.914,250 | | |
| Zaman | 3 | 274,069 | 91,356 | 0,90 |
| $Z \times G$ | 3 | 1377,819 | 459,273 | 4,51* |
| $Z \times D$ | 3 | 779,919 | 259,973 | 2,55 |
| $Z \times G \times D$ | 3 | 476,419 | 158,806 | 1,56 |
| Z × Kt gruplar içi** | 108 | 11.006,025 | 101,908 | |
| Toplam | 159 | 35.404,594 | | |

*$p < ,05$

** Çıkarma işlemi yoluyla elde edilmiştir.

**Tablo 14.8** A, B ve C sabit ile beklenen ortalama kareler

| Kaynak | *sd* | *E(KO)* |
|---|---|---|
| Gruplar arası | $abn - 1$ | |
| *A* | $a - 1$ | $\sigma_e^2 + c\sigma_\pi^2 + nbc\theta_\alpha^2$ |
| *B* | $b - 1$ | $\sigma_e^2 + c\sigma_\pi^2 + nac\theta_\beta^2$ |
| *AB* | $(a - 1)(b - 1)$ | $\sigma_e^2 + c\sigma_\pi^2 + nc\theta_{\alpha\beta}^2$ |
| Gruplar içi Kt | $ab(n - 1)$ | $\sigma_e^2 + c\sigma_\pi^2$ |
| Gruplar içi | $abn(c - 1)$ | |
| *C* | $c - 1$ | $\sigma_e^2 + \sigma_{\gamma\pi}^2 + nab\theta_\gamma^2$ |
| *AC* | $(a - 1)(c - 1)$ | $\sigma_e^2 + \sigma_{\gamma\pi}^2 + nb\theta_{\alpha\gamma}^2$ |
| *BC* | $(b - 1)(c - 1)$ | $\sigma_e^2 + \sigma_{\gamma\pi}^2 + na\theta_{\beta\gamma}^2$ |
| *ABC* | $(a - 1)\,(b - 1)(c - 1)$ | $\sigma_e^2 + \sigma_{\gamma\pi}^2 + n\theta_{\alpha\beta\gamma}^2$ |
| $C \times$ Gruplar içi Kt | $ab(n - 1)(c - 1)$ | $\sigma_e^2 + \sigma_{\gamma\pi}^2$ |
| Toplam | $N - 1$ | |

© Cengage Learning 2013

Varyans analizi ile ilgili özet tablo Tablo 14.7c'de verilmiştir. Bu tabloda **, terimlerin çıkarma yoluyla elde edildiğini gösterir.

$$KT_{\text{denekler içi}} = KT_{\text{toplam}} - KT_{\text{denekler arası}}$$

$$KT_{Kt\ \text{gruplar içi}} = KT_{\text{denekler arası}} - KT_G - KT_D - KT_{GD}$$

$$KT_{ZxKt\ \text{gruplar içi}} = KT_{\text{denekler içi}} - KT_Z - KT_{ZG} - KT_{ZD} - KT_{ZGD}$$

Sondaki iki terim sırasıyla denekler arası ve denekler içi etkilere ait hata terimidir. Bu hata terimlerinin uygun olduğu, Tablo 14.8'de[5] verilen beklenen kareler ortalamaları incelenerek gösterilebilir. Tesadüfî ve karma modeller için beklenen kareler toplamı hesaplamaları için Kirk (1968) veya Winer'ın (1971) çalışmalarına bakabilirsiniz.

Tablo 14.7c'deki özet tabloda *F* sütunundan, Cinsiyetin asıl etkisinin ve Zaman x Grup etkileşiminin manidar olduğu görülür. Bu sonuçların her ikisi de manidardır. Hatırlayacağınız gibi, bağımlı değişken prezervatif kullanarak cinsel ilişkide bulunma sayısıydı (log(frekans+1)). Ortalamalar incelendiğinde yetişkin bayanların yetişkin erkeklere göre daha düşük bir kullanım frekansına sahip olduğu ortaya çıkmıştır. Bunun sebebi onların daha az cinsel ilişkide bulunmaları veya daha az kullanma yüzdesine sahip olmaları olabilir. St. Lawrence ve ark. tarafından verilen tamamlayıcı veriye göre, yetişkin kadınlar yetişkin erkeklere göre daha düşük bir kullanım yüzdesi belirtmişlerdir. Burada açık bir biçimde gördüğümüz şey, bildirilen cinsel ilişkide bulunma sıklığıdır.

Özet tablodaki en önemli sonuç Zaman × Grup etkileşimidir. Bu sonuç aslında tam bizim aradığımız sonuçtur. Ön testte grupların eşit olmasını istediğimizden Grup etkisi olup olmaması ile ilgilenmiyoruz. Bu eşitlik grup farklılıklarının etkisini azaltır. Aynı zamanda, Zamanın asıl etkisi ile de ilgilenmeyiz. Çünkü Kontrol grubu zaman geçtikçe fark edilebilir değişiklik göstermez ve bu Zaman etkisinin etkisini azaltır. Bizim istediğimiz şey DBE grubunun zamanla kullanımını arttırması ve Eğt. Kon. grubunun ise aynı kalmasıdır. Bu etkileşimdir ve tam bizim bulduğumuz sonuçtur.

5 Daha önce beklenen kareler ortalamaları tablosunda, tesadüfî terimlerin varyansı için $\sigma^2$ ve sabit terimlerin değişkenliğini belirtmek için de $\theta^2$ sembollerini kullandık. Bu çalışmada iki ana bağımsız değişken sabit olmasına rağmen denekler daima tesadüfî olarak düşünülür.

## Karmaşık Tekrarlı Ölçüm Desenleri İçin Basit Etkiler

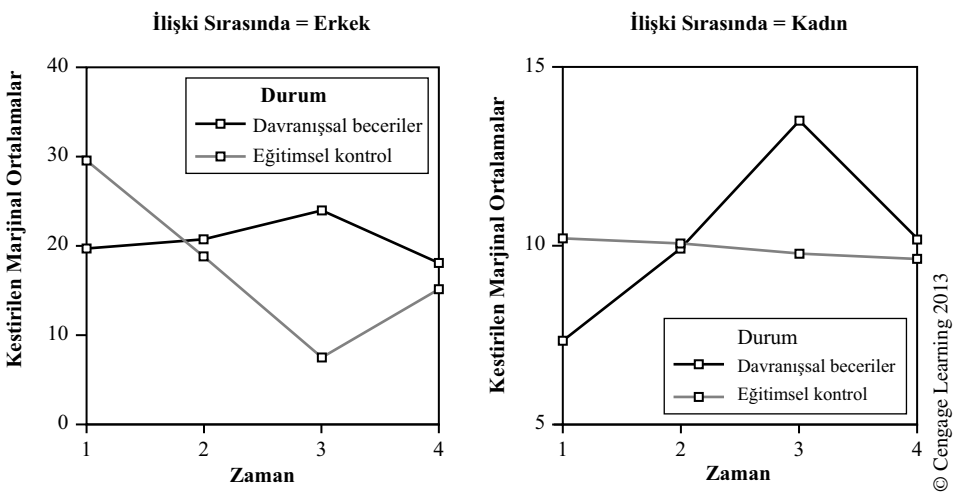
Bir önceki örnekte denekler-içi etkiler üzerinde yapılan testlerin küresellik varsayımı ihlâlleri yüzünden bazen sekteye uğradığı görülmüş ve bu problemin üstesinden gelmek için adımlar atılmıştı. Şimdi incelenecek örnekte de çoğunlukla aynı problem mevcuttur.

Şekil 14.3'te grafiklerini gördüğünüz hücre ortalamaları, iki uygulama koşulu altında, prezervatif kullanım sıklığının zamanla değişimini kadınlar ve erkekler için ayrı ayrı göstermektedir. Şekilden de anlaşılacağı gibi, verinin açıkladığı durum basit değildir.

Burada yine denekler arası faktörler üzerindeki basit etkiler ile denekler-içi faktörler üzerindeki basit etkiler arasında bir ayrım yapmamız gerekecektir. Denekler-arası basit etkilerle işe başlayalım. İncelenebilecek üç farklı denekler-arası basit etki vardır - Zaman değişkeninin her düzeyi için Durum ve Cinsiyet'in basit ana etkileri ve yine Zaman'ın her düzeyi için Cinsiyet × Koşul basit etkileşim etkisi. Örneğin, ön testte iki koşul (DBE ve Eğt. Kon.) arasında fark olmadığı beklentimizi kontrol etmek isteyebiliriz. Ayrıca İT6 veya İT12'de de aralarında fark olduğu beklentimizi test etmek isteyebiliriz. Burada aslında, Tablo 14.7'den manidar olduğunu bildiğimiz Durum × Zaman etkileşim etkisi ayrıntılı biçimde incelenmektedir.

Denekler-arası etkileri test etmenin en iyi yolu, Zaman değişkeninin her düzeyi için ayrı bir iki-yönlü (Durum × Cinsiyet) analiz yapılmasıdır. Bu dört analiz, küçük bir çaba ile her Zaman düzeyindeki üç basit etkinin tamamını elde etmemizi sağlar. Böylece, burada kolaylık olsun diye yapıldığı gibi, siz de bu analizlerden elde edilen *F* değerlerini kabul edebilir veya dört ayrı analizin hata terimlerini birleştirip elde ettiğiniz birleştirilmiş hata terimini söz konusu etki için kareler ortalamasını test ederken kullanabilirsiniz. Eğer bu terimler heterojen ise, bunları birleştirmek akıllıca olmaz. Diğer taraftan terimler eğer homojen varyans kaynaklarını temsil ediyorsa, birleştirilebilir; böylece hata için daha fazla serbestlik derecesi elde edilir. Bu etkiler için küreselliği göz önüne almaya gerek yoktur çünkü her basit etki tekrarlı ölçüm değişkeninin sadece bir düzeyinde ölçülür. (Burada bir uyarı yapmak gerekir. Zaman değişkeninin her düzeyinde Durum × Cinsiyet etkileşimine bakabiliyor olmanız, bakmanız gerektiği anlamına gelmez. Ne kadar çok test yapılırsa, I. Tip hata olasılığı o kadar artar. Bu yüzden sadece araştırma soruları için önemli olan etkiler test edilmelidir.)

Denekler-arası basit etkiler de neredeyse aynı şekilde ele alınır. Örneğin, tam müdahalenin olmadığı durumda Eğt. Kon. Durumunun zaman içinde değişip değişmediğini görmek için, her Durum için ayrı ayrı Zaman değişkenine ait basit etkilere bakmak gerekebilir. Benzer şekilde, DBE durumunun zaman içinde nasıl değiştiği incelenebilir. Fakat hata




© Cengage Learning 2013

**Şekil 14.3** Cinsiyet ve Durum fonksiyonlarına göre prezervatif kullanma frekansları

**Tablo 14.9** Ana etkiler analizi

(a) **Ön testte denekler arası etkiler (Durum, Cinsiyet ve Durum x Cinsiyet)**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Durum | 1 | 403,225 | 403,225 | 1,45 |
| Cinsiyet | 1 | 2544,025 | 2544,025 | 9,13* |
| Durum x Cinsiyet | 1 | 119,025 | 119,025 | 0,43 |
| Hata | 36 | 10027,100 | 278,530 | |
| Toplam | 39 | 13093,375 | | |

(b) **DBÖ durumunda denekler içi etkiler (Cinsiyet, Zaman, Zaman x Cinsiyet)**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 19 | 7849,13 | | |
| Cinsiyet | 1 | 2173,61 | 2173,61 | 6,89* |
| Hata (denekler arası) | 18 | 5675,52 | 315,30 | |
| Denekler içi | 60 | 3646,26 | | |
| Zaman | 3 | 338,94 | 112,98 | 1,88 |
| Zaman x Denekler | 3 | 54,54 | 18,18 | 0,30 |
| Hata (denekler içi) | 54 | 3252,78 | 60,24 | |
| Toplam | 79 | 11495,39 | | |

*$p < .05$

© Cengage Learning 2013

terimini gereksiz yere artırmamak için Cinsiyet değişkeninin bir etki olarak her iki analize de eklenmesi gerekir. Durumlar değişkenine ait hata terimlerini birleştirmek yerine, her bir analiz için farklı bir hata terimi kullanılması tercih edilir.

Diğer değişkenin bir düzeyindeki basit etkilere ait analizler Tablo 14.9'da verilmiştir. Diğer düzeyler için de benzer şekilde testler yapılır, fakat burada yer kaplamaması için verilmemiştir. Bu tabloda, ilgilendiğimizden daha fazla basit etki olmasına rağmen, testlerin nasıl inşa edildiğini aydınlatması açısından az bir kısmı sunulmuştur. Bu yaklaşımla yapılabilecek tüm testleri göz önünde bulundurmak akıllıca olmayacaktır, çünkü bu grup aile-boyu hata oranını ciddi biçimde artırır. Analize başlamadan önce neye bakmak istediğimize karar vermeli ve bu karara sadık kalmalıyız. Eğer gerçekten çok sayıda basit etkiye bakmak isteniyorsa, Bölüm 12'de sözü edilen Bonferroni yaklaşımlarından biri kullanılmalıdır.

Tablo 14.9a'daki denekler-arası analizde Zaman değişkeninin 1. düzeyinde (ön testte) kadın ve erkek arasında manidar bir fark olduğu görülmekte (kadınların kullanma sıklığı daha düşüktür), fakat ne Durum etkisi, ne de Durum x Cinsiyet etkileşimi gözlenmemektedir. Her Durum düzeyinde, erkekler kadınları aşağı yukarı aynı miktarda geçmişlerdir. Hiç Durum etkisi olmaması güven vericidir, çünkü farklı durumlardaki denekler arasında uygulama öncesi fark olsaydı içimiz rahat etmezdi.

Tablo 14b'deki sonuçlardan, DBE durumunda, yine cinsiyete göre manidar bir farklılık olduğunu, fakat Zaman etkisinin ve Zaman × Cinsiyet etkileşiminin olmadığını görmekteyiz. Bu sonuç hayal kırıklığı yaratabilir, çünkü cinsiyet değişkenine göre ortalama aldığımızda, yapılan uygulamanın prezervatif kullanım sıklığında bir değişikliğe sebep olmadığını gösterir. Bu durum, genel analiz sonucunda gözlenen manidar Durum × Zaman etkileşimine dayanarak varılabilecek sonuçla çelişmektedir, bu da basit etkileri incelemenin değerli olduğunu gösterir. Görmek istediğimiz etkinin manidar olması, manidarlığın istediğimiz yönde olacağı anlamına gelmeyebilir.

## 14.9 İki Denekler-İçi Değişken ve Bir Denekler-Arası Değişken

Biraz önce incelediğimiz desen bir denekler-içi, bir denekler-arası değişken olduğu durumun basit bir uzantısı olarak düşünülebilir. Özet tabloya eklememiz gereken, bir başka ana etki ve ona karşılık gelen etkileşimlerdir. Buna karşın iki denekler-içi ana etkisi olan bir deseni ele aldığımızda, yeni hata terimlerinin varlığı nedeniyle problem biraz daha karmaşık hale gelir. Daha genel bir gösterim kullanarak bağımsız değişkenleri *A*, *B* ve *C* olarak etiketleyelim.

Bir önceki çalışmanın bir çeşitlemesi olarak, *A* değişkeninin (Cinsiyet) iki düzeyi için farklı denekler kullandığımızı, fakat her deneği *B* (durum) ve *C* (Uygulamalar) değişkenlerinin tüm kombinasyonlarına tabi tuttuğumuzu varsayalım. Desene ait diyagram aşağıdaki gibi gösterilebilir:

| | $A_1$ | | | $A_2$ | | |
|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | $C_1$ | $C_2$ | $C_3$ |
| $B_1$ | $G_1$ | $G_1$ | $G_1$ | $G_2$ | $G_2$ | $G_2$ |
| $B_2$ | $G_1$ | $G_1$ | $G_1$ | $G_2$ | $G_2$ | $G_2$ |
| $B_3$ | $G_1$ | $G_1$ | $G_1$ | $G_2$ | $G_2$ | $G_2$ |

© Cengage Learning 2013

Örneğe geçmeden önce, bu desene ait beklenen kareler ortalamalarını inceleyelim. Denekler dışında tüm faktörlerin sabit tutulduğu (denekler tesadüfî faktör olarak belirlenmiştir) model durumu için bu değerler Tablo 14.10'da gösterilmiştir. Beklenen kareler ortalamalarından, bu desen için dört hata terimimiz olacağı anlaşılmaktadır. Daha önce olduğu gibi, $KO_{\text{Kt gruplar içi}}$ denekler-arası etkiyi test etmek için kullanılır. Denekler-içi terimlere geldiğimizde, *B* ve *B*'nin etkileşimi ile *AB* x *Kt* gruplar içi ile; *C* ve *C*'nin etkileşimi ile *AC x Kt* gruplar içi ile; *BC* ve *BC*'nin etkileşimi ile *ABC* x *Kt* gruplar içi ile test edilmektedir. Bunların neden gerekli olduğu, beklenen kareler ortalamalarından anlaşılmaktadır.

**Tablo 14.10** Beklenen kareler ortalamaları

| Kaynak | *sd* | *E(KO)* |
|---|---|---|
| Denekler arası | $an - 1$ | |
| *A* (Gruplar) | $a - 1$ | $\sigma_e^2 + bc\sigma_\pi^2 + nbc\theta_\alpha^2$ |
| Gruplar içi Kt | $a(n - 1)$ | $\sigma_e^2 + bc\sigma_\pi^2$ |
| Denekler içi | $na(bc - 1)$ | |
| *B* | $b - 1$ | $\sigma_e^2 + c\sigma_{\beta\pi}^2 + nac\theta_\beta^2$ |
| *AB* | $(a - 1)(b - 1)$ | $\sigma_e^2 + c\sigma_{\beta\pi}^2 + nc\theta_{\alpha\beta}^2$ |
| *B* × Gruplar içi Kt | $a(b - 1)(n - 1)$ | $\sigma_e^2 + c\sigma_{\beta\pi}^2$ |
| *C* | $c - 1$ | $\sigma_e^2 + b\sigma_{\gamma\pi}^2 + nab\theta_\gamma^2$ |
| *AC* | $(a - 1)(c - 1)$ | $\sigma_e^2 + b\sigma_{\gamma\pi}^2 + nb\theta_{\alpha\gamma}^2$ |
| *C* × Gruplar içi Kt | $a(c - 1)(n - 1)$ | $\sigma_e^2 + b\sigma_{\gamma\pi}^2$ |
| *BC* | $(b - 1)(c - 1)$ | $\sigma_e^2 + n\sigma_{\beta\gamma\pi}^2 + na\theta_{\beta\gamma}^2$ |
| *ABC* | $(a - 1)\,(b - 1)(c - 1)$ | $\sigma_e^2 + n\sigma_{\beta\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$ |
| *BC* × Gruplar içi Kt | $a(b - 1)(c - 1)(n - 1)$ | $\sigma_e^2 + n\sigma_{\beta\gamma\pi}^2$ |
| Toplam | $N - 1$ | |

© Cengage Learning 2013

## Koşullu Bastırma ile İlgili Bir Veri Analizi

Radyolu saatinizin yüksek ve sinir bozucu sesli alarmı çalmadan hemen önce saatten minik bir "tık" sesi geldiğini varsayalım. Zamanla, bu "tık" sesi (psikologlar bunu Koşullu Uyaran -*KU*- olarak adlandırır) normalde alarma verilen tepkileri ortaya çıkarabilir (Koşulsuz Uyaran - *KsU*). (Kolejde bir oda arkadaşım hiç saat kurmazdı çünkü saatin "tık" sesi onu uyandırırdı.) Dahası, sadece "tık" sesini sunmak, sonrasında alarm duyulmasa bile, süregelen bir davranışın bastırılmasına neden olabilir. (Eğer orada yatarken bir şey okuyorsanız, okumaya ara verebilirsiniz.) Tık sesinin süregelen bir davranışı nasıl etkilediğini (bastırdığını) araştıran bir laboratuar çalışmasında Bouton ve Swartzentruber (1985) önceden bir şok ile eşleştirilmiş olan ses tonunun farelerin süregelen kaldıraca basma tepki oranını ne derecede bastırdığını incelemişlerdir. Bastırma, ses tonunu izleyen 1 dakikalık test süresinde kaldıraca basma sayısının bir temel düzey ve test süreleri boyunca toplam kaldıraca basma sayılarının toplamına oranı olarak ölçülmüştür. Tüm gruplar için davranış iki evrede değerlendirilmiştir: Şok evresi (ses tonunun ardından şok verilmesi) ve şoksuz evre (ses tonundan sonra şok verilmemesi). Bu evreler dörder kez tekrarlanarak deneyin dört Döngüsünü oluşturmuştur.

İlk önce Tablo 14.11'e bir göz atmak araştırma deseninin anlaşılmasını kolaylaştıracaktır. Birinci Evre'de *A-B* Grubu *A* Kutusu'na konulmuştur. Hayvanın yiyecek için kaldıraca bastığı 1 dakikalık temel düzey süresinden sonra, 1 dakika boyunca ses tonu verilmiş ve bunu hafif bir şok izlemiştir. Ses tonu varken kaldıraca basma tepkisinin bastırılma derecesi (normal korku tepkisi) kaydedilmiştir. Daha sonra döngünün ikinci evresi için hayvan *B* Kutusu'na konulmuştur. Bu safhada 1 dakikalık başlangıç seviyesi kaldıraca basma hareketinden sonra sadece ses tonu uyarıcı olarak verilmiştir. Ses tonuna daha önceden şok eşlik ettiği için kaldıraca basma davranışının kısmen bastırılması beklenir. *A-B* döngülerinin bir süre tekrarlanmasıyla, hayvan ikinci evrede hiç şok verilmediğini, yani *B* Kutusu'nun "güvenli kutu" olduğunu öğrenir. Böylece, sonraki döngülerde şok olmayan uygulamalarda daha az bastırma olması beklenir.

*L-A-B* grubuna yapılan uygulama *A-B* grubuna yapılanın tek bir fark dışında aynısıdır: Bu gruptaki hayvanlar daha önceden şoka ses tonu yerine ışığın eşlik ettiği bir deneyim yaşamışlardır. Bu deneyimlerinden dolayı, araştırmacılar hayvanları özellikle ilk bir-iki döngüde diğer gruptan biraz daha iyi bir performans (ikinci evrede daha az bastırma) göstermelerini bekliyorlardı.

*A-A* Grubu'na yapılan uygulamada da *A-B* Grubu ile aynı yöntem izlenmiştir, fakat bu seferki evrede aynı kutu içerisinde -*A* Kutusu'nda- gerçekleştirilmiştir. Test kutuları arasında ipucu olacak farklılıklar olmadığından (yani hayvanların şok evresi ile şoksuz evreyi ayırt etmesi mümkün olmadığından) şoksuz evrede en çok bu gurubun bastırma göstermesi beklenmelidir.

Bouton ve Swartzentruber, genele bakıldığında şu değişkenlere ait ana etkilerin olacağını öngörmüşlerdi: Evre (yani şok evresi ve şoksuz evre arasında bir fark olacağı), Grup (*A-B* ve *L-A-B* gruplarında *A-A* grubuna göre daha az bastırma gözleneceği) ve Döngü (*B* kutusuna konulan hayvanların zaman içinde bu kutunun güvenli bir yer olduğunu öğreneceği). Araştırmacılar ayrıca her bir etkileşimin manidar olacağını da öngörmüşlerdi. (Bu örneğin seçilmesinin bir sebebi, açık olarak tanımlanması zor da olsa, tüm etkilerin manidar ve akla yatkın olacağını öngören ender çalışmalardan olmasıdır.)

Bu çalışmaya ait veri ve varyans analizi Tablo 14.11'de verilmiştir. Analizin ayrıntılarına girilmemiştir çünkü çoğunlukla daha önceden nasıl yapılacağını bildiğiniz adımları içermektedir. Kolaylık olması bakımından sonuçlar grafik olarak Şekil 14.4'te gösterilmiştir. Grafiklerden, sonuçların çoğunun gayet açık ve öngörülen yönde olduğu anlaşılmaktadır. Bu veri için düşük puanın daha fazla bastırma, yani hayvanların daha yavaş tepki vermesi anlamına geldiğini hatırlatalım.

**Tablo 14.11** Koşullanmış bastırma analizi (Daha düşük puanlar daha çok bastırmayı temsil eder.)

($a_1$) Data

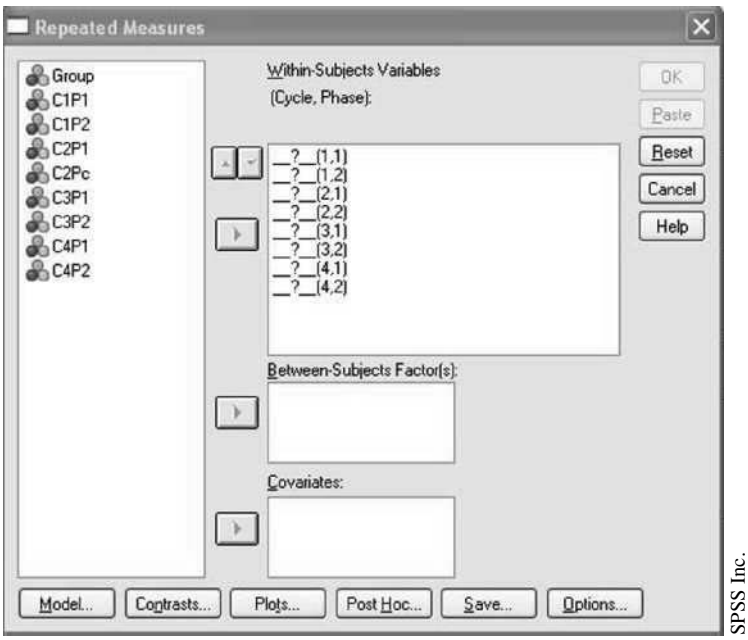
| | Döngü | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | | 4 | | |
| | Evre | | Evre | | Evre | | Evre | | Denek |
| Grup | I | II | I | II | I | II | I | II | Ortalaması |
| A-B | 1* | 28 | 22 | 48 | 22 | 50 | 14 | 48 | 29,125 |
| | 21 | 21 | 16 | 40 | 15 | 39 | 11 | 56 | 27,375 |
| | 15 | 17 | 13 | 35 | 22 | 45 | 1 | 43 | 23,875 |
| | 30 | 34 | 55 | 54 | 37 | 57 | 57 | 68 | 49,000 |
| | 11 | 23 | 12 | 33 | 10 | 50 | 8 | 53 | 25,000 |
| | 16 | 11 | 18 | 34 | 11 | 40 | 5 | 40 | 21,875 |
| | 7 | 26 | 29 | 40 | 25 | 50 | 14 | 56 | 30,875 |
| | 0 | 22 | 23 | 45 | 18 | 38 | 15 | 50 | 26,375 |
| **Ortalama$_{AB}$** | 12,625 | 22,750 | 23,500 | 41,125 | 20,000 | 46,125 | 15,625 | 51,750 | 29,188 |
| A-A | 1 | 6 | 16 | 8 | 9 | 14 | 11 | 33 | 12,250 |
| | 37 | 59 | 28 | 36 | 34 | 32 | 26 | 37 | 36,125 |
| | 18 | 43 | 38 | 50 | 39 | 15 | 29 | 18 | 31,250 |
| | 1 | 2 | 9 | 8 | 6 | 5 | 5 | 15 | 6,375 |
| | 44 | 25 | 28 | 42 | 47 | 46 | 33 | 35 | 37,500 |
| | 15 | 14 | 22 | 32 | 16 | 23 | 32 | 26 | 22,500 |
| | 0 | 3 | 7 | 17 | 6 | 9 | 10 | 15 | 8,375 |
| | 26 | 15 | 31 | 32 | 28 | 22 | 16 | 15 | 23,125 |
| **Ortalama$_{AA}$** | 17,750 | 20,875 | 22,375 | 28,125 | 23,125 | 20,750 | 20,250 | 24,250 | 22,188 |
| L-A-B | 33 | 43 | 40 | 52 | 39 | 52 | 38 | 48 | 43,125 |
| | 4 | 35 | 9 | 42 | 4 | 46 | 23 | 51 | 26,750 |
| | 32 | 39 | 38 | 47 | 24 | 44 | 16 | 40 | 35,000 |
| | 17 | 34 | 21 | 41 | 27 | 50 | 13 | 40 | 30,375 |
| | 44 | 52 | 37 | 48 | 33 | 53 | 33 | 43 | 42,875 |
| | 12 | 16 | 9 | 39 | 9 | 59 | 13 | 45 | 25,250 |
| | 18 | 42 | 3 | 62 | 45 | 49 | 60 | 57 | 42,000 |
| | 13 | 29 | 14 | 44 | 9 | 50 | 15 | 48 | 27,750 |
| **Ortalama$_{LAB}$** | 21,625 | 36,250 | 21,375 | 46,875 | 23,750 | 50,375 | 26,375 | 46,500 | 34,141 |
| **Toplam** | 17,333 | 26,625 | 22,417 | 38,708 | 22,292 | 39,083 | 20,750 | 40,833 | 28,505 |

*Ondalık kısım tabloda gösterilmemiştir fakat hesaplamalara dahil edilmiştir.

© Cengage Learning 2013

Çok az insanın sonuna kadar sabredip inceleyeceği, üç sayfa sürecek tablo ve hesaplamaları buraya koymak yerine analizin SPSS'te yapılması tercih edilmiştir.[6] Veri programa Tablo 14.11'de göründüğü gibi, solda gruplar için bir sütun açılarak girilir. Açılan listelerden sırası ile Analyze, General Linear Model, RepeatedMeasures seçilerek iki tekrarlı ölçüm olduğu belirtilir [4 düzeyi olan Döngü (Cycle) ve 2 düzeyi olan Evre (Phase)]. Daha sonra Define tıklanarak her bir hücre için ilişkili değişkenler ve denekler arası faktörleri (Between-Subject Factor(s)) tanımlayan değişken(ler) belirlenir. Beliren diyalog kutusunda C1P1 – C4P2 (Döngü1Evre1 - Döngü4Evre2) etiketli değişkenler denekler-içi değişkenler (Within-Subjects Variables) kutusuna,"Grup" (Group) değişkeni ise denekler arası faktör(ler) (Between-Subjects Factors) kutusuna taşınır.

6 Hesaplamaları görmek isteyenler, önceki basımlardaki ilgili sayfalara www.uvm.edu/~dhowell/methods8/Supplements adresinden ulaşabilirler.

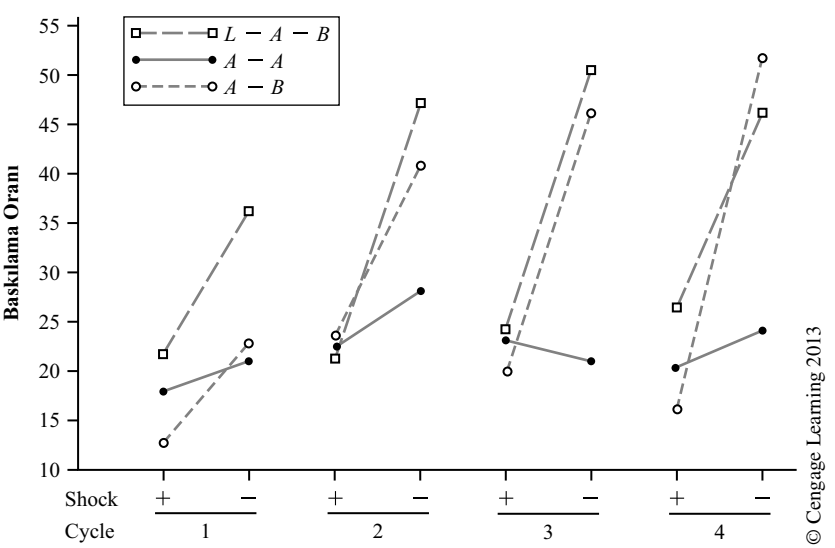



Diyalog kutusunun alt satırında hangi grafiklerin görülmesi, hangi karşılaştırmaların yapılması ve hangi betimsel istatistiklerin çıktısının alınması isteniyorsa belirtilir. Sonra *OK* tıklanarak analiz çalıştırılır.

Kısaltılmış özet tablo bir sonraki sayfada yer almaktadır. Dikkat ederseniz SPSS programı Denekler-Arası ve Denekler-İçi faktörler için farklı tablolar üretiyor. Ben ikisinin bir tabloda, uygun girintiler kullanılarak verilmesini tercih ederim.

Tabloda birden fazla hata terimi gözüktüğüne dikkat edin. Grup etkisi, Denekler-Arası tablosundaki Hata terimi ile test ediliyor. Sonra, Döngü ve Döngü x Grup, Hata (Döngü) ile, Evre ve Evre x Grup Hata (Evre) ile, Döngü x Evre ve Döngü x Evre x Grup ise Hata (Döngü x Evre) ile test edilmektedir.

Tablo 14.12'de verilen özet tabloya bakıldığında, neredeyse tüm öngörülerin desteklendiği açıktır. Manidar olmayan tek etki Grup değişkeninin ana etkisidir, fakat bu etki gerekli değildir çünkü şok evresi ve şoksuz evre üzerinde bir ortalamayı temsil etmektedir. Araştır-




**Şekil 14.4** Koşullanmış baskılama verisi

**Tablo 14.12** Koşullu bastırılmış veri analizi SPSS çıktısı

**Denekler Arası Etkiler Testi**

**Ölçü: ÖLÇÜM_1**
**Dönüştürülmüş Değişken: Ortalama**

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | F | Manidarlık | Kısmî Eta Kare |
|---|---|---|---|---|---|---|
| Kesişme | 156009,005 | 1 | 156009,005 | 208,364 | ,000 | ,908 |
| Grup | 4616,760 | 2 | 2308,380 | 3,083 | ,067 | ,227 |
| Hata | 15723,359 | 21 | 748,731 | | | |

**Denekler İçi Etkilere Ait Testler**

**Ölçüm: Ölçüm_1**

| Kaynak | | Tip III Kareler Toplamı | sd | Kareler Ortalaması | F | Man. Düzeyi | Kısmî Eta Kare |
|---|---|---|---|---|---|---|---|
| Döngü | Küresellik Varsayıldı | 2726,974 | 3 | 908,991 | 12,027 | ,000 | ,364 |
| | Greenhouse-Geisser | 2726,974 | 2,362 | 1154,735 | 12,027 | ,000 | ,364 |
| | Huynh-Feldt | 2726,974 | 2,934 | 929,571 | 12,027 | ,000 | ,364 |
| | Alt Sınır | 2726,974 | 1,000 | 2726,974 | 12,027 | ,002 | ,364 |
| Döngü*Grup | Küresellik Varsayıldı | 1047,073 | 6 | 174,512 | 2,309 | ,044 | ,180 |
| | Greenhouse-Geisser | 1047,073 | 4,723 | 221,691 | 2,309 | ,061 | ,180 |
| | Huynh-Feldt | 1047,073 | 5,867 | 178,463 | 2,309 | ,046 | ,180 |
| | Alt Sınır | 1047,073 | 2,000 | 523,536 | 2,309 | ,124 | ,180 |
| Hata (Döngü) | Küresellik Varsayıldı | 4761,328 | 63 | 75,577 | | | |
| | Greenhouse-Geisser | 4761,328 | 49,593 | 96,009 | | | |
| | Huynh-Feldt | 4761,328 | 61,605 | 77,288 | | | |
| | Alt Sınır | 4761,328 | 21,000 | 226,730 | | | |
| Evre | Assumed | 11703,130 | 1 | 11703,130 | 129,855 | ,000 | ,861 |
| | Greenhouse-Geisser | 11703,130 | 1,000 | 11703,130 | 129,855 | ,000 | ,861 |
| | Huynh-Feldt Lower- | 11703,130 | 1.000 | 11703,130 | 129,855 | ,000 | ,861 |
| | bound Sphericity | 11703,130 | 1,000 | 11703,130 | 129,855 | ,000 | ,861 |
| Evre*Grup | Küresellik Varsayıldı | 4054,385 | 2 | 2027,193 | 22,493 | ,000 | ,682 |
| | Greenhouse-Geisser | 4054,385 | 2,000 | 2027,193 | 22,493 | ,000 | ,682 |
| | Huynh-Feldt | 4054,385 | 2,000 | 2027,193 | 22,493 | ,000 | ,682 |
| | Alt Sınır | 4054,385 | 2,000 | 2027,193 | 22,493 | ,000 | ,682 |
| Hata*Evre | Küresellik Varsayıldı | 1892,609 | 21 | 90,124 | | | |
| | Greenhouse-Geisser | 1892,609 | 21,000 | 90,124 | | | |
| | Huynh-Feldt | 1892,609 | 21,000 | 90,124 | | | |
| | Alt Sınır | 1892,609 | 21,000 | 90,124 | | | |
| Döngü*Evre | Küresellik Varsayıldı | 741,516 | 3 | 247,172 | 4,035 | ,011 | ,161 |
| | Greenhouse-Geisser | 741,516 | 2,137 | 346,969 | 4,035 | ,022 | ,161 |
| | Huynh-Feldt | 741,516 | 2,613 | 283,767 | 4,035 | ,015 | ,161 |
| | Alt Sınır | 741,516 | 1,000 | 741,516 | 4,035 | ,058 | ,161 |
| Döngü*Evre* | Küresellik Varsayıldı | 1273,781 | 6 | 212,297 | 3,466 | ,005 | ,248 |
| Grup | Greenhouse-Geisser | 1273,781 | 4,274 | 298,013 | 3,466 | ,013 | ,248 |
| | Huynh-Feldt | 1273,781 | 5,226 | 243,728 | 3,466 | ,008 | ,248 |
| | Alt Sınır | 1273,781 | 2,000 | 636,891 | 3,466 | ,050 | ,248 |
| Hata (Döngü* | Küresellik varsayıldı | 3859,078 | 63 | 61,255 | | | |
| Evre) | Greenhouse-Geisser | 3859,078 | 44,880 | 85,987 | | | |
| | Huynh-Feldt | 3859,078 | 54,875 | 70,324 | | | |
| | Alt Sınır | 3859,078 | 21,000 | 183,766 | | | |

Adapted from output by SPSS, Inc.

macılar da şok evresinde gruplar arasında hiç fark olmayacağını veya çok küçük bir fark olacağını öngörmüşlerdir. Bu bağlamda Evre x Grup etkileşimi daha ilgi çekicidir ve onun da manidar çıktığı açıkça görülmektedir.

## 14.10 Sınıflar İçi Korelasyon

Hesaplarla dolu karmaşık desenleri bir kenara bırakıp daha rahat baş edilebilir bir konuya döneceğiz. Herhangi bir alanda deney (araştırma) deseni oluşturulurken önem verilen konulardan biri ölçüm araçlarının güvenirliğidir. Birçok kişi güvenirlik konusuna rastlanabilecek son yerin varyans analizinin tartışıldığı bölüm olduğunu düşünebilir, fakat bakılması gereken yer tam burasıdır. (Sınıf içi korelasyon ile ilgili daha ayrıntılı bilgi için http//www.uvm.edu/~dhowell/StatPages/More_Stuff/icc/icc.html adresine bakınız.)

Değerlendirmecilerin küçük çocuklarda olumlu sosyal davranışı derecelendirmelerine dair güvenirliği ölçmekle ilgilendiğimizi varsayalım. Güvenirliği araştırmak için iki veya daha fazla değerlendirmecinin her birinden bir grup çocuğa ait davranış örneklemini derecelendirmeleri istenebilir. Değerlendirmeciler, her bir davranış örneklemindeki olumlu sosyal davranış miktarını yansıtmak üzere 1 ile 10 arasında bir puan verirler. Bu uygulamaya örnek vermek için, bir noktaya dikkat çekmek amacıyla abartılmış veri kullanılacaktır. Şimdi Tablo 14.13'teki veri kümesine bakalım.

Tablo 14.13a'da değerlendirmeciler neredeyse mükemmel uyum içindeler. Hepsi çocuklar arasında büyük farklar görüyor, hangi çocuğun yüksek düzeyde, hangi çocuğun düşük düzeyde olumlu sosyal davranış gösterdiği konusunda aynı düşüncedeler ve bu düzeylerin ne kadar düşük, ne kadar yüksek olduğu konusunda da neredeyse aynı düşüncedeler. Bu durumda, veri kümesindeki değişkenliklerin neredeyse tamamı çocuklar arasındaki farklardan kaynaklanır; değerlendirmeciler arasında hemen hemen hiç fark görülmez ve neredeyse hiç tesadüfî hata yoktur.

Tablo 14.13b'de ise benzer bir durum söz konusu olsa da bir fark görüyoruz. Değerlendirmeciler, genel olarak çocuklar arasındaki farkları görüyorlar ve hangi çocukların davranışın en yüksek (ve en düşük) düzeylerini gösterdikleri konusunda aynı düşüncededirler. Fakat değerlendirmeciler, gördükleri olumlu sosyal davranışın miktarında anlaşmazlık içindeler. İkinci değerlendirmeci birinciden biraz daha az davranış gözlüyor (ortalaması 1 puan daha düşük); üçüncü değerlenmeci ise diğer ikisine göre daha fazla davranış gözlüyor. Başka bir deyişle, değerlendirmeciler çocukların sıralaması konusunda aynı fikirdeler fakat düzeyler konusunda fikir ayrılıkları var. Bu durumda veri hem çocuklar arasında, hem de değerlendirmeciler arasında değişkenlik gösterir. Buna rağmen, tesadüfî hata bileşeni hâlâ çok küçüktür. Genelde, insanların davranışı nasıl derecelendirdiğine dair en gerçekçi model budur, çünkü her birimiz, örneğin 7 puan hak etmek için, ne kadar davranış göstermemiz gerektiği konusunda farklı anlayışa sahibizdir. Bir derecelendirme sisteminin güvenirliğini değerlendirirken normalde değerlendirmeciler arası değişkenliği göz önüne almamız gerekir.

**Tablo 14.13** Sınıf içi korelasyon örneklerine ait veri

| | (a) Değerlendirmeci | | | (b) Değerlendirmeci | | | (c) Değerlendirmeci | | |
|---|---|---|---|---|---|---|---|---|---|
| Çocuk | I | II | III | I | II | III | I | II | III |
| 1 | 1 | 1 | 2 | 1 | 0 | 3 | 1 | 3 | 7 |
| 2 | 3 | 3 | 3 | 3 | 2 | 5 | 3 | 1 | 5 |
| 3 | 5 | 5 | 5 | 5 | 4 | 7 | 5 | 7 | 4 |
| 4 | 5 | 6 | 6 | 5 | 4 | 7 | 5 | 5 | 5 |
| 5 | 7 | 7 | 7 | 7 | 6 | 8 | 7 | 6 | 7 |

© Cengage Learning 2013

Son olarak Tablo 14.13c'ye baktığımızda, değerlendirmecilerin sadece düzeyde değil, çocukların sıralanmasında da fikir ayrılığına düştüğünü görüyoruz. Bu verideki değişkenliğin büyük bir kısmı hata değişkenliğidir. O zaman güvenirlikten söz etmek istediğimizde ne yapmamız gerekir? Değerlendirmecilerin sadece birkaç düzey veya kategori kullandığı durumlarda güvenirliği ölçmenin bir yolu iki değerlendirmecinin aynı puanlamayı yaptığı durumların yüzdesini hesaplamaktır, fakat bu ölçüm yanlıdır çünkü bir veya iki kategorinin üstünlüğü durumunda, şansa bağlı fikir birliği düzeyi yüksek olur. (Cohen kapa hakkındaki bölüme bakınız.) Yaygın başka bir yaklaşım ise iki değerlendirmecinin verdiği puanlar arasında korelasyon hesabı yapmak ya da eğer ikiden fazla değerlendirmeci varsa ikili korelasyonların ortalamasını almaktır. Ama bu yöntem değerlendirmeciler arasındaki farkları hesaba katmayacaktır. (Eğer bir değerlendirmeci her zaman diğerinden 5 puan fazla veriyorsa korelasyon 1,00 olur, fakat bu durumda aslında değerlendirmeciler denekler hakkında farklı şeyler söylemektedirler). Üçüncü bir yol ise, değerlendirmeciler arası farkları göz önüne alan, **sınıf içi korelasyon** olarak adlandırılan hesabı yapmaktır. Şimdi bunun nasıl yapıldığına bakalım.

**sınıf içi korelasyon**

Sınıf içi korelasyon katsayısını farklı şekillerde hesaplayabilirsiniz. Bu değerlendirmecileri sabit ya da tesadüfî değişken olarak varsaymanıza ve değerlendirmecilerin aynı ya da farklı denekleri puanlıyor olmasına göre değişir. Sınıf içi korelasyon konusunun klâsik referansı Shrout ve Fleiss'ın (1979) birkaç alternatif yaklaşımdan söz ettikleri çalışmalarıdır. Burada sadece en yaygın yaklaşım olan, değerlendirmecilerin, kullanabileceğimiz tüm değerlendirmeciler arasından rastgele seçilmiş bir örneklem olduğu ve hepsinin aynı denek grubunu birer kez değerlendirdiği durumu ele alacağım. (Bu kısımda değerlendirmecilerin "denekleri" derecelendirdiği, fakat resim, araba veya şehirlerin yaşanabilirliğini de değerlendirebileceklerini varsayacağım.) Değerlendirilen her ne ise onun için "denek" genel terimi kullanılacaktır.

Tablo 14.13'teki veri kümesinin aşağıdaki model ile temsil edilebileceğini varsaymakla işe başlayacağız:

$$X_{ij} = \mu + \alpha_i + \pi_j + \alpha\pi_{ij} + e_{ij}$$

Bu modelde, *i*'yinci değerlendirmecinin etkisi $\alpha_i$, *j*'yinci deneğin (kişinin) etkisi $\pi_j$, *i*'yinci değerlendirmecinin *j*'yinci denek ile etkileşimi (değerlendiricinin denekle yüzleştirildiğinde puanlama sistemini ne derecede değiştirdiği) $\alpha\pi_{ij}$ ile temsil edilmekte, $e_{ij}$ de o puanlamaya ait hatayı gösterir. Her değerlendirmeci her deneğe bir kez puan verdiği için $\alpha\pi_{ij}$ ve $e_{ij}$ terimlerini ayrı ayrı kestirmak mümkün olmasa da modelde ayrı tutulmaları gerekir.

Bir önceki bölüme bakarsanız, ölçme etkisinin büyüklüğünü hesapladığımızda (aslında $r^2$-aile ölçüsü idi) ilgili etkinin (bu durumda denekler arası farklar) varyans kestirimi ile varyansın birkaç kaynağına dair tahminlerin toplamına oranlamıştık. Şimdi de tam olarak bunu yapacağız.

Sınıf içi korelasyon = $\sigma_\pi^2/(\sigma_\alpha^2 + \sigma_\pi^2 + \sigma_{\alpha\pi}^2 + \sigma_e^2)$

Eğer veri kümesindeki değişkenliğin çoğu denekler arasındaki farklardan ve sadece küçük bir kısmı değerlendirmeciler arası farklardan, değerlendirmeci-denek etkileşiminden ve hatadan kaynaklanıyorsa, bu oran 1,00'e yakın olur. Eğer değerlendirmeciler insanlara genel olarak verdikleri düşük-yüksek puanlar açısından birbirlerinden farklıysa veya değerlendirmeci-denek etkileşimi varsa (farklı değerlendirmeciler farklı insanlara farklı puanlar verirler) ya da puanlamalarda çok hata varsa payda paydan çok daha büyük olacak ve oran 1,00'den oldukça uzaklaşacaktır.

Sınıf içi büyüklüğünü hesaplamak için önce Denekler x Değerlendirmeciler için varyans analizi yapacağız. Tekrarlı ölçümümüz de Değerlendirmeciler olacak. Her Değerlendirmeci, her deneğe sadece bir kez puan verdiği için bir bağımsız hata tahmini olmayacak ve hata terimi olarak Denek x Değerlendirmeci etkileşimini kullanmamız gerekecek. Sonuçta oluşacak

özet tablo için, sınıf içi korelasyon tahminimizi $j$ değerlendirmeci sayısını, $n$ de denek sayısını göstermek üzere aşağıdaki eşitlikle hesaplayacağız:

$$\text{Sınıf içi korelasyon} = \frac{KO_{\text{denekler}} - KO_{\text{Değ.} \times \text{Den.}}}{KO_{\text{denekler}} + (j-1)KO_{\text{Değ.} \times \text{Den.}} + j(KO_{\text{Değ.}} - KO_{\text{Değ.} \times \text{Den.}})/n}$$

Örnek olarak, Tablo 14.13b'deki veri için varyans analizini yaptım, bu veri kümesi kasıtlı olarak hem denekler hem değerlendirmeciler arasında farklılık gösterecek şekilde düzenlenmiştir. Analizin özet tablosu aşağıda verilmiştir:

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler Arası | 4 | 57,067 | 14,267 | |
| Denekler İçi | 10 | 20,666 | 2,067 | |
| Değerlendirici | 2 | 20,133 | 10,067 | 150,25 |
| Değerlendirici × Denekler | 8 | 0,533 | 0,067 | |
| Toplam | 14 | 77,733 | | |

© Cengage Learning 201

Şimdi sınıf içi korelasyonu aşağıdaki gibi hesaplayabiliriz:

$$\text{Sınıf içi korelasyon} = \frac{14,267 - 0,067}{14,267 + (3-1)0,067 + 3(10,067 - 0,067)/5}$$

$$= \frac{14,200}{14,267 + 0,134 + 6} = \frac{14,2}{20,401} = ,70$$

Böylece güvenirlik ölçümüzü ,70 olarak elde ettik; bu muhtemelen görmek istediğimiz kadar iyi değil. Ama hesaplamalara bakınca, güvenirliğin düşük olmasının en büyük nedeninin hata değil değerlendirmeciler arası farklar olduğu gözüküyor. Buradan çıkarılabilecek sonuç şudur: Değerlendirmeciler bir araya gelip tutarlı bir ölçüme karar vermeliler; örneğin, "7" her değerlendirmeci için aynı şeyi ifade etmeli.

# 14.11 Tekrarlı Ölçüm Analizlerine Dair Diğer Düşünceler

## Sıra Etkisi

**sıra etkileri (sequence effects)**

**aktarma etkileri (carryover effects)**

Tekrarlı ölçüm desenleri **sıra etkileri**ne **(sequence effects)** ve **aktarma etkileri**ne **(carryover effects)** açık olması ile ünlüdür. Bir uygulamaya katılmış olmanın diğer uygulamaya verilecek tepkiyi etkileyebileceği durumlarda araştırmacının tekrarlı-ölçüm desenini kullanmaya karar verirken çok dikkatli olması gerekir. Bazı çalışmalarda aktarma etkisi özellikle tercih edilir. Örneğin, öğrenme ile ilgili araştırmalarda ana veri, bir uygulamadan diğerine aktarılanları temsil eder. Diğer durumlarda ise bu etkiler (özellikle ayırıcı aktarma etkileri) sıkıcı bir problemdir, bunlardan kaçınılması gerekir.

## Eşit Olamayan Grup Büyüklükleri

Tekrarlı ölçüm desenlerinin hoş özelliklerinden biri, deneklerden biri bir uygulamaya gelmediğinde, o kişiye ait diğer hücrelerin de boş bırakılmasıdır. Bu sayede hücre büyüklüklerinin eşit olmasa bile orantılı kalması sağlanır. Eğer elinizdeki bir deneğe ait verinin bir kısmı eksik ise, yaygın yöntem bu deneği analizden tamamen çıkarmaktır. Ama eğer sadece bir veya iki puan eksik ise bu puanların yerine yaklaşık (tahminî) değerler konulması mümkündür ve bu çoğu durumda tatmin edici bir yaklaşımdır. Bu konu ile ilgili ayrıntılı bilgi edinmek için Fedeer (1955, s. 125-126, 133ff), Howell (2008) ve özellikle Little ve Rubin'in (1987) çalışmalarını inceleyebilir, Kesim 14.12'deki tartışmaya bakabilirsiniz.

## Eşleştirilmiş Örneklemler ve İlgili problemler

tesadüfî blok deseni

Tekrarlı ölçümlerden söz ederken, aynı deneğe ait birden çok ölçümü kastediyorduk. Bu desenler en çok böyle durumlarda kullanılsa da başka durumlar da söz konusu olabilir. Bir deneğin birkaç kez test ediliyor olmasının aslında konuyla hiçbir ilgisi yoktur. Teknik olarak tekrarlı ölçüm desenlerini (veya daha genel olarak, **tesadüfî blok deseni**ni – tekrarlı ölçümler bu desenin özel bir halidir) $n$ sayısı eşit olan genel faktöriyel desenlerden ayıran şey, tekrarlı ölçümlerde $\sum$'nın köşegen dışındaki elemanlarının sıfır olmasının beklenmemesi, yani işlemler arasında korelasyon olmasıdır. Aynı deneğin tekrarlı kullanımı bu korelasyonların ortaya çıkmasına neden olur, ama **eşleştirilmiş örneklemler**e ait deneklerin kullanılması da aynı sonucu verir. Böylece, örneğin her biri üç denekten oluşan 10 grubumuz olsa, denekler otomobil kullanma deneyimine göre eşleştirilse, blok desen ve her uygulamadaki ilk denek aynı eşleştirilmiş üçlüden gelse, uygulamalar arasında korelasyonlar oluşur ve tekrarlı-ölçüm deseni elde etmiş oluruz. Sıfırdan farklı korelasyonlara (veya kovaryanslara) yol açan başka herhangi bir veri toplama yöntemi de tekrarlı-ölçüm deseni olarak değerlendirilebilir.

# 14.12 Tekrarlı Ölçüm Desenleri İçin Karma Modeller

Bölümün başlarında standart tekrarlı-ölçüm varyans analizinin varyans-kovaryans matrisi ile ilgili, *küresellik* olarak adlandırılan bir varsayım gerektirdiğini söylemiştik. Küresellik, *bileşik simetri*nin özel bir biçimidir. $\omega'$ ve $\hat{\omega}$ hakkında tartışırken, küresellik varsayımının ihlâl edilmesi ile ortaya çıkacak bazı problemlerden kurtulmak için serbestlik derecelerine uygulayabileceğimiz düzeltme faktörleri ile ilgilenmiştik.

Tekrarlı-ölçüm analizlerine ve bu analizlerin temellerindeki varsayımların ihlâline karşı ne kadar dayanıklı olduklarına dair oldukça fazla çalışma vardır. Greenhouse ve Geisser ile Huynh ve Feldt tarafından önerilen ayarlamaların başarılı olduğuna dair evrensel bir fikir birliği yoktur. Buna rağmen, genel ana etki veya etkileşim etkisine odaklanıldığı veya (genel hata terimleri kullanmak yerine) sadece belirli basit etkilerle ilgili verinin kullanıldığı durumlarda bu ayarlamalar mâkul ölçüde iyi sonuçlar vermektedir. Birleştirilmiş hata terimleri kullanılarak bireysel karşılaştırmalar veya birleştirilmiş hata terimleri kullanılarak basit etki analizleri yapıldığında, ciddî sorunlarla karşılaşılır. Boik (1981) bu durumlarda, her karşılaştırma için ayrı hata terimi kullanılmadığı sürece, tekrarlı-ölçüm analizinin küresellik varsayımı ihlâllerine karşı oldukça duyarlı olduğunu göstermiştir. Bununla birlikte, kovaryans matrisi ile ilgili varsayımlarla baş etmenin başka bir yolu daha vardır, ki o da böyle varsayımlarda bulunmamaktır. Fakat bunu yapabilmek için analizin kendisine farklı bir şekilde yaklaşmamız gerekir.

Standart tekrarlı-ölçümler varyans analizi ile ilgili uzun yıllardır aşılamamış ve muhtemelen bir süre daha aşılamayacak iki problem vardır. (Bu yüzden böyle analizler üzerinde çok fazla zaman harcamak yerine "daha iyi bir yolu olabilir" dedim.) Hem birleşik simetri (veya küresellik) hem de verinin eksiksiz olduğu varsayılır. Eğer bir katılımcı bir test oturumuna gelmemişse, diğerlerinin hepsine gelse bile analizden çıkarılması gerekir. Tekrarlı ölçüm desenlerine yönelik, küresellik varsayımı veya eksiksiz veri engellerine takılmayan alternatif bir yaklaşım vardır. Bu analiz genellikle **karma modeller**, **çok düzeyli modelleme** veya **hiyerarşik modelleme** olarak adlandırılır. Burada küçük bir karışıklık olabilir çünkü "karma modeller" terimini daha önce hem sabit hem de tesadüfî faktörler içeren tüm deneysel desenler için kullanmıştık. Ama burada olduğu gibi bir analiz yönteminden bahsettiğimizde "karna modeller"den kastımız, farklı bir hesaplama için farklı bir yaklaşım kullanan, hem sabit hem de tesadüfî faktörler içeren belirli bir çözüm

karma modeller
çok düzeyli modelleme
hiyerarşik modelleme

tipidir. Daha açık bir deyişle, bir kişi analizini karma model kullanarak yaptığını iddia ediyorsa, şu ana kadar odaklandığımız ve önümüzdeki iki bölümde de inceleyeceğimiz[7] en küçük kareler yaklaşımı yerine **en yüksek olabilirlik-EYO (maxsimum likelihood-ML)** veya, daha büyük ihtimalle, **sınırlandırılmış en yüksek olabilirlik-SEYO (restricted maximum likelihood-REML)** yöntemini kullanan bir çözümden bahsediyordur. (Eksik verimiz varsa analizimiz genellikle verinin neden eksik olduğuna dayalıdır. İdeal durum, eksikliğin tesadüfî olmasıdır.)

**en yüksek olabilirlik (EYO)**

**sınırlandırılmış en yüksek olabilirlik (SEYO)**

Bu bölümde, oldukça geniş olan hiyerarşik veya çok düzeyli model konusunun küçük bir kısmından söz edeceğim. Bu modeller için tekrar eden ölçüm (yani Zaman veya İşlem) sabit, denekler tesadüfî faktördür. Denekler-arası faktör de genellikle sabit bir faktördür. Probleme yaklaşmada, parametre kestirme yöntemi olarak sınırlandırılmış en yüksek olabilirliği kullanmak, analizin en başından itibaren bir veya daha fazla faktörün sabit, bir veya daha fazla faktörün de tesadüfî olduğunun çözüm tarafından fark edilmesini sağlar. Standart varyans analizinde en küçük kareler yöntemiyle ulaşılan çözümler *F* istatistiği için hata terimlerini belirleme aşamasına kadar tüm faktörleri sabit kabul eder.

Kimse karma model analizini ciddî olarak elde yapmaya kalkışmaz. Bu analizi yapabilmek için bilgisayar yazılımı kullanmak gerekir. Mevcut birçok yazılım vardır, hatta bunların bir kısmı da ücretsizdir. En ulaşılabilir olanlar muhtemelen SPSS Mixed ve SAS Proc Mixed isimli programlardır. Biz örneğimiz için SPSS kullanacağız, fakat SAS Proc Mixed programı muhtemelen daha esnektir.

Alternatif desenlerin analizi hakkında daha kapsamlı bilgi http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Mixed Models for Repeated Measures.pdf adresinde bulunabilir. Örnek olarak bir denekler-arası bir de denekler-içi değişkeni olan bir desen seçtim. Örnekte veri eksikliği var, çünkü standart varyans analizinin uygun olmadığı bir durumu yansıtıyor.

## Veri

Birtakım özellileri olan bir veri kümesi oluşturdum. Kontrol grubu ve İşlem grubu olmak üzere iki grup var ve bu gruplar dört farklı zamanda ölçülüyor. Ölçüm zamanlarının etiketlerini 0 (ön test), 1 (1 ay sonra son test), 3 (3 ay sonra izleme testi) ve 6 (6 ay sonra izleme testi) olarak belirtim. Depresyonun tedavisine yönelik bir çalışma düşünüyordum, bu yüzden uygulama grubunu son testte keskin bir düşüş gösterecek ve bu düşüşü (hafif bir gerileme ile) 3. ve 6. aylarda sürdürecek şekilde oluşturdum. Kontrol grubu da 4 zaman aralığı süresince yavaş bir düşüş gösteriyor, fakat İşlem grubunun düşük seviyelerine ulaşamıyor.

Veri Tablo 14.14'te görülebilir. Eksik değerler "…" sembolü ile gösterilmiştir. Karma modellere ait veri dosyalarının diğerlerinden bir farkı, genellikle "uzun form" olarak adlandırılan veri düzenlemelerinin kullanılmasıdır. Her bir deneğe ait veri için bir satır oluşturmak yerine, bağımlı değişkenin her değeri için ayrı bir satır açılır. Buna göre veri dosyamız Tablo 14.15'te görülene benzer şekilde yapılandırılır.

SPSS'in grafik ara yüzünü nasıl kullanmanız gerektiğini göstermek yerine, sadece komutlara ait söz dizimi veriyorum.[8] Veri kümesini girdikten sonra, yeni bir "söz dizimi Syntax)" penceresi açınız, aşağıdaki komutları yapıştırınız ve araç çubuğundan "Run" seçene-

---

7 Kitabın önceki baskılarında bileşik simetri varsayımlarından kaçınmak için SPSS'te "Univariate/Repeated measures" seçenekleri altındaki MANOVA yaklaşımını kullanmıştım. Bu yaklaşım birleşik simetri gerektirmez, fakat dengeli desenler gerektirir. Bu yaklaşımdan vazgeçip karma model kullanmaya başlamamın sebebi karma modelin eksik veri ile çok daha iyi başa çıkmasıdır.

8 Menü seçimlerinin kullanılmasına ait basit bir tanım şöyledir: "Analysis/karma/doğrusal seçimlerini yapın. "Denekler (Subjects)" kutusuna "Denek", "Repeated" kutusuna "Zaman" kelimelerini girin. "Devam (Continue)" tuşuna basarak bir sonraki ekrana geçin. Bağımlı değişkeni (bd) ve faktörleri (Grup ve Zaman) belirleyin. Sonra kutunun altındaki "sabit (fixed)" seçeneğini seçerek, "Grup" ve "Zamanı" işaretleyip "topla (add)" sonra da "devam (continue)" tuşlarına basın. Şimdi "tesadüfî (random)" tuşunu tıklayın ve alttaki kutuya "Denek"i ekleyin. Söz diziminizin (syntax) benim verdiğime benzediğinden emin olmak için "yapıştır (paste)" tuşuna basın.

**Tablo 14.14** Karma model analizinde kullanılan veri kümesi

| Grup | Denek | Zaman0 | Zaman1 | Zaman3 | Zaman6 |
|---|---|---|---|---|---|
| 1 | 1 | 296 | 175 | 187 | 242 |
| 1 | 2 | 376 | 329 | 236 | 126 |
| 1 | 3 | 309 | 238 | 150 | 173 |
| 1 | 4 | 222 | 60 | 82 | 135 |
| 1 | 5 | 150 | ... | 250 | 266 |
| 1 | 6 | 316 | 291 | 238 | 194 |
| 1 | 7 | 321 | 364 | 270 | 358 |
| 1 | 8 | 447 | 402 | ... | 266 |
| 1 | 9 | 220 | 70 | 95 | 137 |
| 1 | 10 | 375 | 335 | 334 | 129 |
| 1 | 11 | 310 | 300 | 253 | ... |
| 1 | 12 | 310 | 245 | 200 | 170 |
| 2 | 13 | 282 | 186 | 225 | 134 |
| 2 | 14 | 317 | 31 | 85 | 120 |
| 2 | 15 | 362 | 104 | ... | ... |
| 2 | 16 | 338 | 132 | 91 | 77 |
| 2 | 17 | 263 | 94 | 141 | 142 |
| 2 | 18 | 138 | 38 | 16 | 95 |
| 2 | 19 | 329 | ... | ... | 6 |
| 2 | 20 | 292 | 139 | 104 | . |
| 2 | 21 | 275 | 94 | 135 | 137 |
| 2 | 22 | 150 | 48 | 20 | 85 |
| 2 | 23 | 319 | 68 | 67 | ... |
| 2 | 24 | 300 | 138 | 114 | 174 |

© Cengage Learning 2013

**Tablo 14.15** Uzun formda yeniden düzenlenmiş veri

| Denek | Zaman | Grup | *dv* |
|---|---|---|---|
| 1 | 0 | 1 | 296 |
| 1 | 1 | 1 | 175 |
| 1 | 3 | 1 | 187 |
| 1 | 6 | 1 | 242 |
| … | … | … | … |
| 24 | 3 | 2 | 114 |
| 24 | 6 | 2 | 174 |

© Cengage Learning 2013

ğini seçiniz. İnce ayarlarla ilgili birkaç komutu burada belirtmedim fakat verilen komutlar analizinizi gerçekleştirecektir.

```
MIXED
dv BY Group Time
/FIXED = Group Time Group*Time | SSTYPE(3)
/METHOD = REML
/PRINT = DESCRIPTIVES SOLUTION
/REPEATED = Time | SUBJECT(Subj) COVTYPE(CS)
/EMMEANS = TABLES(Group)
/EMMEANS = TABLES(Time)
/EMMEANS = TABLES(Group*Time)
```

Burada sadece çıktının en önemli kısımlarını veriyorum, gerisini görmek için analizi kendiniz yapabilirsiniz. (Veri kümesi kitabın internet sayfasında WicksellLongMiss.dat isimli dosyadadır.)

Bilgi kriterleri[a]

| | |
|---|---|
| –2 Restricted Log Likelihood | 905,398 |
| Akaike'ın Bilgi Kriteri (AIC) | 909,398 |
| Hurvich ve Tsai'nin Kriteri (AICC) | 909,555 |
| Bozdogan'ın Kriteri (CAIC) | 916,136 |
| Schwarz'ın Bayesian Kriteri (BIC) | 914,136 |

Belirtilen bilgi kriterlerinde küçük olan daha iyi formdadır.

[a] Bağımlı Değişken: dv.

Sabit Etkiler
Tip III Sabit Etkiler Testi[a]

| Kaynak | Pay için sd | Payda için sd | F | Mnd. düzeyi |
|---|---|---|---|---|
| Kesim | 1 | 22,327 | 269,632 | ,000 |
| Grup | 1 | 22,327 | 16,524 | ,001 |
| Zaman | 3 | 58,646 | 32,453 | ,000 |
| Grup * Zaman | 3 | 58,646 | 6,089 | ,001 |

[a] Bağımlı değişken: dv

Kovaryans Parametreleri
Kovaryans Parametreleri Kestirmeleri[a]

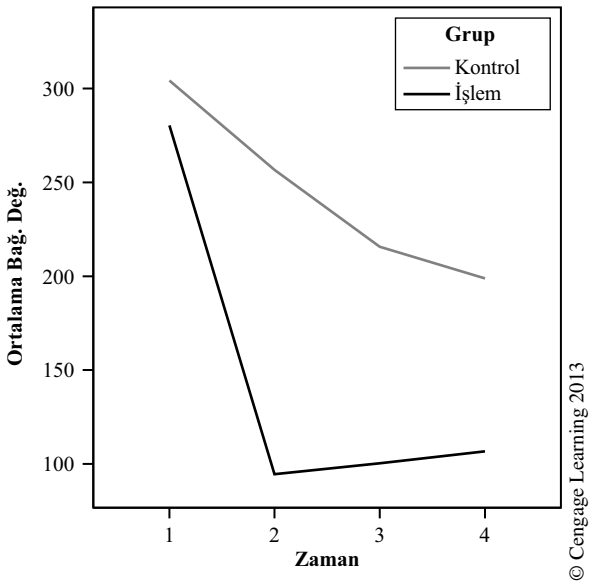
| | Parametre | Kestirim | Std. Hata |
|---|---|---|---|
| Tekrarlı Ölçümler | Köşegen uzaklığı | 2954,544 | 551,1034 |
| | Kovaryans | 2558,656 | 1026,581 |

[a] Bağımlı değişken: dv.

SPSS, Inc. çıktısından adapte edilmiştir.

"Bilgi Kriterleri" başlıklı kısmı burada tartışmayacağım, fakat farklı modellerin uyumlarını karşılaştırdığım zaman o konuya geri döneceğim. Tablonun sabit etkiler kısmı tıpkı birçok varyans analizinde görülenlere benzer, tek fark kareler toplamı ve kareler ortalamasını içermemesidir. Bunun sebebi en büyük olabilirlik çözümlerinin probleme yaklaşımıdır. Bazı yazılımlarda bunları çıktıya eklemek mümkündür. Sabit (y eksenini kesim noktası) kısmındaki teste dikkat edin. Bu sadece büyük ortalamanın "0" olduğu bir testtir ve bizim için ilginç değildir. Diğer üç etkinin hepsinin de manidar olduğu gözüküyor. İki ana etki ile çok ilgilenmiyoruz. Gruplar başlangıçta ön testte eşitler ve farkların sıfır olması gruplara herhangi bir genel ana etkiyi etkiler. Benzer şekilde Zaman etkisi ile çok ilgilenmiyoruz çünkü iki grubun da zaman içerisinde farklı davranması beklenen bir durumdur. Bizi esas ilgilendiren etkileşimdir. Etkileşim bize, görmeyi beklediğimiz gibi, iki grubun Zaman'a göre farklı davranışlar sergilediğini söylüyor. Bu etkiyi Şekil 14.5'te görebilirsiniz.

Çıktıda, üzerinde durulması gereken iki sonuç daha vardır. "Kovaryans Parametreleri" başlıklı kısım modelin tesadüfî olan kısmıdır. "CS köşegen uzaklığı" etiketli terim artık varyansı temsil eder ve dengeli desenlerde bu terim, denekler-içi testler için hata terimidir. "CS kovaryans" terimi ise kesenlerin varyansıdır. Eğer her denek için bağımlı değişkenin zamana göre grafiğini çizersek bu doğruların kesenleri arasındaki farklar deneklerden




**Şekil 14.5** İki koşul için deneme ortalamaları

**Tablo 14.16** Denemeler arası korelasyonlar

**Denek 1 İçin Kestirilen *R* Korelasyon Matrisi**

| Satır | Sütun 1 | Sütun 2 | Sütun 3 | Sütun 4 |
|---|---|---|---|---|
| 1 | 1,0000 | 0,5121 | 0,4163 | −0,08840 |
| 2 | 0,5121 | 1,0000 | 0,8510 | 0,3628 |
| 3 | 0,4163 | 0,8510 | 1,0000 | 0,3827 |
| 4 | −0,0884 | 0,3628 | 0,3827 | 1,0000 |

© Cengage Learning 2013

kaynaklanan farkları temsil eder (bazı doğrular diğerlerinden daha yüksektir). Elimizdeki varyans da budur. Birçoğumuz için varyans o kadar da önemli değildir, ama önemli olduğu çalışmalar da vardır.

Daha önce sözünü ettiğimiz gibi, karma model analizlerinde birleşik simetri varsayımı gerekmemektedir. SPSS ile korelasyon veya kovaryanslarla ilgili herhangi bir varsayımda bulunmadan problemi çözebiliriz. (Bu, özünde MANOVA yaklaşımının tekrarlı ölçümlere yaptığı şeydir.) Bu yaklaşımla ilgili sorun, çözümün korelasyonlar için tahmini değerler üretmek zorunda olmasıdır. Bu durum serbestlik derecelerinin belki de gereksiz yere azalmasına neden olur. Sadece kısmen cahilken, tamamen cahil olduğunuzu söylemenize gerek yoktur. Başka bir yaklaşım ise kovaryans matrisinin özel (ama farklı) bir biçimini varsaymaktır. Örneğin, otomatik olarak azalan çözüm ifadesiyle adlandırılan yaklaşımı kullanabiliriz. Bu çözüm yönteminde, iki gözlem arasındaki zaman aralığı arttıkça, bu gözlemler arasındaki korelasyonların azalacağı varsayılır. Bir varsayım da her korelasyonun sadece bir önceki korelasyona ve biraz (belki de çok) hataya bağlı olduğudur. Eğer birbirini izleyen iki uygulama arasındaki ortalama korelasyon, örneğin (şu anda üzerinde durduğumuz örnekte olduğu gibi) 0,5121 ise, aralarında iki zaman birimi olan gözlemlerin korelasyonunun $,5121^2 = ,2622$, üç zaman birimi olan gözlemlerin korelasyonunun $,5121^3 = ,1343$ olduğu varsayılır. Böylece, gözlemler arasındaki zaman arttıkça düzenli olarak azalan korelasyonlardan oluşan bir matris elde edilir. Zaman içinde depresyon ölçümleri yaptığımız-

da da böyle bir sonuç elde etmek mantıklı bir beklenti gibi gözüküyor. Şimdilik otomatik olarak azalan kovaryans yapısını göz önüne alacağız.

Korelasyon (veya kovaryans) yapısına karar verdikten sonra, SPSS'e sadece bu yapıyı kullanmasını ve problemi önceki gibi çözmesini söylememiz yeterli olacaktır. Yapılacak tek değişiklik repeated komutunda covtype(cs)" yerine covtype(AR1) yazmaktır.

```
MIXED
dv BY Group Time
/FIXED = Group Time Group*Time | SSTYPE(3)
/METHOD = REML
/PRINT = DESCRIPTIVES SOLUTION
/REPEATED = Time | SUBJECT(Subj) COVTYPE(AR1)
/EMMEANS = TABLES(Group)
/EMMEANS = TABLES(Time)
/EMMEANS = TABLES(Group*Time) .
```

Bilgi Kriterleri[a]

| | |
|---|---|
| –2 Sınırlı Log Olabilirlik | 895,066 |
| Akaike'nin Bilgi Kriteri (AIC) | 899,066 |
| Hurvich ve Tsai'nin Kriteri (AICC) | 899,224 |
| Bozdogan'ın Kriteri (CAIC) | 905,805 |
| Schwarz'ın Bayesian Kriteri (BIC) | 903,805 |

Bilgi kriterleri küçük-olan-daha iyidir formunda gösterilmiştir.

[a] Bağımlı değişken: *dv*.

SabitEtkiler

Sabit Etkilerin Tip III Testleri[a]

| Kaynak | Payın *sd*'si | Paydanın *sd*'si | *F* | Mnd. |
|---|---|---|---|---|
| Kesim Noktası | 1 | 26,462 | 270,516 | ,000 |
| Grup | 1 | 26,462 | 17,324 | ,001 |
| Zaman | 3 | 57,499 | 30,821 | ,000 |
| Grup * Zaman | 3 | 57,499 | 7,721 | ,001 |

[a] Bağımlı değişken: *dv*.

Kovaryans Parametreleri

Kovaryans Parametrelerinin Kestirmeleri[a]

| Parametre | | Kestirim | Std. Hata |
|---|---|---|---|
| Tekrarlı Ölçümler | AR1 köşegen | 5349,879 | 1060,035 |
| | AR1 ro (rho) | ,618198 | ,084130 |

[a] Bağımlı değişken: *dv*.

SPSS, Inc'den adapte edilmiştir.

Burada tüm etkilerin hâlâ manidar olduğunu görmek cesaret vericidir. Fakat bu iki modelden -biri kovaryans matrisinde birleşik simetri yapısını, diğeri birinci derece otomatik azalan regresyon (autoregressive) yapısını varsayan- hangisi daha iyi bir seçimdir? Üzerinde durduğumuz örnekte iki model için de aynı sonuca varacağız, fakat genelde bu doğru değildir ve hâlâ hangi modelin daha iyi olduğunu bilmek istiyoruz. Bunu yapmanın bir yolu iki analizin "Bilgi Kriterleri" başlıklı kısımlarını karşılaştırmaktır. İki modele ait bu kısımlar aşağıda verilmiştir.

**Birleşik Simetri**

| Bilgi Kriterleri[a] | |
|---|---|
| –2 Sınırlı Log Olabilirlik | 905,398 |
| Akaike'nin Bilgi Kriteri (AIC) | 909,398 |
| Hurvich ve Tsai'in Kriteri (AICC) | 909,555 |
| Bozdogan'ın Kriteri (CAIC) | 916,136 |
| Schwarz'ı Bayesian Kriteri (BIC) | 914,136 |
| Bilgi kriteri küçük olan daha iyidir formunda gösterilmiştir. | |

[a] Bağımlı değişken: *dv*.

**Öz Bağımlı (1)**

| Bilgi Kriterleri[a] | |
|---|---|
| –2 Sınırlı Log Olabilirlik | 895,066 |
| Akaike'nin Bilgi Kriteri (AIC) | 899,066 |
| Hurvich ve Tsai'in Kriteri (AICC) | 899,224 |
| Bozdogan'ın Kriteri (CAIC) | 905,805 |
| Schwarz'ın Bayesian Kriteri (BIC) | 903,805 |
| Bilgi kriteri küçük olan daha iyidir formunda gösterilmiştir. | |

[a] Bağımlı değişken: *dv*.

© Cengage Learning 2013

Modelleri karşılaştırmanın iyi bir yolu Akaike (AIC) veya Bayes Bilgi Kriterleri'ni (BIC) kullanmaktır. Genel olarak, daha küçük bir değere sahip model daha iyidir. Bizim örneklerimizde AIC kriterlerinin değerleri 909,398 ve 899,066'dır. Otomatik azalan regresyon model [Autoregressive (1)] tercih edilmeli gibi duruyor. Bu da kovaryans yapıları hakkında gözlerimizin bize söylediği ile tutarlılık gösteriyor. (Eğer analizi yapılandırılmamış bir kovaryans matrisi ile tekrar yaparsak AIC değerleri 903,691 ile 927,385 olarak elde edilir, bu da yine otomatik azalan regresyon modeli seçmemizi gerektirir.)

Karma modeller, veriyi modellere uydurarak ve altta yatan modelleri karşılaştırarak verimizi en iyi şekilde yorumlayabilmemiz için bize çok büyük imkânlar sağlar. Hem eksik veri olması hem de eksik veri olmaması durumlarında da çok işe yarayabilirler. Fakat uygulaması daha zordur ve gelişmelere rağmen yazılımlar bazı durumlarda hiç kullanışlı değildir. Yine de önümüzdeki on yılda gittikçe daha fazla analizin bu yöne kayacağını ve onları anlamanın önemli olduğunu düşünüyorum.

Overall, Tonidandel ve diğer araştırmacıların yaptığı çalışmalar karma modellerle ilgili sorunlara ışık tutmaktadır. En büyük sorun, modeli doğru tanımlamanın çok zor olmasıdır; farklı tanımlamalar farklı sorunlara yol açar ve bazen modelin gücünü düşürür. Bu konu ile ilgili mükemmel iki çalışma Overall ve Shivakumar (1999) ile Overall ve Tonidandel'a (2007) aittir. Karma model kullanmayı düşünenlerin bu makalelere bakmasını öneririm, fakat bu araştırmalarda analizler için SAS **Proc Mixed** yazılımı kullanılmış ve oluşan modellerin SPSS ile elde edeceğiniz modellerle nasıl bir ilişkisi olacağı açık değildir. Kritik öneme sahip olduğu görünen durum, eksik verinin katılımcının başlangıç düzeyindeki ilk tepkisine bağlı olması ve bu ölçümün kovaryans olarak kullanılmaya çalışılmasıdır.

## Anahtar Terimler

Ayırma (Giriş)
Ayrıştırma (Giriş)
Tekrarlı-ölçümler desenleri (Giriş)
$KT_{\text{denekler arası}}$ (Giriş)
$KT_{\text{denekler içi}}$ (Giriş)
Esas-köşegen (14.3)
Esas-köşegen dışındaki elemanlar (14.3)
Bileşik simetri (14.3)
Kovaryans matrisi ($\sum$) (14.3)
Küresellik (14.3)
Çoklu varyans analizi (MANOVA) (14.3)
Çok değişkenli yöntem (14.3)
$\text{Hata}_{\text{arası}}$ (14.7)
$\text{Hata}_{\text{içi}}$ (14.7)
Sınıflar arası korelasyon (14.10)
Sıra etkileri (14.11)
Devretme etkileri (14.11)
Tesadüfî blok desenleri (14.11)
Eşleştirilmiş örneklemler (14.11)

# Alıştırmalar

14.1 Standart hale getirilmiş testler ile tekrarlanan deneyimlerin, hiçbir müdahale yapılmasa bile, daha iyi puanlar elde edilmesine neden olduğu söylenir. Sekiz deneğimizin olduğunu ve bu deneklere üç hafta boyunca her cumartesi sabahı bir standartlaştırılmış giriş sınavı yaptığımızı varsayalım. Veri kümesi aşağıdadır:

| *Denek* | Birinci | İkinci | Üçüncü |
|---|---|---|---|
| 1 | 550 | 570 | 580 |
| 2 | 440 | 440 | 470 |
| 3 | 610 | 630 | 610 |
| 4 | 650 | 670 | 670 |
| 5 | 400 | 460 | 450 |
| 6 | 700 | 680 | 710 |
| 7 | 490 | 510 | 510 |
| 8 | 580 | 550 | 590 |

© Cengage Learning 2013

a. Bu kümeye ait istatistiksel modeli yazınız.
b. Varyans analizi yapınız.
c. GBO sınavına ait, eğer varsa, deneme etkileri için, nasıl bir sonuca varabilirsiniz?

14.2 Alıştırma 14.1'deki veri kümesini kullanarak;
a. Üçüncü oturuma dair veri kümesini silerek birinci ve ikinci oturum arasında (eşleştirilmiş örneklemler) *t*-testi uygulayınız.
b. Şimdi (a) seçeneğinde kullandığınız veri kümesine tekrarlı-ölçümler varyans analizi yapınız. Bulduğunuz *F* değerini (a) seçeneğinde elde ettiğiniz *t* değeri ile karşılaştırınız.

14.3 Temel öğrenim kuramlarının pratik uygulamalarını göstermeyi amaçlayan, davranış değiştirme ile ilgilenen bir psikolog, ileri derecede gelişim bozukluğu olan çocuklara öz bakım becerileri kazandırmak amacı ile tasarlanan bir çalışma için veri toplamıştır. Deney grubuna öz bakım ile ilgili etkinlikler için pekiştirme verilmiştir. İkinci bir grup ise eşit miktarda ilgi görmüş, fakat pekiştirme almamıştır. Bir değerlendirmeci, çocuklara 10 puanlık öz yeterlilik ölçeği üzerinden (kör değerlendirme ile) puan vermiştir. Değerlendirmeler başlangıç oturumunda ve eğitimin sonunda olmak üzere iki kere yapılmıştır. Veri kümesi aşağıdadır:

| Pekiştireç Verildi | | Pekiştireç Verilmedi | |
|---|---|---|---|
| Başlangıçta | Eğitim Sonrası | Başlangıçta | Eğitim Sonrası |
| 8 | 9 | 3 | 5 |
| 5 | 7 | 5 | 5 |
| 3 | 2 | 8 | 10 |
| 5 | 7 | 2 | 5 |
| 2 | 9 | 5 | 3 |
| 6 | 7 | 6 | 10 |
| 5 | 8 | 6 | 9 |
| 6 | 5 | 4 | 5 |
| 4 | 7 | 3 | 7 |
| 4 | 9 | 5 | 5 |

© Cengage Learning 2013

Uygun analizi yapıp hangi sonuçlara vardığınızı açıklayınız.

14.4 İstatistik eğitimi çok az olan bir araştırmacı, alıştırma sorusu 14.3'teki veriyi alıp bağımsız örneklemler *t* testi yapmıştır. Ham veri olarak puanlar arasındaki farkı (eğitim–temel düzey) kullanmıştır.

a. Bu analizi siz de yapınız.
b. *t* değerinin karesini alıp soru 14.3'te bulduğunuz *F* değerleriyle karşılaştırınız.
c. Gruplar için $t^2$ değerinin neden *F* değerine eşit olmadığını açıklayınız.

14.5 İleri derecede gelişim bozukluğu olan öğrencilerin eğitimi ile ilgili deneyde (Alıştırma 14.3) tam olarak ne olduğunu anlayabilmek için esas araştırmacımız, diğer iki grupla aynı zamanlarda üçüncü bir grubu daha değerlendirmiş, fakat bu gruba hiçbir müdahalede bulunmamıştır. Başka bir deyişle bu çocuklar pekiştirme almadıkları gibi, kontrol grubuna gösterilmiş olan fazladan ilgiyi de görmemişlerdir. Gruba ait veri kümesi aşağıdadır:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Başlangıçta:** | 3 | 5 | 8 | 5 | 5 | 6 | 6 | 6 | 3 | 4 |
| **Eğitim Sonrası:** | 4 | 5 | 6 | 6 | 4 | 7 | 7 | 3 | 2 | 2 |

a. Bu veriyi Alıştırma 14.3'teki veriye ekleyerek analizi tekrar yapınız.

b. Sonuca ait grafiği çiziniz.

c. (a) ve (b) seçeneklerinde bulduklarınızdan nasıl bir sonuç çıkarıyorsunuz?

d. Bu üç gruplu deney bağlamında, soru 14.3'ten aldığınız iki durum için karşılaştırma analizi yapınız.

e. (d) seçeneğinde bulduğunuz karşılaştırma sonucu için etki büyüklüğünü hesaplayınız.

14.6 Çocuk ve yetişkinlerin öyküleri nasıl özetlediklerini inceleyen bir çalışmada, 10 beşinci sınıf öğrencisi ve 10 yetişkin seçilmiştir. Bu kişiler de kendi aralarında okuması iyi ve zayıf olanlar olmak üzere (okuması iyi ve zayıf olanların öykü ile ilgili bilgileri farklı şekillerde depoluyor veya hatırlıyor olabilecekleri hipotezinden yola çıkarak) ikiye ayrılmıştır. Tüm denekler 10 öykü okumuştur. Okumaları biter bitmez deneklerden öyküyü kendi cümleleri ile özetlemeleri istenmiştir. Tüm özetlere içerik analizi uygulanmış, Yer/Zaman, Amaç ve çıkarılan Sonuçlar ile ilgili ifadelerin sayısı kaydedilmiştir. Veri 10 öykü üzerinden daraltılmıştır:

| | Yetişkinler | | | Çocuklar | | |
|---|---|---|---|---|---|---|
| Yaş Maddeleri | Yer/Zaman | Amaç | Sonuç | Yer/Zaman | Amaç | Sonuç |
| **Okuması iyi olanlar** | 8 | 7 | 6 | 5 | 5 | 2 |
| | 5 | 6 | 4 | 7 | 8 | 4 |
| | 5 | 5 | 5 | 7 | 7 | 4 |
| | 7 | 8 | 6 | 6 | 4 | 3 |
| | 6 | 4 | 4 | 4 | 4 | 2 |
| **Okuması zayıf olanlar** | 7 | 6 | 3 | 2 | 2 | 2 |
| | 5 | 3 | 1 | 2 | 0 | 1 |
| | 6 | 6 | 2 | 5 | 4 | 1 |
| | 4 | 4 | 1 | 4 | 4 | 2 |
| | 5 | 5 | 3 | 2 | 2 | 0 |

© Cengage Learning 2013

Uygun analizi yapınız.

14.7 Soru 14.6'ya bakarak

a. Çocuklar için okuma becerisinin basit etkisini hesaplayınız.

b. Okuması iyi yetişkinler için maddelerin basit etkisini hesaplayınız.

14.8 Soru 14.6'daki veri için gruplar içi kovaryans matrislerini hesaplayınız.

14.9 Deneklerimizden, özetlerini 10 kelime ile sınırlamalarını istediğimizi varsayalım. Soru 14.6'daki veri üzerinde bunun nasıl bir etkisi olabilir?

14.10 Sigara içimi ile ilgili bir çalışmada, araştırmacı üç farklı sigara bırakma yöntemini (yavaş yavaş azaltma, birdenbire bırakma, isteksizlik terapisi) karşılaştırmaya karar vermiştir. Her gruptan beşer denek alarak onlardan "şu anda" iki farklı ortamda (ev ve iş) sigara içmeyi ne kadar istediklerini 10 puan üzerinden değerlendirmelerini istemiştir. Bu değerlendirme sigarayı bırakmadan önce ve sonra olmak üzere iki kere yapılmıştır. Böylece elimizde bir denekler arası (İşlem grubu), iki de denekler içi (Ortam ve Önce/Sonra) değişken bulunmaktadır.

| | Önce | | Sonra | |
|---|---|---|---|---|
| | Ev | İş | Ev | İş |
| **Yavaş bırakma** | 7 | 6 | 6 | 4 |
| | 5 | 4 | 5 | 2 |
| | 8 | 7 | 7 | 4 |
| | 8 | 8 | 6 | 5 |
| | 6 | 5 | 5 | 3 |
| **Ani bırakma** | 8 | 7 | 7 | 6 |
| | 5 | 5 | 5 | 4 |
| | 7 | 6 | 6 | 5 |
| | 8 | 7 | 6 | 5 |
| | 7 | 6 | 5 | 4 |
| **Kaçınma** | 9 | 8 | 5 | 4 |
| | 4 | 4 | 3 | 2 |
| | 7 | 7 | 5 | 3 |
| | 7 | 5 | 5 | 0 |
| | 8 | 7 | 6 | 3 |

© Cengage Learning 2013

a. Uygun varyans analizi yöntemini uygulayınız.

b. Sonuçları yorumlayınız.

14.11 Soru 14.10'da bulduğunuz sonuçların grafiğini çiziniz.

14.12 Sonuçları açık ve kesin hale getirmek için soru 14.10'daki veri kümesine basit etkiler analizi yapınız.

14.13 Gösterim 14.1'de görülen kısaltılmış çıktı Alıştırma 14.5'teki veri kümesinin analizini temsil etmektedir.

a. Bu çıktıyı, Alıştırma 14.5'te bulduğunuz sonuçlarla karşılaştırınız.

b. Ortalama (MEAN) için $F$ değerinin manidar çıkması bize ne söyler?

c. $KO_{\text{hücre içi}}$ ile hücre standart sapmaları tablosunu ilişkilendiriniz.

```
BMDP2V – ANALYSIS OF VARIANCE AND COVARIANCES
WITH REPEATED MEASURES.

PROGRAM CONTROL INFORMATION
/PROBLEM    TITLE IS 'BMDP2V ANALYSIS OF EXERCISE 14.5'.
/INPUT      VARIABLES ARE 3.
            FORMAT IS '(3F2.0)'.
            CASES ARE 30.

/VARIABLE   NAMES ARE GROUP, PRE, POST.
/DESIGN     DEPENDENT ARE 2, 3.
            LEVELS ARE 2.
            NAME IS TIME.
            GROUP = 1.
/END
          CELL MEANS FOR 1-ST DEPENDENT VARIABLE
                                                        MARGINAL
     GROUP      =   * 1.0000   * 2.0000   *3.0000
               TIME
PRE              1    4.80000   4.70000   5.10000   4.86667
POST             2    7.00000   6.40000   4.60000   6.00000
MARGINAL              5.90000   5.55000   4.85000   5.43333
COUNT                   10        10        10        30
```

SPSS, Inc.'dan uyarlanmıştır.

STANDARD DEVIATIONS FOR 1-ST DEPENDENT VARIABLE

| GROUP | 5 | * 1,0000 | * 2,000 | *3,0000 |
|---|---|---|---|---|
| TIME | | | | |
| PRE | 1 | 1,68655 | 1,76698 | 1,52388 |
| POST | 2 | 2,16025 | 2,45855 | 1,89737 |

| SOURCE | SUM OF SQUARES | DEGREES OF FREEDOM | MEAN SQUARE | F | TAIL PROBABILITY |
|---|---|---|---|---|---|
| MEAN | 1771,26667 | 1 | 1771,26667 | 322,48 | 0,0000 |
| GROUP | 11,43333 | 2 | 5,71667 | 1,04 | 0,3669 |
| 1 ERROR | 148,30000 | 27 | 5,49259 | | |
| TIME | 19,26667 | 1 | 19,26667 | 9,44 | 0,0048 |
| TG | 20,63333 | 2 | 10,31667 | 5,06 | 0,0137 |
| 2 ERROR | 55,10000 | 272,04074 | | | |

**Gösterim 14.1**

14.14 Gösterim 14.2'de görünen SPSS çıktısı Alıştırma 14.6'daki verinin, tüm değişkenlerin denekler-arası değişken olarak kabul edilerek (verinin standart üç-faktöriyel deseni temsil ettiği varsayılarak) değerlendirilmesi sonucu ortaya çıkmıştır. Doğru analizdeki hata terimlerinin, faktöriyel analizdeki hata teriminin bir ayrışımı olduğunu gösteriniz.

**Denekler-Arası Etki Testi**

Bağımlı Değişken: DV

| Kaynak | Tip III Kareler Toplamı | sd | Kareler Ortalaması | F | Man. |
|---|---|---|---|---|---|
| Düzeltilmiş Model* | 170,800[a] | 11 | 15,527 | 9,001 | ,000 |
| Kesim Noktası | 1058,400 | 1 | 1058,400 | 613,565 | ,000 |
| YAŞ | 68,267 | 1 | 68,267 | 39,575 | ,000 |
| OKUMA TİPİ | 29,400 | 1 | 29,400 | 17,043 | ,000 |
| KISIM | 60,400 | 2 | 30,200 | 17,507 | ,000 |
| YAŞ*OKUMA TİPİ | 3,267 | 1 | 3,267 | 1,894 | ,175 |
| YAŞ*KISIM | ,933 | 2 | ,467 | ,271 | ,764 |
| OKUMA TİPİ*KISIM | ,000 | 2 | ,000 | ,000 | 1,000 |
| YAŞ*OKUMA TİPİ*KISIM | 8,533 | 2 | 4,267 | 2,473 | ,095 |
| Hata | 82,800 | 48 | 1,725 | | |
| Toplam | 1312,000 | 60 | | | |
| Düzeltilmiş Toplam | 253,600 | 59 | | | |

[a] R Kare = ,674 (Düzeltilmiş R Kare = ,599)

**Gösterim 14.2**

14.15 *A* ve *B* üzerinde tekrarlı, *C* ve *D* üzerinde de bağımsız ölçümlere sahip bir *A* x *B* x *C* x D desenine ait bir özet tabloyu ana hatlarıyla belirleyiniz.

14.16 Foa, Rothbaum, Riggs ve Murdock (1991), travma sonrası stres bozukluğu (TSSB) tedavisinde kullanılan farklı yöntemleri karşılaştıran bir çalışma yapmışlardır. Çalışmada üç grup (bir de bekleme durumundaki kontrol grubu) kullanılmıştır. Bir gruba Stres Aşılama Terapisi (StAT), ikinci gruba Uzun Süre Maruz Bırakılma tedavisi (UMB), üçüncü gruba da standart Destek Danışmanlığı (DD) uygulanmıştır. Tüm hastalar üzerinde Tedavi Öncesi, Tedavi Sonrası ve 3,5 Ay Sonra İzleme olmak üzere üçer kez ölçme yapılmıştır. Aşağıdaki veri, araştırmacıların topladıklarına oldukça yakındır. Bağımlı değişken ise TSSB ölçümüdür.

| StAT | | | UMB | | | DD | | |
|---|---|---|---|---|---|---|---|---|
| Önce | Sonra | İzleme | Önce | Sonra | İzleme | Önce | Sonra | İzleme |
| 19 | 6 | 1 | 20 | 5 | 0 | 12 | 14 | 18 |
| 28 | 14 | 16 | 21 | 18 | 21 | 27 | 18 | 9 |
| 18 | 6 | 8 | 36 | 26 | 17 | 24 | 19 | 13 |
| 23 | 6 | 11 | 25 | 11 | 9 | 32 | 21 | 11 |
| 21 | 6 | 13 | 26 | 2 | 7 | 26 | 20 | 18 |
| 24 | 10 | 8 | 30 | 31 | 10 | 18 | 20 | 26 |
| 26 | 10 | 7 | 19 | 6 | 11 | 38 | 35 | 34 |
| 15 | 6 | 13 | 19 | 7 | 5 | 26 | 22 | 22 |
| 18 | 8 | 6 | 22 | 4 | 4 | 23 | 10 | 8 |
| 34 | 13 | 8 | 22 | 17 | 20 | 22 | 19 | 19 |
| 20 | 10 | 16 | 24 | 19 | 1 | 34 | 27 | 23 |
| 34 | 10 | 1 | 28 | 22 | 16 | 22 | 15 | 12 |
| 29 | 16 | 23 | 29 | 23 | 20 | 27 | 18 | 13 |
| 33 | 19 | 39 | 27 | 15 | 20 | 23 | 21 | 19 |
| 22 | 7 | 16 | 27 | 7 | 3 | 26 | 18 | 13 |

© Cengage Learning 2013

a. Bu veriye tekrarlı ölçümler varyans analizi uygulayınız.

b. İlgili sonuçları çıkarınız.

14.17 Alıştırma 14.16'daki veriyi kullanarak, kovaryans matrisi için uygun bir form belirleyip SPSS ile karma modeller varyans analizi yapınız. Sonuçları daha önce elde ettiklerinizle karşılaştırınız.

14.18 Aşağıdaki veri kümesi Alıştırma 14.16 ile aynıdır fakat bazı gözlemler silinmiştir. (999 eksik gözlemleri temsil etmektedir.)

| StAT | | | UMB | | | DD | | |
|---|---|---|---|---|---|---|---|---|
| Önce | Sonra | İzleme | Önce | Sonra | İzleme | Önce | Sonra | İzleme |
| 19 | 6 | 1 | 20 | 5 | 0 | 12 | 14 | 18 |
| 28 | 14 | 16 | 999 | 999 | 21 | 27 | 18 | 9 |
| 18 | 6 | 8 | 36 | 26 | 17 | 24 | 999 | 13 |
| 999 | 6 | 11 | 25 | 11 | 9 | 32 | 21 | 11 |
| 21 | 6 | 13 | 26 | 999 | 7 | 26 | 20 | 18 |
| 24 | 10 | 8 | 30 | 31 | 10 | 18 | 20 | 26 |
| 26 | 10 | 999 | 19 | 6 | 11 | 38 | 35 | 34 |
| 15 | 6 | 13 | 19 | 7 | 999 | 26 | 22 | 999 |
| 18 | 8 | 6 | 22 | 4 | 999 | 23 | 10 | 8 |
| 34 | 13 | 8 | 22 | 17 | 20 | 22 | 19 | 19 |
| 20 | 999 | 999 | 24 | 19 | 1 | 34 | 999 | 999 |
| 34 | 10 | 1 | 28 | 22 | 16 | 22 | 15 | 12 |
| 29 | 16 | 23 | 29 | 23 | 20 | 27 | 18 | 13 |
| 33 | 19 | 39 | 27 | 15 | 20 | 23 | 21 | 19 |
| 22 | 7 | 16 | 27 | 7 | 3 | 26 | 18 | 13 |

© Cengage Learning 2013

a. Bu veriyi standart tekrarlı ölçümler varyans analizi kullanarak analiz ediniz.

b. Elde ettiğiniz sonuçlar, Alıştırma 14.20'de bulduğunuz sonuçlardan nasıl ayrılır?

14.19 Bu defa 14.18'deki veriyi, uygun bir kovaryans matrisi formu ile karma modeller yaklaşımını kullanarak analiz ediniz. Elde ettiğiniz sonuçlar ile 14.18'de bulduğunuz sonuçlar arasında nasıl bir fark var?

14.20 Web sitesinde bulunan "Stress.dat" isimli dosyadaki veri, kanser hastaları ve eşleri tarafından, tanı konduktan kısa bir süre sonra ve 3 ay sonra olmak üzere iki kez bildirilen stres düzleyleridir. Veri kümesinde katılımcının cinsiyeti de belirtilmektedir. Genelde olduğu gibi, "." işareti eksik veri anlamına gelmektedir. 686. sayfadaki "EK: Bilgisayar Veri Kümeleri" başlığı altındaki açıklamalara bakabilirsiniz.

a. Herhangi bir istatistik paket programı kullanarak tekrarlı ölçümler varyans analizi yapınız. Burada, Cinsiyet ve Rol (hastaya karşı eşi) denekler-arası değişkenler, Zaman da tekrarlı ölçüm olarak alınacaktır. (Eksik veri için kullanılan sembolü programda tanıtmanız gerekebilir, olmazsa başka bir sembol kullanınız.)
b. Programın hücre ortalamalarının çıktısını oluşturmasını sağlayınız ve yorum yapmanıza yardımcı olması için bu ortalamaların grafiğini çizdiriniz.
c. Bu analizde manidar çıkan bir üçlü etkileşim var. Bunu ana etkilerle beraber yorumlayınız.

14.21 Everitt genç kızlarda anoreksiya hastalığını tedavide kullanılan üç yöntem üzerine bir çalışma yapmıştır. Çalışmada ele alınan tedavi yöntemleri bilişsel davranışçı terapi, terapisiz kontrol durumu ve aile terapisi durumudur. İlgili veri kümesi aşağıda verilmiştir:

| Grop | Ön test | Son test | Kazanç | Grup | Ön test | Son test | Kazanç |
|---|---|---|---|---|---|---|---|
| 1 | 80.5 | 82.2 | 1.7 | 2 | 75.1 | 86.7 | 11.6 |
| 1 | 84.9 | 85.6 | .7 | 2 | 80.6 | 73.5 | –7.1 |
| 1 | 81.5 | 81.4 | –.1 | 2 | 78.4 | 84.6 | 6.2 |
| 1 | 82.6 | 81.9 | –.7 | 2 | 77.6 | 77.4 | –0.2 |
| 1 | 79.9 | 76.4 | –3.5 | 2 | 88.7 | 79.5 | –9.2 |
| 1 | 88.7 | 103.6 | 14.9 | 2 | 81.3 | 89.6 | 8.3 |
| 1 | 94.9 | 98.4 | 3.5 | 2 | 78.1 | 81.4 | 3.3 |
| 1 | 76.3 | 93.4 | 17.1 | 2 | 70.5 | 81.8 | 11.3 |
| 1 | 81.0 | 73.4 | –7.6 | 2 | 77.3 | 77.3 | 0.0 |
| 1 | 80.5 | 82.1 | 1.6 | 2 | 85.2 | 84.2 | –1.0 |
| 1 | 85.0 | 96.7 | 11.7 | 2 | 86.0 | 75.4 | –10.6 |
| 1 | 89.2 | 95.3 | 6.1 | 2 | 84.1 | 79.5 | –4.6 |
| 1 | 81.3 | 82.4 | 1.1 | 2 | 79.7 | 73.0 | –6.7 |
| 1 | 76.5 | 72.5 | –4.0 | 2 | 85.5 | 88.3 | 2.8 |
| 1 | 70.0 | 90.9 | 20.9 | 2 | 84.4 | 84.7 | 0.3 |
| 1 | 80.4 | 71.3 | –9.1 | 2 | 79.6 | 81.4 | 1.8 |
| 1 | 83.3 | 85.4 | 2.1 | 2 | 77.5 | 81.2 | 3.7 |
| 1 | 83.0 | 81.6 | –1.4 | 2 | 72.3 | 88.2 | 15.9 |
| 1 | 87.7 | 89.1 | 1.4 | 2 | 89.0 | 78.8 | –10.2 |
| 1 | 84.2 | 83.9 | –.3 | 3 | 83.8 | 95.2 | 11.4 |
| 1 | 86.4 | 82.7 | –3.7 | 3 | 83.3 | 94.3 | 11.0 |
| 1 | 76.5 | 75.7 | –.8 | 3 | 86.0 | 91.5 | 5.5 |
| 1 | 80.2 | 82.6 | 2.4 | 3 | 82.5 | 91.9 | 9.4 |
| 1 | 87.8 | 100.4 | 12.6 | 3 | 86.7 | 100.3 | 13.6 |
| 1 | 83.3 | 85.2 | 1.9 | 3 | 79.6 | 76.7 | –2.9 |
| 1 | 79.7 | 83.6 | 3.9 | 3 | 76.9 | 76.8 | –0.1 |
| 1 | 84.5 | 84.6 | .1 | 3 | 94.2 | 101.6 | 7.4 |
| 1 | 80.8 | 96.2 | 15.4 | 3 | 73.4 | 94.9 | 21.5 |
| 1 | 87.4 | 86.7 | –.7 | 3 | 80.5 | 75.2 | –5.3 |
| 2 | 80.7 | 80.2 | –.5 | 3 | 81.6 | 77.8 | –3.8 |
| 2 | 89.4 | 80.1 | –9.3 | 3 | 82.1 | 95.5 | 13.4 |
| 2 | 91.8 | 86.4 | –5.4 | 3 | 77.6 | 90.7 | 13.1 |
| 2 | 74.0 | 86.3 | 12.3 | 3 | 83.5 | 92.5 | 9.0 |
| 2 | 78.1 | 76.1 | –2.0 | 3 | 89.9 | 93.8 | 3.9 |
| 2 | 88.3 | 78.1 | –10.2 | 3 | 86.0 | 91.7 | 5.7 |
| 2 | 87.3 | 75.1 | –12.2 | 3 | 87.3 | 98.0 | 10.7 |

a. Alınan Kilo puanları arasındaki grup farklarına bakmak için varyans analizi yapınız.
b. Aynı analizi bu sefer tekrarlı ölçüm deseni kullanarak yapınız. Tekrarlı ölçüm olarak Ön Test ve Son Test puanlarını alınız.

© Cengage Learning 2013

c. (b) seçeneğinin cevabı ile (a) seçeneğinin cevabı arasında nasıl bir ilişki vardır?

d. Her grup için ayrı ayrı Ön Test ve Son Test puanları arasındaki ilişkiyi gösteren dağılım grafiğini çizdiriniz. Bu grafikler neleri gösteriyor?

e. Kontrol grubuna ait Alınan Kilo değerinin 0,00 olacağına dair sıfır hipotezini test ediniz. Şaşırdınız mı?

f. İki deney grubunda alınan kilolara ait gözlenen değerler için manidar sonuçlar neden kontrol grubu olmadan yorumlanamaz?

## Tartışma Soruları

14.22 Alıştırma 14.20'de deneklerimizin aynı aileden çiftler olması durumunu göz ardı ettik.

a. Böyle yapmak neden yanlıştır?

b. Hangi koşullar altında bu problemi göz ardı etmek kabul edilebilir?

c. Alternatif olarak hangi analizleri önerirsiniz?

14.23 Alıştırma 14.20'de Zaman 2'de birçok gözlemin eksik olduğunu fark etmişsinizdir. (Bunun sebeplerinden biri birçok hasta için tanının üzerinden henüz 3 ay geçmemiş olmasıdır.)

a. Zaman 2'de, veri bulunan ve bulunmayan deneklerin Zaman 1'deki ortalamalarını karşılaştırınız.

b. Eğer (a) seçeneğinde bir fark bulduysanız, bu sonuç veri kümesi hakkında akla ne getirir?

Çocuklarda davranış problemleriyle ilgili bir çalışmada 3 değerlendirmeciden 20 çocuğun saldırgan davranış düzeyini belirlemeleri istenmiştir. Bu değerlendirmeciler çocuğun ebeveyni, öğretmeni ve kendisidir. Veri kümesi aşağıdadır.

| **Çocuk** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ebeveyn** | 10 | 12 | 14 | 8 | 16 | 21 | 10 | 15 | 18 | 6 | 22 | 14 | 19 | 22 | 11 | 14 | 18 | 25 | 22 | 7 |
| **Öğretmen** | 8 | 13 | 17 | 10 | 18 | 24 | 9 | 16 | 18 | 8 | 24 | 19 | 15 | 20 | 10 | 18 | 19 | 30 | 20 | 10 |
| **Kendisi** | 12 | 17 | 16 | 15 | 24 | 24 | 13 | 17 | 21 | 13 | 29 | 23 | 16 | 20 | 15 | 17 | 21 | 25 | 25 | 14 |

Bu veri kümesi, Bölüm 14.10'da gördüklerimizden bir şekilde farklıdır. O örnekte her çocuğu aynı kişiler değerlendiriyordu. Burada ise Ebeveyn ve Kendi doğal olarak çocuktan çocuğa değişiyor. Bu durumu şimdilik göz ardı edip, bir şekilde tüm değerlendirmeleri aynı ebeveynin ve "kendi"nin (çocuğun) yapabildiğini varsayacağız.

14.24 Sınıf içi korelasyon katsayısı bakımından bu veri kümesinin güvenirliği nedir?

14.25 Hesaplamalarınız bu veri kümesindeki değişkenliğin kaynağı ile ilgili size ne söylüyor?

14.26 Bir kaynağın çocukları diğer bir kaynaktan sistematik olarak daha yüksek veya daha düşük değerlendirmesi ile ilgili hiçbir endişenizin olmadığını varsayınız. Güvenirliği daha farklı nasıl değerlendirirsiniz?

14.27 Değerlendirmeciler arasındaki farklarla hangi koşullar altında ilgilenmeyebilirsiniz?

14.28 Ebeveyn değerlendirmesini yapan "ebeveyn"in, çocuktan çocuğa değişiyor olmasının önemi nedir?

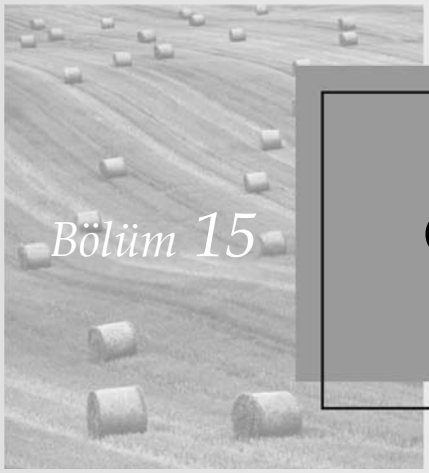
14.29 Strayer, Drews ve Crouch (2006) (Alıştırma 11.32'de denekler arası desen olarak kurulduğunu gördüğümüz) çalışmalarında cep telefonu kullanımının otomobil sürme becerisi üzerine olan etkilerini incelemişlerdir. Çalışma için 40 sürücü cep telefonu ile konuşurken, yasal alkol limitini (%0,08) aşmadan ve normal koşullar altında otomobil kullanmışlardır. (Koşullarda sürücüler arasında karşıt dengeleme yapılmıştır.) Bu çalışmaya ait veri www.uvm.edu/~dhowell/methods8/DataFiles/Ex14-29.

dat adresinde bulunmaktadır. (Maddeler arası korelasyonların yaklaşık olarak ,45 olduğunu varsaydım.) Başkalarının çalışmalarına dayandırdıkları hipotezleri, otomobil kullanırken cep telefonu ile konuşmanın, otomobil kullanırken alkollü olmak kadar etkili olduğu idi. Bu örnekteki bağımlı değişken "fren tepki süresi"dir. Örnekteki veri kümesine ait ortalama ve standart sapma, çalışmada araştırmacıların buldukları ile aynıdır.

a. Tekrarlı ölçümler deseni için varyans analizi yapınız.

b. Uygun karşılaştırma yöntemini kullanarak üç koşulu karşılaştırınız. Bulduğunuz sonuçlar araştırmacıların öngörülerini destekliyor mu?

14.30 Alıştırma 11.37'de Teri'ye (1997) ait, Alzheimer hastalarının farmakolojik olmayan yöntemlerle tedavisini inceleyen bir çalışma görmüştük. Orada her grup için sadece değişim puanlarını vermiştik. Ön ve son test puanlarını www.uvm.edu/~dhowell/methods8/DataFiles/Teri.dat adresinde bulabilirsiniz. Tekrarlı ölçümler varyans analizi yaparak elde ettiğiniz sonuçları Ünitede 11'de oluşturduğunuz özet tablo ile karşılaştırınız.



Bölüm 15

# Çoklu Regresyon

## Hedefler

Birden fazla yordayıcı değişkeni aynı anda analize dâhil ederek bir yordanan değişkene ait kestirimde bulunmaya yönelik işlemlerin tanıtılması ve söz konusu işlemlerin doğasında var olan problemlere dikkat çekilmesi.

## Konular


**15.1** Çoklu Doğrusal Regresyon
**15.2** Ek Yordayıcıların Kullanılması
**15.3** Regresyon Katsayılarının Standart Hatası ve Test Edilmesi
**15.4** Yeniden Örnekleme Yaklaşımı
**15.5** Artık Varyans
**15.6** Dağılım Varsayımları
**15.7** Çoklu Korelasyon Katsayısı
**15.8** Kısmi ve Yarı Kısmi Korelasyon
**15.9** Bastırıcı Değişkenler
**15.10** Regresyona Hazırlık
**15.11** Regresyon Eşitliğinin Oluşturulması
**15.12** Her Bir Değişkenin "Önemi"
**15.13** Yaklaşık Regresyon Katsayılarının Kullanılması
**15.14** Aracı Değişken ve Şartlı Değişken İlişkileri
**15.15** Lojistik Regresyon

BÖLÜM 9'da, bir ölçüt değişkenin (*Y*) tek bir yordayıcı değişken *(X)* ile yordanması amaçlanmıştır. Bu bölümde, Bölüm 9'dan farklı olarak birden fazla yordayıcı değişken analize dâhil olmaktadır. Daha açık bir deyişle, analizde yine tek bir ölçüt değişken bulunmakta *(Y)*, ancak birden fazla yordayıcı değişken ($X_1$, $X_2$, $X_3$,…., $X_p$) yer almaktadır. Uygulanan analiz sayesinde, *p* sayıda yordayıcı değişken *aynı anda* analize alınarak *Y* ölçüt değişken yordanacaktır. Bundan dolayı, Bölüm 9'da yer verilen regresyon analizi, bu bölümde sunulan çoklu regresyonun özel bir hali olarak düşünülebilir. Buna bağlı olarak, bu bölümde daha önceki bölümlerden aşina olunan korelasyon katsayısı, eğim, tahminin standart hatası ve $KT_{\text{regresyon}}$ gibi kavramların kullanılmasına devam edeceğiz.

Hesaplamalara odaklanmanın pek bir getirisi olmayacağından, bu bölümde yazılım çözümleri üzerinde duracağız. Kendimizi regresyon analizi ile ilgili hesaplamalardan uzak tuttuğumuzda, uygun regresyon çözümünü seçmenin altında yatan ve gerçekten önemli olan konulara odaklanmamız daha kolay olacaktır. Genellikle formüller üzerinde çok az tartışma vardır. Diğer taraftan; en uygun yordayıcı değişken sayısı, regresyon tanımının kullanılması, yordayıcı değişkenlerin bağıl ve yordayıcı değişkenlerin seçimi hakkındaki sorular için herkes tarafından kabul gören cevaplar yoktur. Bu bölüme başlangıç yaparken, sunulan açıklamaların birer görüş niteliği taşıdığı ve tartışmaya açık olduğu konusunda bir ön ikazda bulunmak yerinde olacaktır. Cohen, Cohen, West ve Aiken (2003) ile Stevens (1992) çoklu regresyon analizini çalışmak için önerilebilecek kaynaklar arasında yer almaktadır. Regresyon analizine ilişkin varsayımların test edilmesine yönelik en iyi kaynaklardan biri ise Fox'tur (2002).

## 15.1 Çoklu Doğrusal Regresyon

Regresyon analizinde, *Y* (bazen $X_0$ ile gösterilir) ölçüt değişkenin *p* sayıda yordayıcı değişken ($X_1$, $X_2$, $X_3$,, $X_p$) ile yordanmasına yönelik bir regresyon eşitliğinin elde edilmesi amaçlanır. Örneğin; lisansüstü eğitim başarısını (*Y*) yordamak için lisans not ortalaması ($X_1$), mezuniyet sırasında uygulanan standart testlerden alınan puan ($X_2$), temel disiplin ile ilgili alınan derslerin sayısı ($X_3$) ve tavsiye mektuplarının olumluluğuna ilişkin puanları ($X_4$) yordayıcı değişken olarak alabiliriz. Bu yordayıcı değişkenleri kullanarak lisansüstü eğitim başarısının yordanmasını hedef alabiliriz. Benzer şekilde, şehirdeki herhangi bir noktadan diğer bir noktaya gitmenin ne kadar zaman alacağını; trafik ışıklarının sayısı ($X_1$), hız limiti ($X_2$), kırmızı ışık yanarken dönüşün serbest olup olmaması ($X_3$) ve trafik yoğunluğu ($X_4$) yordayıcı değişkenleri aracılığıyla yordamak isteyebiliriz. İlk örnekte, başvuru yapan öğrencilerden hangilerinin lisansüstü eğitim programlarında daha başarılı olacağına yönelik doğru yordamalar yapılması amaçlanmakta ve bireysel başvurulara odaklanılmaktadır. İkinci örnekte ise, süre hakkında yapılacak yordamalardan çok, her bir yordayıcı değişkenin oynadığı rol ile ilgilenilmektedir. Bununla birlikte, sunulan örneklerin her ikisi de aynı şekilde analiz edilir. Gerçekte çoklu regresyon analizinin kullanılmasının başlıca nedeni, analiz sonucunda elde edilen eşitliği kullanarak bir yordama yapmanın ötesinde değişkenler arasındaki ilişkilerin anlaşılmasıdır.

### Regresyon Eşitliği

**regresyon katsayıları**

Bölüm 9'a, $\hat{Y} = bX + a$ doğru denklemi ve bu denklemdeki *a* ile *b* değerlerinin $\Sigma(Y - \hat{Y})^2$ değerlerini minimum yapacak çözümü ile başlamıştık. Çoklu regresyon analizinde de aynı işlemi uygulayacağız. Ancak, çoklu regresyon analizinde, $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p$ eşitliğinin çözülmesini amaçlayacağız. Bu eşitlikte $b_0$ kesim noktasını (regresyon doğrusunun *y* eksenini kestiği nokta) ve $b_1$, $b_2$, $b_3$,…, $b_p$ ise sırasıyla $X_0$, $X_1$, $X_2$, $X_3$, .., $X_p$ yordayıcı değişkenlerine ait **regresyon katsayıları**dır. Bu eşitliğin, $\Sigma(Y - \hat{Y})^2$ ifadesini minimum yapacak şekilde çözülmesi amaçlanmaktadır. Böylelikle, yordanan *Y* değerlerinin gerçek *Y* değerlerine yakın olması sağlanmaya çalışılır.[1] Yordayıcı değişken sayısı arttıkça $b_i$ değerlerinin kestirilmesi için gerekli hesaplamaların yapılması zorlaşır. Burada bu

1 Standart en küçük kareler ölçütüne alternatif yöntemler vardır. Bu yöntemler bazı açılardan en küçük kareler yöntemine göre daha üstün yordayıcılar üretmektedir. Standart en küçük kareler yöntemine kıyasla daha az yaygın olan bu yöntemler genellikle Dayanıklı Regresyon olarak adlandırılır. Rousseeuw ve Leroy (1987), bu yöntemlerin çoğunu içermektedir.

hesaplamalarla ilgilenmeyip bir örnekle başlayacağız ve gerekli çözümün SPSS, Minitab veya SAS gibi bilgisayar programlarından biri yardımıyla elde edildiğini varsayacağız. Burada kullanılan veri dosyası bu kitaba ait internet sitesinde Table 15-1.dat adı ile yer almaktadır, oradan alınabilir.

Burada kullanılacak örnek, Guber'ın (1999) çalışmasından alınmıştır. Ancak, regresyon analizini bu bölümün sonundaki alıştırmalara taşıyabilmek için veri dosyasına bazı değişkenler eklenmiştir. İlköğretim ve ortaöğretimin kalitesini arttırmak için neler yapılabileceği hakkında uzun zamandır süregelen tartışmalar bulunmaktadır. Genel olarak eğitime daha fazla para harcandığında, öğrencilerin daha iyi yetişeceği kabul edilmektedir. Bununla birlikte, bu sadece bir varsayımdır. Bu soruya cevap bulabilmek için Amerika Birleşik Devletleri'ndeki (ABD) 50 eyaletten veri toplanmıştır. Bu kapsamda, her eyalet için eğitime harcanan para, öğrenci/öğretmen oranı (öğretmen başına düşen öğrenci sayısı), ortalama öğretmen maaşı, SAT'a katılan öğrencilerin yüzdesi (PctSAT) ve toplam SAT puanları kaydedilmiştir. İnternet sitesinde ayrı ayrı bulunan Sözel ve Matematik puanları birleştirilerek ele alınmıştır. Her eyalet için ACT'ye giren öğrenci yüzdesi ve ortalama ACT puanları da veri dosyasına eklenmiştir. Veri kümesi Tablo 15-1'de gösterilmiştir. Guber'ın çalışmasının özeti ile tam metnine http://www.amstat.org/publications/jse/v7n2_abstracts.html internet adresinden erişilebilir.

**Tablo 15.1** Eğitim harcamaları ve performansa ilişkin veri kümesi

| Eyalet | Harcama | Öğren./ Öğret. Oranı | Maaş | SAT'a Giren Öğr. Yüzdesi | SAT Ort. | ACT'ye Giren Öğr. Yüzdesi | ACT Ort. |
|---|---|---|---|---|---|---|---|
| Alabama | 4,405 | 17,2 | 31,144 | 8 | 1029 | 61 | 20,2 |
| Alaska | 8,963 | 17,6 | 47,951 | 47 | 934 | 32 | 21,0 |
| Arizona | 4,778 | 19,3 | 32,175 | 27 | 944 | 27 | 21,1 |
| Arkansas | 4,459 | 7,1 | 28,934 | 6 | 1005 | 66 | 20,3 |
| California | 4,992 | 24,0 | 41,078 | 45 | 902 | 11 | 21,0 |
| Colorado | 5,443 | 18,4 | 34,571 | 29 | 980 | 62 | 21,5 |
| Connecticut | 8,817 | 14,4 | 50,045 | 81 | 908 | 3 | 21,4 |
| Delaware | 7,030 | 16,6 | 39,076 | 68 | 897 | 3 | 21,0 |
| Florida | 5,718 | 19,1 | 32,588 | 48 | 889 | 36 | 20,4 |
| Georgia | 5,193 | 16,3 | 32,291 | 65 | 854 | 16 | 20,2 |
| Hawaii | 6,078 | 17,9 | 38,518 | 57 | 889 | 17 | 21,6 |
| Idaho | 4,210 | 19,1 | 29,783 | 15 | 979 | 62 | 21,4 |
| Illinois | 6,136 | 17,3 | 39,431 | 13 | 1048 | 69 | 21,2 |
| Indiana | 5,826 | 17,5 | 36,785 | 58 | 882 | 19 | 21,2 |
| Iowa | 5,483 | 15,8 | 31,511 | 5 | 1099 | 64 | 22,1 |
| Kansas | 5,817 | 15,1 | 34,652 | 9 | 1060 | 74 | 21,7 |
| Kentucky | 5,217 | 17,0 | 32,257 | 11 | 999 | 65 | 20,1 |
| Louisiana | 4,761 | 16,8 | 26,461 | 9 | 1021 | 80 | 19,4 |
| Maine | 6,428 | 13,8 | 31,972 | 68 | 896 | 2 | 21,5 |
| Maryland | 7,245 | 17,0 | 40,661 | 64 | 909 | 11 | 20,7 |
| Massachusetts | 7,287 | 14,8 | 40,795 | 80 | 907 | 6 | 21,6 |
| Michigan | 6,994 | 20,1 | 41,895 | 11 | 1033 | 68 | 21,3 |
| Minnesota | 6,000 | 17,5 | 35,948 | 9 | 1085 | 60 | 22,1 |
| Mississippi | 4,080 | 17,5 | 26,818 | 4 | 1036 | 79 | 18,7 |
| Missouri | 5,383 | 15,5 | 31,189 | 9 | 1045 | 64 | 21,5 |

*(devam ediyor.)*

© Cengage Learning 2013

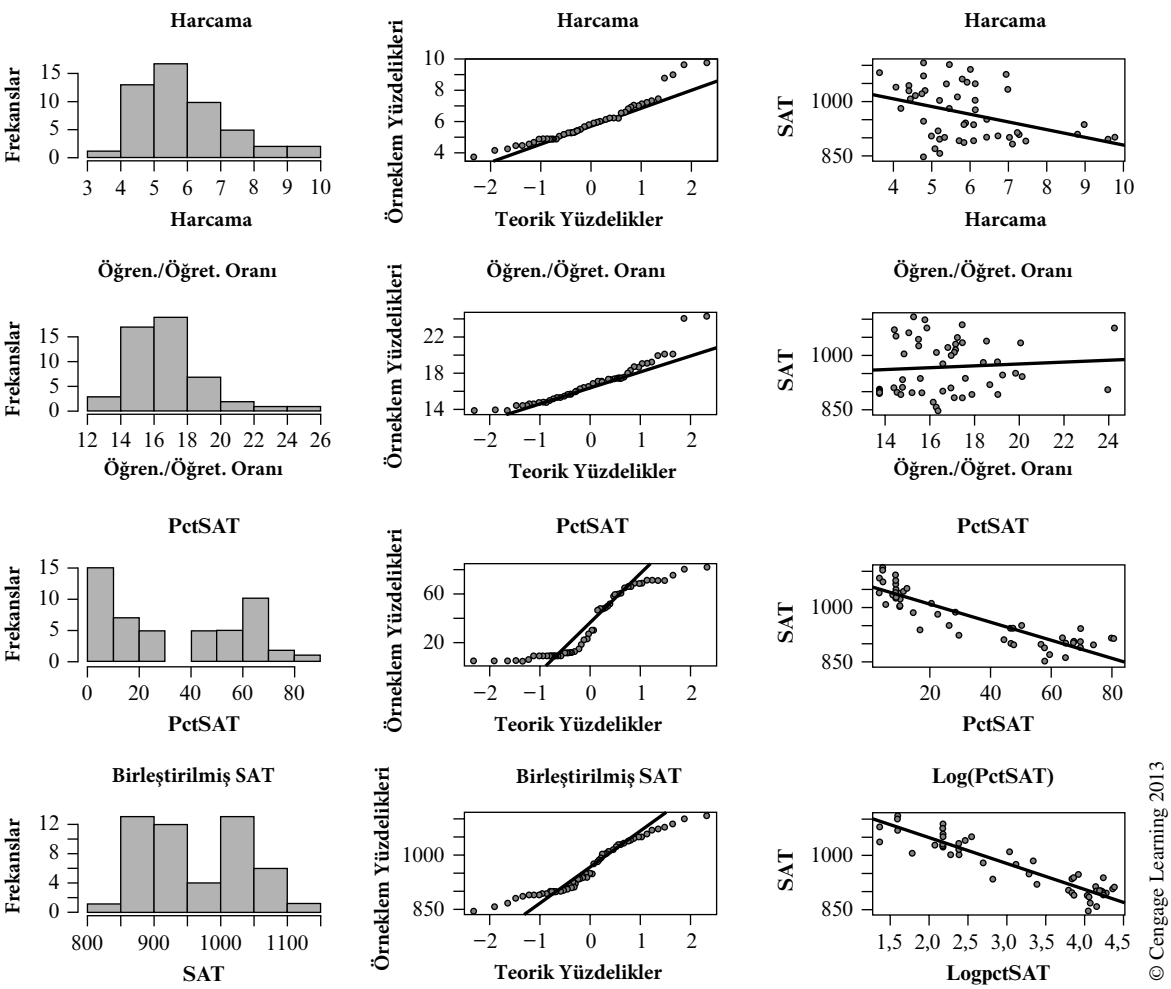
**Tablo 15.1** Eğitim harcamaları ve performansa ilişkin veri kümesi *(devamı)*

| Eyalet | Harcama | Öğren./ Öğret. Oranı | Maaş | SAT'a Giren Öğr. Yüzdesi | SAT Ort. | ACT'ye Giren Öğr. Yüzdesi | ACT Ort. |
|---|---|---|---|---|---|---|---|
| Montana | 5,692 | 16,3 | 28,785 | 21 | 1009 | 55 | 21,9 |
| Nebraska | 5,935 | 14,5 | 30,922 | 9 | 1050 | 73 | 21,7 |
| Nevada | 5,160 | 18,7 | 34,836 | 30 | 917 | 39 | 21,3 |
| New Hampshire | 5,859 | 15,6 | 34,720 | 70 | 935 | 4 | 22,3 |
| New Jersey | 9,774 | 13,8 | 46,087 | 70 | 898 | 3 | 20,8 |
| New Mexico | 4,586 | 17,2 | 28,493 | 11 | 1015 | 59 | 20,3 |
| New York | 9,623 | 15,2 | 47,612 | 74 | 892 | 16 | 21,9 |
| North Carolina | 5,077 | 16,2 | 30,793 | 60 | 865 | 11 | 19,3 |
| North Dakota | 4,775 | 15,3 | 26,327 | 5 | 1107 | 78 | 21,4 |
| Ohio | 6,162 | 16,6 | 36,802 | 23 | 975 | 60 | 21,3 |
| Oklahoma | 4,845 | 15,5 | 28,172 | 9 | 1027 | 66 | 20,6 |
| Oregon | 6,436 | 19,9 | 38,555 | 51 | 947 | 12 | 22,3 |
| Pennsylvania | 7,109 | 17,1 | 44,510 | 70 | 880 | 8 | 21,0 |
| Rhode Island | 7,469 | 14,7 | 40,729 | 70 | 888 | 2 | 21,4 |
| South Carolina | 4,797 | 16,4 | 30,279 | 58 | 844 | 13 | 18,9 |
| South Dakota | 4,775 | 14,4 | 25,994 | 5 | 1068 | 68 | 21,3 |
| Tennessee | 4,388 | 18,6 | 32,477 | 12 | 1040 | 83 | 19,7 |
| Texas | 5,222 | 15,7 | 31,223 | 47 | 893 | 30 | 20,2 |
| Utah | 3,656 | 24,3 | 29,082 | 4 | 1076 | 69 | 21,5 |
| Vermont | 6,750 | 13,8 | 35,406 | 68 | 901 | 7 | 21,9 |
| Virginia | 5,327 | 14,6 | 33,987 | 65 | 896 | 6 | 20,7 |
| Washington | 5,906 | 20,2 | 36,151 | 48 | 937 | 16 | 22,4 |
| West Virginia | 6,107 | 14,8 | 31,944 | 17 | 932 | 57 | 20,0 |
| Wisconsin | 6,930 | 15,9 | 37,746 | 9 | 1073 | 64 | 22,3 |
| Wyoming | 6,160 | 14,9 | 31,285 | 10 | 1001 | 70 | 21,4 |

Bu veri dosyasını seçtim, çünkü bunu kullanarak regresyon analizine dair çok sayıda özelliğin gösterilebileceğini düşündüm. İlk olarak, bu dosya güncel bir problemi konu edinen gerçek bir veri kümesini içerir. Bu durum, veri kümesindeki değişkenlerin, tesadüfî sayılar üreticisinden oluşturduğum değişkenler kadar düzgün bir dağılıma sahip olmadığı anlamına gelir. Gerçekten sayılar biraz karışıktır. Ayrıca, veri kümesi başlangıçta, oldukça şaşırtıcı sonuçlar göstermekte ve daha sonra bu sonuçları mantıklı bir şekilde açıklama imkânı sunmaktadır. Veri kümesindeki yordayıcı değişkenlerden sadece biri kullanıldığında karşılaşılan sonuç ile analize iki yordayıcı değişken dâhil edildiğinde ortaya çıkan sonuç oldukça farklıdır. Veri dosyası bu yönüyle çoklu regresyon analizinin yararlarına örnek teşkil edebilecek bir niteliğe sahiptir. Son olarak; bu veri kümesi ölçümler üzerinde dikkatlice düşünülmesi gerektiğini göstermekte ve elde edilen ölçümlerin gerçekten ölçülmek istenen özelliği yansıttığı şeklindeki bir varsayımın hiç sorgulanmadan kabul edilmesinin yanıltıcı olabileceğine işaret etmektedir.

Bu kitap ABD ve Kanada dışındaki birçok kişi tarafından kullanılmaktadır. Dolayısıyla sunulan örnekte yer alan değişkenler hakkında bilgi verilmesi konunun daha iyi anlaşılmasına yardımcı olacaktır. SAT ve ACT üniversite başvurularında kullanılan iki ayrı standart testtir. SAT puanları 200 ile 800 arasında değişebilirken ACT'den alınabilecek puanlar 1 ile 36 arasında uzanmaktadır. SAT bir yetenek testi olarak nitelendirilmekte, ACT ise okulda işlenen konular ile ilgili bir başarı testi şeklinde tanımlanmaktadır. Ortalamasının çok daha düşük olduğu dikkate alındığında bile, ACT'ye ilişkin standart sapma SAT'ın standart sapmasına kıyasla önemli ölçüde daha düşüktür (SAT ve CAT'a ait değişkenlik katsayıları sırasıyla 0,077 ve 0,042'dir). İki sınav arasındaki en önemli fark, SAT ağırlıklı olarak ülkenin kuzeydoğusundaki ve batısındaki prestiji yüksek üniversitelerce kullanılmaktadır. Bu iki kıyı dışındaki yerlerde yaşayan öğrenciler bu kıyılardaki Harvard, Princeton, Berkeley veya Stanford gibi okullara başvurmayı düşünmedikçe ACT'ye başvurma olasılıkları daha yüksektir. Örnek olarak sunulan regresyon analizine ilişkin sonuçların doğru yorumlanabilmesi için bu farkın dikkate alınması oldukça önemlidir.

Regresyon çözümünü düşünmeden önce, her bir değişkenin dağılımına bakılması gerekir. Bu değişkenlerin her birine ait dağılımlar; histogram, *qq* (yüzdelik-yüzdelik ve dağılım diyagramı olarak Şekil 15.1'de gösterilmiştir. Örnekte yer alan değişkenlerin normal dağılıma sahip olmadığı bu grafiklerden açıkça görülmektedir. Yine bu grafiklere bakıldığında, ölçüt (yordanan) değişken ile yordayıcı değişkenlerden üç tanesinin iki modlu bir dağılıma sahip olduğu anlaşılmaktadır. Sat'a girenler yüzdesi değişkeninin iki modlu olduğu, bu değişkene ait histogramda açık bir şekilde görülmektedir. Buna göre, herhangi bir eyalet için ya öğrencilerin büyük bir kısmı SAT'a girmekte ya da SAT'a giren öğrenci sayısı oldukça az olmaktadır. Diğer bir deyişle, SAT'a giren öğrenci sayısının orta düzeyde olduğu eyalet oldukça sınırlı sayıdadır. Öte yandan, SAT'a giren öğrenci sayısının fazla ya da az olduğu eyaletlerin sayısı oldukça fazladır. SAT puanlarındaki iki modlu dağılımın bir diğer göstergesi; SAT'a giren öğrenci sayısı yüzdesi ile SAT puanları arasında eğrisel bir ilişki bulunmasıdır. SAT'a giren öğrenci yüzdesi değişkeninin doğal logaritması alındığında, bu dağılım kısmen düzelmektedir. Ayrıca sağ alt köşede bulunan dağılım diyagramında da görüldüğü gibi, SAT'a girenlerin yüzdesi değişkeninin doğal logaritması (LogPctSAT) alındığı takdirde, bu değişken ile SAT puanları arasındaki ilişki nispeten daha doğrusal hale gelmektedir. Regresyon analizine ilişkin örnekte, SAT'a giren öğrenci yüzdesi yerine bu değişkenin logaritmik dönüşümü kullanılacaktır. Çünkü dağılım iki modlu olsa da, söz konusu logaritmik dönüşüm bu problemin üstesinden gelmeye katkı sağlamaktadır. Son olarak; Şekil 15.1'in sol alt köşesine bakıldığında toplam SAT puanlarının geniş bir dağılıma sahip olduğu gözlenmektedir.




© Cengage Learning 2013

**Şekil 15.1** Örnekte sunulan değişkenler için histogram, Q-Q grafikleri ve dağılım grafikleri

## İki Değişkenli İlişkiler

Bu veri kümesiyle yapılacak en belirgin şey, eğitim harcamaları ile eğitim çıktıları arasındaki ilişkilerin incelenmesidir. Başlangıçta, ACT veri kümesi göz ardı edilerek, SAT puanları ve eğitim harcamaları arasındaki ilişkiye odaklanılacaktır. Bu ilişkiye bakarken, test puanları üzerinde etkili olabilecek muhtemel diğer değişkenler arasındaki ilişkilere de bakılacaktır. Eğitime harcanan para arttıkça öğrencilerin performansının da artacağı sonucuyla karşılaşılacağımızı öngörmekteyiz. Bunun yanı sıra, çok tartışılan konulardan biri olan öğrenci/öğretmen oranın ve birçok kişinin tartıştığı öğretmen maaşlarının eğitim çıktıları üzerindeki rolüyle de ilgileneceğiz. Burada, SAT puanlarının eğitim çıktılarına ilişkin bir ölçüm olarak alındığı unutulmamalıdır. SAT puanları okul öğrenmeleri için iyi bir gösterge olmasa da, sunulan örnek için uygun bir ölçüm niteliğindedir.

Şekil 15.1'in sağ üst köşesindeki şekil, SAT puanları ile eğitim harcamaları arasındaki ilişkinin grafiğini göstermektedir. Ayrıca, Tablo 15.2'de değişkenler arasındaki Pearson Momentler çarpımı korelasyon katsayıları verilmiştir. Tablodaki en şaşırtıcı noktalardan biri, regresyon doğrusunun azalan olması, SAT puanları ile eğitim harcamaları arasındaki negatif ilişkidir. SAT puanları ile eğitim harcamaları arasındaki korelasyon katsayısı negatif olup (r = -,381, $p$ = ,006), bu korelasyon eğitime yapılan harcama arttıkça, öğrencilerin başarılarının düştüğü anlamına gelir. Bu korelasyon katsayısı çok yüksek olmasa da, istatistiksel açıdan manidardır ve göz ardı edilemeyecek bir sonuçtur. Böyle bir sonuç, varlıklı okullardan gelen öğrencilerin SAT'ta daha az bir başarı ortaya koyma eğiliminde olduklarına işaret eder. Neden böyledir? Öğrenci/öğretmen oranı ile öğrenci başarısı arasında bir ilişki bulunmaması tablodaki bir diğer şaşırtıcı sonuçtur. Bu sonuç nasıl yorumlanmalıdır?

SAT testi hakkında daha önce söylediklerimin dikkate alınması, karmaşık görünen bu sonuçların anlamlandırılmasına yardımcı olabilir. Tüm kolejler ve üniversiteler için öğrencilerin SAT'a girmesi bir zorunluluk değildir. Dolayısıyla, sadece en iyi öğrencileri alan daha prestijli üniversitelere girmek isteyen öğrenciler SAT'a katılmaktadır. Ek olarak, SAT'a giren öğrenci sayısı açısından eyaletler arasında büyük farklılıklar bulunmaktadır. Örneğin; Connecticut eyaletindeki öğrencilerin %81'i, Utah eyaletindeki öğrencilerin ise sadece %4'ü SAT'a başvurmaktadır. Yüzdenin en düşük olduğu eyaletler orta batıda, en yüksek olduğu eyaletler ise kuzey batıda yer almaktadır. Sınava giren öğrenci sayısının düşük olduğu eyaletlerde, sadece en iyi okullara girmeyi amaç edinen ve başarılı olma olasılıkları yüksek olan

**Tablo 15.2** Seçilen değişkenler arasındaki korelasyonlar

**Korelasyonlar**

| | | Harcama | Öğrenci/ Öğretmen Oranı | Maaş | PctSAT | SAT Puanları Ortalaması | LogPctSAT |
|---|---|---|---|---|---|---|---|
| Harcama | Pearson Korel. | 1 | -,371 | ,870 | ,593 | -,381 | ,561 |
| | Man.(2-yönlü) | | ,008 | ,000 | ,000 | ,006 | ,000 |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |
| Öğrenci/ Öğretmen Oranı | Pearson Korel. | -,371 | 1 | -,001 | -,213 | ,081 | -,132 |
| | Mnd.(2-yönlü) | ,008 | | ,994 | ,137 | ,575 | ,361 |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |
| Maaş | Pearson Korel. | ,870** | -,001 | 1 | ,617** | -,440** | ,613** |
| | Mnd.(2-yönlü) | ,000 | ,994 | | ,000 | ,001 | ,000 |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |
| SAT'a Katılan Öğrenci Yüzdesi (PctSAT) | Pearson Korel. | ,593** | -,213 | ,617** | 1 | -,887** | ,961** |
| | Man.(2-yönlü) | ,000 | ,137 | ,000 | | ,000 | ,000 |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |
| SAT Puanları Ortalaması | Pearson Korel. | -,381** | ,081 | -,440** | -,887** | 1 | -,926** |
| | Mnd. (2-yönlü) | ,006 | ,575 | ,001 | ,000 | | ,000 |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |
| LogPctSAT | Pearson Korel. | ,561** | -,132 | ,613** | ,961** | -,926** | 1 |
| | Man. (2-yönlü) | ,000 | ,361 | ,000 | ,000 | ,000 | |
| | N | 50 | 50 | 50 | 50 | 50 | 50 |

** Korelasyon ,001 düzeyinde manidardır (iki yönlü).

öğrenciler SAT'a başvurmaktadır. Öğrencilerin birçoğunun sınava girdiği Massachusetts ve Connecticut gibi eyaletlerde ise, düşük seviyedeki öğrenciler de SAT'a başvurmaktadır. Dolayısıyla bu eyaletlerde sınava giren düşük seviyedeki öğrenciler sadece en iyi öğrencilerin sınava girdiği eyaletlere kıyasla bulundukları eyaletin ortalamasını düşürmektedir. Eğer bu doğru ise, sınava giren öğrencilerin yüzdesi ile eyaletlerin ortalamaları arasındaki korelasyonun negatif olmasını bekleriz. Şekil 15.1'in sağ alt köşesinde yer alan ve SAT puanları ile LogPctSAT arasındaki korelasyona baktığımızda gördüğümüz durum budur.

## Yordayıcılardan Birini Kontrol İçin Diğer Yordayıcılar Arasındaki İlişkinin İncelenmesi

Şimdi akla gelen soru; öğrencilerin SAT puanlarının yordayıcıları olarak SAT'a giren öğrenci yüzdesinin doğal logaritması (LogPctSAT) ve Eğitim Harcamaları değişkenleri aynı anda analize alınırsa nasıl bir sonuçla karşılaşılacağıdır. Bunun anlamı çok açık olmasa da burada, *Log*PctSAT değişkenini kontrol için, Eğitim harcamaları ve SAT puanları arasındaki ilişki incelenecektir. (Ayrıca, Eğitim Harcamaları değişkeni *kontrol altında tutularak LogPct*SAT ile SAT puanları arasındaki ilişki incelenecektir). LogPctSAT değişkeninin kontrol altına alınmasıyla, değişkenler arasındaki ilişkiler incelenirken LogPctSAT'ın sabit olarak tutulması kastedilmektedir. Sadece 50 eyalet yerine, ABD'de binlerce eyalet olduğunu ve SAT'a katılan öğrenci yüzdesinin eşit olduğu (örneğin, %60) eyaletlerin çalışmaya dâhil edildiğini kabul edelim. Ardından, SAT'a katılan öğrenci yüzdesinin eşit olduğu, bu eyaletler için SAT puanlarını Eğitim Harcamalarından kestirebilmek için, SAT puanları ile Eğitim Harcamaları arasındaki korelasyon ve regresyon katsayılarının hesaplandığını varsayalım. Bu işlemin ardından, SAT'a katılma oranının %40 olduğu eyaletlerden bir örneklem oluşturarak, SAT puanları ile eğitim harcamaları arasındaki ilişkiye ait korelasyon ve regresyon katsayılarını sadece bu eyaletler için hesaplayabiliriz. Burada, iki korelasyon ve iki regresyon katsayısı hesapladım ve iki durumda da SAT'a giren öğrenci yüzdesi değişkenini sabit tuttum (%40 veya %60). Binlerce eyalet bulunduğunun kabul edildiği bu süreçte, PctSAT değişkeninin belirli bir değerde sabit tutulmasına yönelik işlemlerin birçok kez tekrarlanacağı varsayılmaktadır. Bu analizlerin her biri için SAT puanları ile Eğitim Harcamaları arasındaki ilişkiye ait bir regresyon katsayısı elde edilecektir. Elde edilen regresyon katsayılarının ortalaması, incelenecek regresyon katsayısına oldukça yakın olacaktır. Bunların, genellikle regresyon katsayısı olarak ifade edilmesine rağmen, kısmi regresyon katsayısı olarak adlandırılması daha doğrudur. Çünkü bu regresyon analizinde, herhangi bir yordayıcı değişken ile ölçüt değişken arasındaki ilişki incelenirken, diğer yordayıcı değişkenler kontrol altında tutulmaktadır. Aynı şekilde hesaplanan korelasyon katsayılarının ortalamasının alınması, kısmi korelasyon katsayısının elde edilmesini sağlayacaktır. (Daha karmaşık bir model tanıtılmadan önce, SAT ile Eğitim Harcamaları arasındaki ilişki ne olursa olsun, LogPctSAT'ın her bir düzeyi için aynı olduğu varsayılmaktadır).

Açık bir şekilde ABD'de binlerce eyalet yoktur. Verilen örnekte, sadece 50 eyalet bulunmaktadır. Bununla birlikte eyalet sayısının sınırlı olması, yordayıcıların kontrol altına alınmasına yönelik bu sürecin gerçekleşmiş olduğunun varsayılmasına ve süreç sonunda elde edilecek katsayının ne olacağının matematiksel olarak kestirilmesine engel teşkil etmez. Yordayıcıların kontrol altına alınmasına yönelik bu işlemler çoklu regresyonun ne anlama geldiğini tam olarak yansıtır.

## Çoklu Regresyon Denklemi

Çoklu regresyonu düşünmenin bir veya daha fazla değişkenin düzeylerini sabit tutmanın dışında başka yolları da vardır. Bu yolları tartışmadan önce, bu veri kümeleri üzerinde çoklu regresyon analizi yaptım. Regresyon analizi için SPSS paket programını kullandım. Sonuçlar Gösterim 15.1'de gösterilmiştir. Regresyon analizinde, ilk adımda, sadece Eğitim Harcamaları değişkeninin yordayıcı olarak analize dâhil edilmesi talimatını verdim, daha sonra LogPctSAT değişkenini ekledim ve regresyon analizini bu iki yordayıcı değişken ile gerçekleştirdim. Normal olarak regresyon analizini bu şekilde yapmam (her iki yordayıcı değişken ile tek bir regresyon analizi gerçekleştiririm). Bununla birlikte, bu tür bir yaklaşımın çoklu regresyon analizinin anlaşılmasını kolaylaştıracağını düşündüm. Analiz çıktı-

larında, aşağıda sunulandan daha fazla sayıda tablo bulunmaktadır. Ancak, yer kazanmak amacıyla sonuçlarla ilgili tabloların bir kısmı aşağıda gösterilmemiştir.

Tablolar incelendiğinde, her bir tablonun iki bölümü vardır - birinci bölümünde, regresyon analizinin tek yordayıcı değişkeni olan Eğitim Harcamaları ve diğer bölümünde ise, Eğitim Harcamaları ve LogPctSAT yordayıcı değişkenlerinin ikisi de bulunmaktadır. İlk olarak Model Özeti başlıklı birinci tabloyu tartışmak isterim. Özet tabloda görüldüğü gibi, Eğitim Harcamaları tek yordayıcı değişken olduğunda, eğitim harcamaları ile SAT puanları arasındaki korelasyon katsayısı, daha önce gördüğünüz gibi -,381'dir. Ancak, LogPctSAT değişkeni ikinci yordayıcı olarak analize dahil edildiğinde, korelasyon katsayısının değeri -,381'den oldukça uzaklaşarak ,941'e yükselmektedir.

Burada, çoklu regresyon analizine ilişkin birçok husus mevcuttur. İlk olarak, çoklu regresyon analizinde korelasyon katsayısı daima pozitif olacaktır. Öte yandan, basit Pearson korelasyonu, pozitif veya negatif olabilir. Bu durumun güzel bir nedeni vardır, fakat bunun ayrıntısına şimdi girmek istemiyorum. (Eğer korelasyonlar daima pozitif olursa ilişkinin negatif olduğunu nasıl bileceğiz? Regresyon katsayısının işaretine bakacağız, kısa zaman içinde buna geleceğim.) Tablo 15.2 hatırlanırsa, SAT ve LogPctSAT değişkenleri arasındaki basit korelasyonun -,93, SAT ile Eğitim Harcamaları arasındaki korelasyon katsayısının ise -,38 olduğu belirlenmişti. Sadece Eğitim Harcamaları değişkeninin dâhil edildiği regresyon analizine LogPctSAT değişkenin eklenmesi, analize önemli bir katkı sunmakta-

**Model Özeti**

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Kestirimin Standart Hatası |
|---|---|---|---|---|
| 1 | ,381[a] | ,145 | ,127 | 69,909 |
| 2 | ,941[b] | ,886 | ,881 | 25,781 |

[a] Yordayıcılar: (Sabit), Eğitim Harcamaları
[b] Yordayıcılar: (Sabit), Eğitim Harcamaları, LogPctSAT

**ANOVA[c]**

| Model | | Kareler Toplamı | sd | Kareler Ortalaması | F | Mnd, |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 39722,059 | 1 | 39722,059 | 8,128 | ,006[a] |
| | Artık | 234585,621 | 48 | 4887,200 | | |
| | Toplam | 274307,680 | 49 | | | |
| 2 | Regresyon | 243065,908 | 2 | 121532,954 | 182,834 | ,000[b] |
| | Artık | 31241,772 | 47 | 664,719 | | |
| | Toplam | 274307,680 | 49 | | | |

[a] Yordayıcılar: (Sabit), Eğitim Harcamaları
[b] Yordayıcılar: (Sabit), Eğitim Harcamaları, LogPctSAT
[c] Bağımlı Değişken: Toplam SAT Puan

**Katsayılar[a]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | | | B için %95 Olasılıklı Güven Aralığı | |
|---|---|---|---|---|---|---|---|---|
| | | B | Standart Hata | Beta | t | Mnd, | Alt Sınır | Üst Sınır |
| 1 | (Sabit) | 1089,294 | 44,390 | | 24,539 | ,000 | 1000,042 | 1178,546 |
| | Eğitim Harcamaları | –20,892 | 7,328 | -,381 | -2,851 | ,006 | –35,627 | -6,158 |
| 2 | (Sabit) | 1147,100 | 16,701 | | 68,684 | ,000 | 1113,50 | 1180,698 |
| | Eğitim Harcamaları | 11,129 | 3,264 | ,203 | 3,409 | ,001 | 4,562 | 17,696 |
| | LogPctSAT | –78,203 | 4,471 | -1,040 | -17,490 | ,000 | -87,198 | -69,208 |

[a] Bağımlı Değişken: Birleştirilmiş SAT Puanı

**Gösterim 15.1** SAT puanlarını Eğitim Harcamaları ve LogPctSAT değerlerinden yordayan çoklu regresyon

dır. Ancak, sadece LogPctSAT değişkeninin kullanıldığı regresyon analizine Eğitim Harcamaları değişkenin eklenmesi regresyon modeline önemli bir katkı sağlamamaktadır. Bir dakikadan daha kısa süre içerisinde bu konuya da değineceğiz.

Model özetinde, korelasyonların kareleri de görülmektedir. Korelasyonların karesi basit regresyon analizinde ne anlama geliyorsa, çoklu regresyon analizinde de aynı anlamı taşır. Eğitim Harcamaları değişkeninin tek yordayıcı olması durumunda SAT puanlarındaki varyansın %14,5'lik bir kısmı açıklanabilmektedir (-,381$^2$ = 14,5). Eğitim Harcamaları ile LogPctSAT değişkenleri aynı anda analize alındığında, SAT puanlarındaki varyansın %88,6'sı bu iki değişken ile açıklanabilmektedir (,941$^2$ = ,886). Bu değerlerin sağ tarafında, düzeltilmiş $R^2$ olarak adlandırılan bir sütun görülmektedir. Bu sütun büyük ölçüde göz ardı edilebilir. Düzeltilmiş $R^2$ değeri, korelasyon katsayının karesinin evrendeki değerinin daha az yanlı bir göstergesidir. Diğer işlemlerde yansız kestirimler tercih edilmesine rağmen, burada düzeltilmiş $R^2$ değeri nadiren rapor edilmektedir. Basit bir biçimde düzeltmiş $R$ yerine, $R$ kullanılmaktadır.

Model özeti tablosundan sonra ANOVA tablosu yer almaktadır. Bu tablo, regresyon katsayısının manidarlığının test edilmesine yönelik varyans analizi sonuçlarını içerir. $F$, çoklu korelasyon katsayısının 0'dan manidar düzeyde farklı olup olmadığını belirlemeye yarayan bir testtir. Bölüm 9 ve Bölüm 10'da $t$ testi kullanılmış olsa da bu test, Bölüm 9 ve Bölüm 10'da verilen test ile aynı türdendür. Tablo 15.2'de görüldüğü gibi, tek bir yordayıcı değişken (Eğitim Harcamaları) olduğunda, korelasyon katsayısı -,38'dir. Bu korelasyon katsayısının 0'dan manidar düzeyde farklı olup olmadığını gösteren $F$ testine ilişkin manidarlık değeri ,006'dır. Bu değer ,05'ten küçük olduğundan, korelasyon katsayısının 0'dan manidar düzeyde farklı olduğu söylenebilir. LogPctSAT değişkeninin de analize dâhil edildiği çoklu regresyon analizi sonuçlarına bakıldığında, cevaplanması gereken iki soruyla karşılaşılmaktadır. Bu sorulardan ilki, her iki yordayıcı değişkenin aynı anda analize dahil edildiği duruma ilişkin korelasyon katsayısının 0'dan manidar düzeyde farklı olup olmadığıdır. Her bir yordayıcı değişkenin, araştırılan ilişkiye şans sonucu gerçekleşmesi beklenenden daha fazla bir katkı sunup sunmadığı cevaplanması gereken ikinci sorudur. ANOVA tablosuna bakıldığında, $F$ = 182,834 ve $p$ = ,000 olarak görülmektedir. Bu sonuç, iki yordayıcı değişkenin kullanılması durumunda korelasyonun 0'dan manidar düzeyde farklı olduğunu gösterir. İlerleyen kesimlerde her bir yordayıcı değişkenin manidarlığı ayrı ayrı değerlendirilecektir.

**regresyon katsayıları**

Analiz çıktılarındaki en ilginç kısım "Katsayılar" tablosudur. Her iki yordayıcı değişkenin aynı anda analize alınması durumuna ilişkin tüm **regresyon katsayıları** bu tabloda görülmektedir. Tıpkı

$$\hat{Y} = bX + a$$

eşitliği ile gösterilen basit regresyonda olduğu gibi, çoklu regresyon denklemi aşağıdaki şekilde yazılabilir.

$$\hat{Y} = b_1X_1 + b_2X_2 + b_0$$

Bu denklemdeki $X_1$ ile $X_2$ yordayıcı değişkenler, $b_0$ ise regresyon doğrusunun y eksenini kesim noktasıdır. Her iki yordayıcı değişkene ait katsayılar tabloda görülmektedir. Tablodaki değerlere bakıldığında, Eğitim Harcamaları değişkenine ($X_1$) ait katsayı ($b_1$) 11,129 ve LogPctSAT değişkenine ($X_2$) ait katsayı ($b_2$) -78,203'tür. Bu katsayıların işareti değişkenler arasındaki ilişkinin yönü (pozitif ya da negatif) hakkında bilgi vermektedir. Bu değerlerin sabit ile toplanmasıyla regresyon eşitliği elde edilir.

$$\hat{Y} = 1147{,}100 + 11{,}129(\text{Eğitim Harcamaları}) - 78{,}203(\text{LogPctSAT})$$

Bu denklemdeki 1147,100 değeri regresyon analizi sonuçlarına ait "sabit (constant)" olup genellikle $b_0$ ile gösterilir. Regresyon denklemindeki sabit, Eğitim Harcamaları ve LogPctSAT değişkenleri sıfır olduğunda SAT puanının alacağı değerdir. Verilen örnek için Eğitim Harcamaları ve LogPctSAT değişkenleri hiçbir zaman sıfıra eşit olmayacaksa da sabit değere ihtiyaç duyulmaktadır. Çünkü sabit, yapılan yordamaların ortalamasının gerçek değerlerin ortalamasına yaklaşmasını sağlar. Ancak, buna gerçek anlamda dikkat edilmesi oldukça nadirdir. Çoklu regresyon analizindeki eşitlik Bölüm 9'da sunulan basit regresyon analizindeki eşitlik ile aynı şekilde kullanılabilmektedir. Herhangi bir eyaletin Eğitim Harcamaları ve LogPctSAT değerleri verildiğinde, bu eyalete ait SAT puanları yor-

danabilir. Örneğin; Eğitim Harcamaları ve LogPctSAT değeri dikkate alındığında Colorada eyaleti için yordanan SAT puanı;

$$\hat{Y} = 1147{,}100 + 11{,}129(\text{Eğitim Harcamaları}) - 78{,}203(\text{LogPctSAT})$$
$$= 1147{,}100 + 11{,}139(5{,}443) - 78{,}203(\text{Log}(29))$$
$$= 1147{,}100 + 11{,}139(5{,}443) - 78{,}203(3{,}367) = 944{,}366$$

olacaktır. Colorado eyaletinin gerçek SAT puanları ortalaması 980'dir, buna göre ortalamayı bir miktar küçük yordamış oluyoruz. Virginia eyaletine ait gerçek SAT puanları ortalaması ile yordanan SAT puanı ortalaması arasındaki fark 980 - 944,366 = 35,648 olup bu değer **artık (kalan) hata** olarak adlandırılır. SAT puanlarının büyüklüğü dikkate alındığında artık hataya ait bu değerin küçük olduğu söylenebilir. Diğer yandan, Batı Virginia eyaleti için bu değer -61,510'dur.

artık (kalan) hata

Eğitim Harcamaları değişkeninin katsayısının pozitif olması, *LogPctSAT değişkeni kontrol altına alındığında*, Eğitim Harcamaları ile SAT performansı arasında pozitif ilişki bulunduğu anlamına gelmektedir. Buna göre, eyaletlerin eğitim harcamaları arttıkça SAT performansları da artmaktadır. Ayrıca Eğitim Harcamaları değişkeni kontrol altına alındığında, LogPctSAT ile SAT değişkenleri arasındaki negatif ilişki bulunduğu şeklinde mantıklı bir sonuç karşımıza çıkmaktadır. Bir eyalette SAT'a giren öğrenci sayısı arttığında, eyaletin SAT başarısı ortalamasının daha düşük olmasının beklendiği ve bu beklentinin nedeni daha önce açıklanmıştı.

Ancak, LogPctSAT değişkeninin SAT değişkeni ile korelasyonunun -,93 olduğu dikkate alınmalıdır. Eğitim Harcamaları değişkeni ise, LogPctSAT ve SAT değişkenleri arasındaki bu ilişkiye önemli bir katkı sunmamaktadır. Eğitim Harcamaları değişkeni analize dahil edildiğinde, değişkenler arasındaki ilişki sadece ,941'e çıkmaktadır. Katsayılar tablosuna bakıldığında, sağ tarafta *t* ve "Mnd." (manidar) başlıklı iki sütun yer almaktadır. Bu sütunlardaki değerler, regresyon katsayılarının manidarlığıyla ilgilidir. Tablodaki "Mnd." sütununa bakıldığında, tüm katsayıların ,05 düzeyinde manidar olduğu ($p < ,05$) görülmektedir. Burada sabitin herhangi bir anlamı yoktur. Çünkü eyaletin eğitime herhangi bir harcama yapmadığı ve eyalette SAT'a giren öğrenci bulunmadığı anlamına gelmektedir. Eğitim harcamaları değişkenine ait katsayı manidardır ve eyalette SAT'a giren öğrencilerin yüzdesi kontrol altında alındıktan sonra, Eğitim Harcamalarının, öğrencilerin SAT başarılarıyla ilgili olduğuna işaret etmektedir. Benzer şekilde, Eğitim Harcamaları değişkeni kontrol altına alındığında, SAT'a giren öğrenci sayısının az olduğu eyaletlerde öğrencilerin SAT ortalamaları daha yüksektir. LogPctSAT yordayıcı değişkeninin yanında eğitim harcamaları değişkeni de analize dahil edildiğinde, korelasyon katsayısı çok yükselmese de, bu değişkenin regresyon analizine sunduğu katkı istatistiksel olarak manidardır.

Yukarıda, çoklu regresyon analizi sonuçlarının ne anlama geldiğini yorumlamanın yollarından biri tartışılmıştır. Eğer diğer değişkenler sabit tutulursa, herhangi bir yordayıcı değişkene ait eğim, bu yordayıcı değişken ile ölçüt değişken arasındaki ilişkiyi yansıtır. Burada bazı değişkenlerin sabit tutulmasıyla, bu değişkenler için tüm katılımcıların aynı puana sahip olması kastedilmektedir. Ancak regresyon analizini düşünmede yararlı olabilecek başka iki yol daha vardır.

## Çoklu Regresyon Analizinin Başka Bir Yorumu

Sadece Eğitim Harcamaları ile SAT puanları arasındaki korelasyona bakılıp LogPctSAT değişkeni tamamen göz ardı edildiğinde, SAT puanlarında doğrudan LogPctSAT değişkeni ile açıklanabilen bir miktar varyans bulunmaktadır. Bu varyans, SAT puanları ile Eğitim Harcamaları arasındaki şaşırtıcı negatif sonuca da kaynaklık eden faktördür. Dolayısıyla, hem SAT puanları hem de Eğitim Harcamaları değişkenleri LogPctSAT değişkeninin etkisinden arındırıldığında, SAT puanları ile eğitim harcamaları arasında nasıl bir korelasyon olacağı merak konusudur. Diğer bir deyişle, SAT puanlarındaki değişkenliğin bir kısmı Eğitim Harcamalarından, bir kısmı da LogPctSAT değişkeninden kaynaklanmaktadır. Hem SAT puanlarında hem de Eğitim Harcamalarında LogPctSAT değişkeninden kaynaklanan farklılıklar kontrol altına alındıktan sonra, düzeltilmiş SAT puanları ile eğitim harcamaları değişkenleri arasındaki korelasyonun ne olduğu sorusunun cevaplanması gerekir. Bu soruyu cevaplandırılabilmek için yapılması gereken işlemler göründüğünden daha

basittir. Çoklu regresyon analizinde doğrudan bu şekilde bir yol izlenmese de, sözü edilen işlemlerin analiz sürecinin anlaşılmasını kolaylaştıracağını düşünmekteyim.

artık

SAT puanlarının sadece LogPctSAT değişkeni ile yordanmasına yönelik bir regresyon analizi gerçekleştirilirse, SAT puanlarında LogPctSAT değişkeni ile açıklanabilen varyans miktarı elde edilir. Eğer yordanan SAT puanları gerçek puanlardan çıkarılırsa, elde edilecek **artık** "ResidSAT" olarak adlandırılır. SAT puanlarında LogPctSAT değişkeni ile açıklanamayan varyansın bir kısmı (ArtıkSAT) değerine karşılık gelir. Aynı işlem Eğitim Harcamalarının LogPctSAT değişkeni ile yordanmasına yönelik olarak da gerçekleştirebilir. Yordanan puanlar ile gerçek puanlar arasındaki fark hesaplandığında elde edilecek değer yine artıklar olacak ve Artık Eğitim Harcamaları (ResidExpend) olarak isimlendirilecektir. Yine bu ResidExpend, LogPctSAT değişkeninden bağımsızdır. Bu işlemlerin ardından, her ikisi de LogPctSAT değişkeninden bağımsız olan ResidSAT ve ResidExpend şeklinde iki artık puana ulaşılacaktır. Dolayısıyla, LogPctSAT değişkeni bu iki artık puan kümesi arasındaki ilişkide bir rol oynayamamaktadır.[2]

Bu işlemlerin ardından, standartlaştırılmış SAT puanlarının Eğitim Harcamaları puanları ile yordanmasına yönelik bir regresyon analizi yapılırsa (yani, ResidSAT ve ResidExpend değişkenleri ile bir regresyon analizi gerçekleştirilirse) aşağıdaki sonuçlar elde edilir.

**Katsayılar[a]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | t | Man. |
|---|---|---|---|---|---|---|
| | | B | Standart Hata | Beta | | |
| 1 | (Sabit) | -3,1E-015 | 3,608 | | ,000 | 1,000 |
| | Standartlaştırılmamış Artıklar | 11,130 | 3,230 | ,445 | 3,446 | ,001 |

[a] Bağımlı Değişken: Standartlaştırılmamış artık
SPSS programında rapor edilen analiz çıktılarının düzenlenmiş halidir.

Tabloya bakıldığında, düzeltilmiş SAT puanlarının düzeltilmiş Eğitim Harcamaları ile yordanmasına yönelik regresyon katsayısının 11,130 olduğu görülmektedir. Bu değer (yuvarlama hatası içerisinde) aynı işlemlerin Eğitim Harcamaları değişkeni ile normal şekilde yapıldığı regresyon analizindeki değere eşittir. Aşağıdaki tabloda, iki düzeltilmiş değişken arasındaki korelasyon katsayısı yer almaktadır. Tabloya göre, LogPctSAT değişkeninden kaynaklanan etkiler dışarıda tutulduktan sonra, Eğitim Harcamaları ile SAT puanları arasındaki korelasyon ,445 olmaktadır (Bu korelasyon katsayısının pozitif olduğuna dikkat edilmelidir).

**Model Özeti**

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Tahminin Standart Hatası |
|---|---|---|---|---|
| 1 | ,445[a] | ,198 | ,182 | 25,1077434 |

a Yordayıcılar: (Sabit), Standartlaştırılmamış Artık
b Bağımlı Değişken: Standartlaştırılmamış Artık
SPSS programında rapor edilen analiz çıktılarının düzenlenmiş halidir.

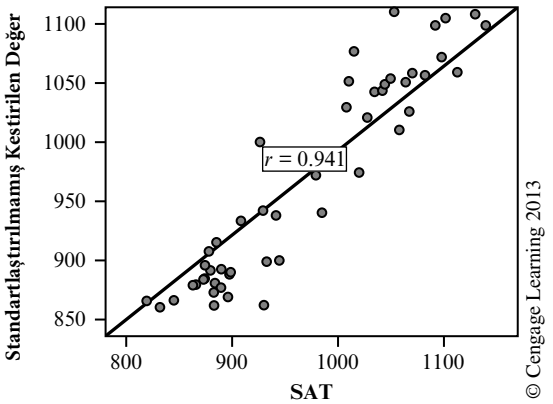
Hiç kimse, regresyon analizini bu şekilde yapmak zorunluluğunu düşünmez. Çoklu regresyon analizinde, çok sayıdaki yordayıcı değişkenin eşitlikteki diğer yordayıcı değişkenler için düzeltildiğini göstermek amacıyla bu tür bir işlem gerçekleştirilmiştir. Burada kullanılan kontrol altına almak, düzeltmek ve sabit tutmak ifadeleriyle aynı şey kastedilmektedir.

## Çoklu Regresyon Analizini Düşünmenin Son Bir Yolu

Çoklu regresyon analizini düşünmenin bazı yönleriyle faydalı bulduğum bir yolu daha vardır. Çoklu regresyon analizinde, aşağıdaki formda olan bir eşitliğin çözümü ile

$$\hat{Y} = b_1X_1 + b_2X_2 + b_0$$

2 SPSS programında bu artık değerlerin elde edilmesi oldukça kolaydır. Analiz gerçekleştirilirken, regresyon penceresinde, "Save" düğmesine basılıp "Unstandardized residuals" işaretlenirse artık değerler veri dosyasına kaydedilecektir.




**Şekil 15.2** SAT Puanı ile SAT puanlarının yordanmasına yönelik regresyon eşitliğindeki yordayıcı değişkenlerin en uygun doğrusal birleşimi arasındaki ilişkiyi gösteren dağılımın grafiği

ilgilenmekteyiz. Yani kullanmakta olduğumuz değişkenler yönünden

$$\widehat{\text{SAT}} = b_1(\text{Eğitim Harcamaları}) + b_2(\text{LogPctSAT}) + b_0$$

eşitliğinin çözülmesini amaçlanmaktayız. Buna göre yordanan SAT puanını hesaplamak için, aşağıdaki eşitliği elde ettim:

$$\widehat{\text{SAT}} = 11{,}129 \times (\text{Eğitim Harcamaları}) - 78{,}203 \times (\text{LogPctSAT}) + 1147{,}100$$

Yordanan SAT puanı, YordSAT şeklinde gösterilmektedir. Eğer gerçek SAT puanı ile yordanan SAT puanı arasındaki korelasyona bakılırsa, çoklu korelasyona karşılık gelen ,941 değeri elde edilir. Bu ilişkiye ait dağılımın grafiği Şekil 15.2'de gösterilmiştir.

Bu son yaklaşımın odak noktası; çoklu korelasyon katsayısının, ölçüt (SAT) ile yordayıcı değişkenlerin en uygun doğrusal birleşimi arasındaki basit korelasyon olarak düşünülebileceğini gösterir. "En uygun doğrusal kombinasyon" ifadesiyle, değişken olan SAT puanlarının yordayıcı değişkenler ile daha iyi bir şekilde kestirilmesine ait bir ağırlıklandırma (kısmi regresyon katsayıları) bulunmadığını kastetmek istiyorum. Bu oldukça önemli bir husustur. İstatistikte, bu kitapta yer verilmemiş oldukça ileri teknikler vardır. Bunlar, bazı değişkenlerin maksimum ağırlıklı toplamından yeni bir değişken oluşturmakta ve elde edilen bu yeni değişkenin, analizin ana parçası olarak kullanılmasını sağlamaktadır. Bu yaklaşım, iki değişken arasındaki ilişki negatif olsa bile çoklu korelasyon katsayılarının neden daima pozitif olduğunu da açıklamaktadır. Kestirilen değerin ölçüt ile pozitif ilişkili olması kesinlikle beklenen bir sonuçtur.

## Özet

Yukarıda açıklananlar, çoklu regresyon analizini düşünmenin farklı yolları olduğunu gösterir. Bu yollardan her biri, regresyon analizinin arkasında yatan mantığı anlamak için farklı bir bakış açısı getirir. Bu yollardan hangisi size daha mantıklı görünürse, o yaklaşıma odaklanmanız önerilir.

- Regresyon katsayısı, dikkate alınan bir değişken dışındaki yordayıcı değişkenler açısından farklılık göstermeyen bir grup eyalet söz konusu olduğunda, elde edilecek katsayı olarak ifade edilebilir. Diğer bir deyişle; bir değişken hariç tüm yordayıcı değişkenler sabit tutulmakta ve sabit tutulmayan bu tek yordayıcı değişkenin yordayıcının oluşturduğu değişkenliğe bakılmaktadır.
- Eğer iki değişkenden birini, kontrol altına almak istediğimiz değişken için düzeltirsek, çoklu regresyon analizindeki regresyon katsayısı, basit regresyon analizindeki katsayı gibi düşünülebilir. Verilen örnek için düzeltme uygulanması anlamına gelir (bu değişken için elde edilen puanlar ile kontrol altına alınacak değişkenden rahatsız edici değişken tahmin edilen puanlar arasındaki fark hesaplanarak). Bu şekilde elde edilen katsayı, çoklu regresyon analizinde hesaplanan katsayıya eşittir.

- Çoklu korelasyon katsayısı, ölçüt değişken ($Y$) ile yordayıcı değişkenlerin en uygun doğrusal kombinasyonundan oluşan değişken ($\hat{Y}$) arasındaki basit Pearson korelasyon katsayısı olarak düşünülebilir.

**NOT**

SAT'ı hazırlayan Eğitimde Ölçme Servisi, eyaletlerin performanslarını değerlendirmek için SAT'ın uygun bir karşılaştırma aracı olmayacağı konusunda uyarıda bulunmaktadır. Yani SAT, eyaletlerin performanslarını karşılaştırmak için adil bir gösterge değildir. Yukarıda verilen örneğe göre; SAT'ın farklı eyaletlerin performanslarını karşılaştırmak için uygun bir gösterge olmamasının nedenlerinden biri, farklı eyaletlerde sınava giren öğrenci sayılarının değişiklik göstermesidir. Diğer bir deyişle, SAT her bir öğrencinin performansını değerlendirmek için uygun bir sınav olsa da, farklı eyaletlerde SAT'a giren öğrenci sayısının değişiklik göstermesi bu sınavı eyaletlerin performanslarını değerlendirmek için uygun bir yol olmaktan uzaklaştırmaktadır. Bu sorunun üstesinden gelebilmek için herhangi bir eyalette SAT'a katılan öğrencilerin yüzdesinin doğal logaritmasının alınmasıyla elde edilen ve LogPctSAT olarak adlandırılan yeni bir değişken oluşturulabilir. Ancak bu yeni değişkenin kullanımı konusunda da dikkatli olunmalıdır. Çünkü bu değişkenin adil olma olasılığı bulunsa da; dikkate alınmayan başka değişkenlerin de bulunması olasılığı vardır.

## 15.2 Ek Yordayıcıların Kullanılması

Çoklu regresyon analizinin diğer özelliklerine bakmadan önce, SAT puanlarını yordamak için ek değişkenler kullanılırsa nasıl bir sonuçla karşılaşacağımız sorusunun cevaplanması gerekir. Verilen örnek için, regresyon analizine dâhil edilemeyen ancak potansiyel olarak kullanılabilecek iki değişken (öğrenci/öğretmen oranı ve öğretmen maaşları) bulunmaktadır. Regresyon analizine bu iki değişken de dâhil edilebilir. Ancak, burada sadece öğrenci/öğretmen oranı değişkeni analize eklenecektir. Geleneksel inançlara göre, öğrenci/öğretmen oranı düşük olduğunda öğrencilerin performansının daha yüksek olması beklenmektedir. Öğrenci/öğretmen oranın düşük olması, eğitim harcamalarındaki artışı beraberinde getirebilir. Dolayısıyla, öğrenci/öğretmen oranı ile Eğitim Harcamaları değişkenlerinin kesiştikleri bir alan bulunmaktadır. Buna bağlı olarak; öğrenci/öğretmen oranı, eğitim harcamaları değişkeniyle örtüşebilir ve Eğitim Harcamalarının ötesinde yeni bir bilgi sağlamayabilir.

Tablo 15.3, eğitim harcamaları, LogPctSAT, öğrenci/öğretmen oranı değişkenleri ile SAT puanlarının yordanmasına ait çoklu regresyon analizi sonuçlarını göstermektedir. Bu tablo hakkında pek çok yorum yapılabilir.

Bu analiz sonucunda elde edilen regresyon eşitliği aşağıdaki gibi yazılabilir:

$$\hat{Y} = 1132{,}033 + 11{,}665(\text{Eğitim Harcamaları}) - 78{,}393(\text{LogPctSAT}) - {,}742(\text{Öğrenci/Öğretmen oranı})$$

Burada, eğitim harcamaları ve LogPctSAT değişkeninin regresyon modeline katkısı hâlâ manidardır (sırasıyla, $t$ = 3,302 ve –17,293). Fakat, öğrenci/öğretmen oranı değişkeninin regresyon modeline katkısı manidar değildir ($t$ = ,418). Buna göre, öğrenci/öğretmen oranı

**Tablo 15.3** Öğrenci/öğretmen oranı değişkenin analize dâhil edildiği regresyon modeli

**Katsayılar[a]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | t | Mnd. | Eş Doğrusallık İstatistikleri | |
|---|---|---|---|---|---|---|---|---|
| | | B | Standart Hata | Beta | | | Tolerans | VIF |
| 1 | (Sabit) | 1132,033 | 39,787 | | 28,452 | ,000 | | |
| | Eğitim Harcamaları | 11,665 | 3,533 | ,212 | 3,302 | ,002 | ,596 | 1,679 |
| | LogPctSAT | -78,393 | 4,533 | -1,042 | -17,293 | ,000 | ,679 | 1,473 |
| | Öğr./Öğret. Oranı | ,742 | 1,774 | ,022 | ,418 | ,678 | ,854 | 1,171 |

[a] Bağımlı Değişken: SAT

SPSS programında rapor edilen analiz çıktılarının düzenlenmiş halidir.

tolerans

Varyans Bükülme Faktörü (VBF)

eşdoğrusallık

çoklu doğrusallık

değişkeninin analize eklenmesi SAT puanlarının yordanmasına ait regresyon modelinde bir iyileşmeye neden olmamaktadır. (Öğrenci/öğretmen oranı ile SAT puanları arasındaki basit korelasyon da manidar ($r$ = ,081) değildir). Tablo 15.3'te, **Tolerans** ve **Varyans Bükülme Faktörü (VBF)** olarak adlandırılan iki yeni sütun göreceksiniz. Yordayıcı değişkenlerin kendi aralarındaki korelasyonlara bakıldığında, **eşdoğrusallık (collinearity)** ve **çoklu doğrusallık (multicollinearity)** kavramları karşımıza çıkmaktadır. Eşdoğrusallık regresyon katsayısındaki standart hatayı artıran, güven aralığını genişleten ve regresyon katsayısına ait $t$ değerini düşüren bir etki oluşturur. VBF sıralanan özellikleri ölçen bir istatistiktir. Ayrıca iki yordayıcı değişken arasındaki korelasyon yüksek olduğunda, bu değişkenlerden biri regresyon denkleminin kararsızlığını arttırdığı gibi, regresyon modeline de bir katkı sağlamamaktadır.

Tolerans katsayısı VBF katsayısının tersi olup $1 - R_j^2$ ifadesiyle hesaplanır. Burada $R_j$, $j$ değişkeni değişken$_j$ ile diğer tüm *yordayıcı* değişkenler arasındaki korelasyon katsayısıdır. Buna bağlı olarak, VBF değerinin düşük, Tolerance faktörünün ise yüksek olmasını isteriz. Tolerans faktörü iki özellik için bir gösterge niteliği taşır. Biri, yordayıcı değişkenler arasında örtüşen varyans miktarını göstererek hangi yordayıcıların ortak hangi yordayıcıların ise bağımsız bilgiler sağladığının anlaşılmasına yardımcı olur. (Yüksek tolerans katsayısı değişkenler arasında çakışan varyans miktarının düşük olduğuna işaret eder.) İki değişkenin kendi başlarına sağladıkları bilginin çakışması, bu değişkenlerden birinin analizden çıkarılması için yeterli bir neden değildir. Ancak böyle bir durum, söz konusu iki değişkenin ortak katkısının beklenenden daha az olabileceğinin bir sinyali olabilir.

İkinci olarak tolerans istatistiği, modeldeki potansiyel kararsızlık sorunlarının görülmesini sağlar. Tolerans değeri çok düşük olduğunda, modelin kararlılığı ve hatta bazen aritmetik doğruluğu tehlikeye girebilmektedir. Bir yordayıcı değişkenin diğer değişkenler ile mükemmel yordanabildiği uç durumlarda, **tekil (singular)** kovaryans (veya korelasyon) matrisi olarak adlandırılan bir durumla karşı karşıya kalırız. Böyle bir durumda birçok program herhangi bir model oluşturmadan analizi sonlandırmaktadır. Analiz çıktılarında "tekil matris (matrix singular)" veya "pozitif tanımlı değil (not positive-definite)" şeklinde bir sonuç ile karşılaşırsanız bunun en muhtemel sebebi tolerance katsayısının 0,00 olması ve yordayıcı değişkenlerden herhangi birinin diğer yordayıcı değişkenler tarafından mükemmel düzeyde tahmin edilebiliyor olmasıdır. Bu durumda, söz konusu etkiden kurtulmak için en azından bir yordayıcı değişkenin analizden çıkarılması gerekir. Böyle bir ilişkinin en sık ortaya çıktığı durumlar, yordayıcı değişkenlerden herhangi birinin diğer yordayıcı değişkenlerin toplamına veya ortalamasına karşılık geldiği veya tüm yordayıcı değişkenlerin bir sabit ile toplandığı durumlardır.

tekil

Ortak yapılan hatalardan biri; $b_1$'nin bağıl büyüklüklerine, her bir yordayıcı değişkenin bağıl önemi gibi davranılmasıdır. Böyle bir mantığa göre, eğitim harcamalarının LogPctSAT değişkeninden daha az önemli olduğu şeklinde bir sonuca varılabilir. Çünkü eğitim harcamaları değişkeninin katsayısı (11,129), LogPctSAT değişkeninin katsayısından (-78,203) önemli ölçüde daha düşüktür. Belki gerçekten eğitim harcamaları daha az önemli bir yordayıcı olabilir. Ancak, regresyon katsayılarına bakarak böyle bir sonuca ulaşamayız. Katsayıların büyüklükleri, kısmen bu katsayılara karşılık gelen değişkenlerin standart sapmalarının bir fonksiyonudur. LogPctSAT değişkeninin standart sapması Eğitim Harcamaları değişkeninin standart sapmasına göre daha düşük olduğundan, bu değişkene ait regresyon katsayısı ($b_2$) Eğitim Harcamaları değişkenine ait regresyon katsayısından daha büyük olma eğilimindedir. Bu eğilim değişkenlerin regresyon denklemindeki öneminden bağımsızdır.

Probleme biraz farklı bakılırsa, bu son noktayı anlamak daha kolay olabilir. (Burada, Eğitim Harcamaları ve öğrenci/öğretmen oranı değişkenlerinin yordayıcı olarak alındığı bir örnek sunacak ve bu örnekte LogPctSAT değişkenine yer vermeyeceğim. Bu, konuyu örneklendirmemi kolaylaştıracaktır.) İki eyaletin eğitim harcamaları arasında bir birimlik bir fark olması halinde, bu iki eyaletin başarıları arasında önemli bir fark olacaktır (eyaletlerin eğitim harcamaları arasındaki fark sadece 6 birimdir). Diğer taraftan öğrenci/öğretmen oranı değişkenindeki bir birimlik fark eyaletlerin SAT başarılarında önemli bir değişikliğe yol açmayacaktır (öğrenci/öğretmen oranı değişkeninin ranjı 77 birimdir). Bu iki yordayıcı değişkenin bağıl önemlerinden bağımsız olarak, önceki bilgimize göre her iki değişken için bir puanlık değişimin SAT puanlarında aynı farkı oluşturması mantıklı bir beklenti olmayacaktır.

## Standartlaştırılmış Regresyon Katsayıları

**önem**

Daha sonra göreceğimiz gibi, değişkenlerin bağıl önemine ilişkin problemin, **önem** ile ne kastedildiğine bağlı olarak değişen cevapları vardır. Bununla birlikte, kendi içinde ayrı bir istatistik olduğundan, önemin bir ölçüsünden burada söz edilmelidir.

Çoklu regresyon eşitliğini elde etmeden önce, her bir değişkenin standartlaştırılmış halinin elimizde mevcut olduğunu varsayalım. Hatırlayacağınız gibi, standartlaştırılmış değişkenlerin ortalaması 0 ve standart sapması 1'dir. Bu ayrıca, standart sapma birimlerindeki sonuçları da ifade eder. (Etki büyüklüklerinin çoğunun standart sapmaya bölünerek standartlaştırıldığı hatırlanmalıdır.) Şimdi tüm değişkenler eşit standart sapmaya (1) sahip olacaktır. Dolayısıyla, bir değişken için herhangi iki eyaletin puanları arasındaki bir birimlik fark, başka bir değişken için söz konusu iki eyalet arasındaki bir birimlik fark ile karşılaştırılabilir hale gelecektir. Eğer standartlaştırılmış değişkenler kullanılarak regresyon eşitliği çözümlenirse aşağıdaki eşitlik elde edilir:

$$\hat{Y}_Z = 0{,}203Z_{\text{Eğitim Harcamaları}} - 1{,}040Z_{\text{LogPctSAT}}$$

**standartlaştırılmış regresyon katsayıları**

Bu eşitlikte, Z standartlaştırılmış değişkenleri gösterir. Bu durumda regresyon katsayıları **standartlaştırılmış regresyon katsayıları** olarak adlandırılmakta ve SPSS çıktılarında Beta olarak adlandırılıp $\beta_i$ ile gösterilmektedir. Buna göre,

$$\beta_1 = 0{,}203$$

$$\beta_2 = -1{,}040$$

olur. Değişkenler standartlaştırıldığında, sabit 0'a eşit olur ve dolayısıyla gösterilmez.

$\beta_i$'nin yukarıdaki değerlerinden hareketle şu sonuca varabiliriz: $Z_1$ için eyaletler arasındaki bir birimlik fark (bir standart sapma büyüklüğündeki fark) $\hat{Y}_Z$'de 0,203 standart sapma büyüklüğünde bir fark yaratırken; $Z_2$ değişkeni için eyaletler arasındaki bir birimlik fark $\hat{Y}_Z$'de -1,040 birimlik bir fark oluşturmaktadır. Bu durumda, LogPctSAT değişkeni eğitim harcamalarından daha önemli bir yordayıcıymış gibi görünebilir. $\beta_i$ değerlerinin bağıl büyüklükleri, değişkenlerin öneminin en iyi göstergesi olmamakla birlikte; bu katsayılar regresyon analizini gerçekleştiren birçok bilgisayar programında rapor edilmekte ve en azından değişkenlerin eşitlikteki bağıl katkılarına ilişkin kaba bir tahmin vermektedir. Standartlaştırılmış regresyon katsayıları, regresyon analizini gerçekleştirmek için kullanılan bilgisayar programlarının neredeyse tümü tarafından rapor edilmektedir.

# 15.3 Regresyon Katsayılarının Standart Hatası ve Test Edilmesi

Regresyon katsayılarını (standartlaştırılmış veya standartlaştırılmamış) elde ettikten sonra, bu katsayıların manidarlığının test edilmesi gerekir. Örneğin, Eğitim Harcamaları değişkenin SAT puanlarının yordanmasına hizmet eden bir değişken olduğunu söyleyebilmek için eğitim harcamalarının SAT puanlarının yordanmasına ait katkısı 0'dan manidar düzeyde büyük olmalıdır. İstatistiksel manidarlık test edilirken, ham puanların ($b_i$) mı yoksa standartlaştırılmış puanların ($\beta_i$) mı test edildiğinin bir önemi yoktur. Ham puanlar ile standart puanlar birbirlerinin doğrusal dönüşümleridir ve hangisi kullanılırsa kullanılsın aynı istatistikler elde edilir.

Regresyon katsayılarını test edebilmek için (veya bu konudaki diğer birçok istatistik), bu istatistiğin standart hatasının bilinmesi gerekir. $b_i$ katsayıları için standart hatalar Gösterim 15.1'de verilmiş ve Std. Hata (Std. Error) olarak adlandırılmıştır. Örneğin, $b_0$ (sabit) için standart hata 16,700, $b_1$ değişkenine ait standart hata ise 3,264'tür. Diğer standart hatalar gibi, regresyon katsayılarının standart hataları da, tekrarlı örneklemler üzerinde regresyon istatistiklerinin değişkenliğine ait bir ölçüdür. Çalışmanın birbirinden bağımsız öğrenci grupları üzerinde birçok kez tekrarlandığını varsayalım (bunun yapılması mümkün değilse de, böyle bir işlemin yapıldığı varsayılabilir). Her bir öğrenci grubu için farklı bir $b_1$ değeri elde edilir. Elde edilen ve evrendeki gerçek değerin yansız bir kestirimi olan bu değerleri $b_1^*$ ile göstereceğiz. Çok sayıda $b_1$ katsayısı değerleri, $b_1$ değişkeninin standart hatası olan 3,264 değeri ile $b_1^*$ etrafında normal bir dağılıma sahip olacaktır.

Regresyon katsayılarına ilişkin *t* testini yapmak için bu standart hataları kullanabiliriz. Özel olarak $N - p - 1$ serbestlik derecesinde[3]

$$t = \frac{b_j - b_j^*}{s_{b_j}}$$

ifadesiyle hesaplanır. Daha sonra $H_0$ hipotezini test etmek için; $H_0$: $b_j^* = 0$

$$t = \frac{b_j}{s_{b_j}}$$

değeri elde edilir. Eğitim Harcamaları değişkenine ait regresyon katsayısını test ettiğimizde,

$$t = \frac{11,129}{3,264} = 3,410$$

değerini elde ederiz. Bu, $N - p - 1 = 50 - 2 - 1 = 47$ serbestlik derecesindeki standart Student *t* dağılımıdır. Ek *t*'deki tabloya göre 47 serbestlik derecesindeki kritik değer 2,01'dir. Buna göre, $H_0$ hipotezini redderiz ve evrendeki regresyon katsayısının 0'a eşit olmadığı sonucuna ulaşırız. Ekte *t*'de kritik *t* değerlerinin bulunduğu ifade edilmiş olsa da, analiz sonuçlarını yorumlamak için bu değerlere bakılmasına gerek yoktur. Çünkü regresyon analizi sonucunda elde edilen çıktılarda, sadece *t* değerleri değil, *t* değerlerinin manidarlık düzeyi (iki yönlü) de rapor edilmektedir. Buna göre, Eğitim Harcamaları için 11,130 büyüklüğündeki bir *b* değeri ,001 düzeyinde manidardır. Bir başka deyişle, Eğitim Harcamaları arttıkça kestirilen *Y* değeri de artmakta ve Eğitim Harcamaları değişkeni SAT puanlarının yordanmasına ait regresyon modeline manidar katkı sağlamaktadır.

Benzer işlemler LogPctSAT değişkeni için tekrarlandığında,

$$t = \frac{-78,203}{4,471} = -17,491$$

elde edilir. Bu sonuç da manidardır ($p =$ ,000). Bu sonuca göre, LogPctSAT değişkeninin de SAT puanlarının yordanmasına ait regresyon modeline Eğitim Harcamalarına göre daha fazla manidar katkı sağladığı söylenebilir. Öğrenci/öğretmen oranı değişkeni modele eklendiğinde, ,481'lik bir *t* değeri elde edilir. Bu *t* değeri manidar değildir. Dolayısıyla bu değişkenin modelden çıkarılması düşünülebilir. İlerleyen kısımlarda bu konu üzerine başka yorumlar da yapılacaktır. Burada, herhangi bir değişkene ilişkin istatistiksel testlerin eşitlikteki diğer tüm değişkenler bağlamında yapıldığının fark edilmesi önemlidir. Herhangi bir yordayıcı değişken ölçüt değişken ile yüksek korelasyona sahip olabilir. Nitekim öğretmen maaşı değişkeni ile SAT puanları arasında $r$ = -0,440'lık ($p$ = ,001) manidar bir korelasyon bulunmasına karşın; öğretmen maaşı değişkeni SAT puanlarının yordanmasına ait regresyon modeline manidar bir katkı sağlamamaktadır. (Öğretmen maaşları ile Eğitim Harcamaları arasındaki korelasyon ,87'dir. Yani, eğitim harcamaları değişkeni dikkate alındığında geriye maaşı açıklayacak fazla bir şey kalmamaktadır.)

Bazı bilgisayar programları, standart regresyon katsayıları ($\beta_i$) için standart hataları rapor ve test etmeyi tercih etmektedir. Bu tercih, yapılan işlemlerde herhangi bir farklılığa neden olmaz. Benzer şekilde, bazı programlarda *t* testi yerine *F* testi (1 ve $N - p - 1$ serbestlik derecesinde) rapor edilmektedir. Bu *F* değeri *t* değerinin karesine eşittir ve yine analiz sonuçlarında bir farklılaşmaya neden olmaz.

## 15.4 Yeniden Örnekleme Yaklaşımı

Bu kitabın çeşitli bölümlerinde Student *t* testi ve Ki-kare gibi geleneksel hipotez test etme yöntemlerini tartıştık, daha sonra, aynı veya benzer hipotezleri test etmek için yeniden örnekleme yaklaşımını ele aldık. Geleneksel testler mâkul bir süre içerisinde yeniden örnekleme yapmanın mümkün olmadığı durumlar için geliştirilmiştir. Yeniden örnekleme

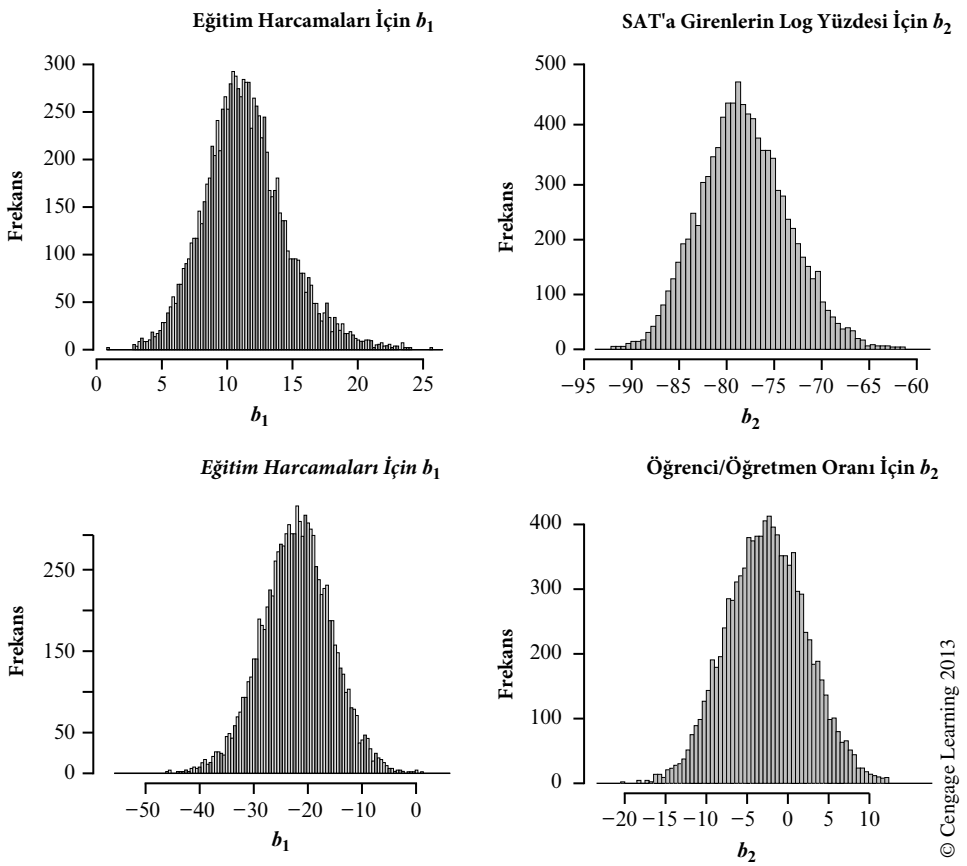
3 Çok sayıda yazar (örneğin, Draper ve Smith, 1981; Huberty, 1989), bu istatistiğin tam anlamıyla Student *t* gibi bir dağılım olmadığına dikkat çekmiştir. Bununla birlikte genel olarak; öyleymiş gibi muamele edilir. Ancak ilgili Student *t* dağılımına ilişkin olasılık harfi harfine alınmamalıdır.

yaklaşımı henüz geleneksel testlerin yerini alacak kadar ön plana çıkmasa da, bu yöne doğru gitmektedir.

Veri kümemizin 50 eyaletten oluşan bir evrene sahip olduğunu düşünelim. Evrenden alınan örneklemi değiştirerek önyükleme (bootstrapping) ile *çekilen yerine konularak* örneklemler oluşturulabilir. Bazı eyaletlerin birden çok defa temsil edileceği ve bazı eyaletlerin dâhil olmayacağı bir örneklem oluşturabiliriz. Daha sonra, veri kümelerinden bir regresyon eşitliği oluşturur ve regresyon katsayılarını kaydederiz. Ardından bu işlemi 9.999 defa tekrarlayarak katsayıları kaydederiz. Sonrasında, elde edilen katsayıların grafiğini oluşturur ve bu katsayıları dikkate alarak herhangi bir yüzdelik için güven aralığı limitlerini hesaplarız. Söz gelimi, örneklem içerisindeki en düşük %2,5'lik değerler ile en yüksek %2,5'lik değerleri alarak %95 olasılıklı güven aralığı oluşturubiliriz. Bu sınırlar büyük ihtimalle, ortalama $b_i$ değerinin etrafında simetrik olmayan bir dağılım gösterecektir. Eğer bu aralık 0'ı içermez ise, regresyon katsayılarının 0'dan manidar düzeyde farklı olduğu sonucuna varabiliriz.

Bu yeniden örnekleme işlemini iki farklı durum için gerçekleştirdim. Bunlardan ilkinde, Eğitim Harcamaları ve LogPctSAT değişkenlerini birleşik SAT puanlarını kestirmek için kullandım. Yapılan analizde her iki değişkenin de birleşik SAT puanlarının manidar birer yordayıcısı olduğu belirlenmiştir. Daha sonra, değişkenlerden herhangi birinin manidar bir yordayıcı olmadığı bir örnek sunmak için Eğitim Harcamaları ile öğrenci/öğretmen oranı değişkenlerini kullanarak regresyon analizine ait işlemleri tekrarladım. Bu sonuçlar Şekil 15.3'te görülmektedir.

Üst satırda, iki manidar yordayıcının kullanıldığı regresyon eşitliğine ait katsayılar bulunmaktadır. Dikkat edilirse yeniden örneklenmiş $b_i$ katsayıları, Eğitim Harcamaları değişkeni için daima pozitif, LogPctSAT değişkeni için de daima negatiftir. Bu beklenen bir sonuçtur. Bu model için %95 olasılıklı güven aralığı Eğitim Harcamaları değişkeni için $5{,}70 \le b^* \le 18{,}56$ ve LogPctSAT değişkeni için $-86{,}88 \le b^* \le -68{,}65$'tir.




**Şekil 15.3** Regresyon katsayılarının örneklem dağılımları

Fakat, eğitim harcamaları değişkeni ile birlikte öğrenci/öğretmen oranı değişkeni yordayıcı olarak kullanıldığında, eğitim harcamaları değişkeninin katsayılarından sadece birkaç tanesi pozitif olmaktadır. Öte yandan, öğrenci/öğretmen oranı değişkenine ait katsayıların pek çoğu negatiftir; birçoğu pozitiftir. Bu ikinci aralık 0'ı içerdiğinden yokluk hipotezi kabul edilir. Bu modelde, %95 olasılıklı güven aralığı Eğitim Harcamaları değişkeni için $-35{,}34 \le b^* \le -9{,}76$ ve öğrenci/öğretmen oranı için $-12{,}04 \le b^* \le 7{,}29$ bulunur. Bu aralıklar, geleneksel yöntemden elde edilen sonuçlara paraleldir.

**NOT**

Önyükleme yaklaşımını kullanma nedenimiz, $b_j$ için daha iyi kestirim elde etmek değildir. Önyükleme yaklaşımı, $b_j$'nin örneklem değişkenliği (standart hata) hakkında tahminde bulunmak amacıyla kullanılmıştır. 10.000 defa yeniden örnekleme sonucunda, regresyon katsayılarının ortalaması, evrendeki regresyon katsayısına değil, orijinal örneklemdeki regresyon katsayına yaklaşmalıdır. Yapılan kestirimlere ait değişkenlikler bu kestirimlerin güven aralığının belirlenmesine imkân tanır.

## 15.5 Artık Varyans

İster geleneksel yöntem isterse yeniden örnekleme yaklaşımı kullanılmış olsun, evrendeki regresyon katsayısı için yapılan kestirimlerin bir örnekleme hatası içerdiği göz önünde bulundurularak standart hata kavramını daha önce ele almıştık. Biraz farklı bir hata türü, *Y*'lerin yordanan değişkenine ait kestirim için de söz konusudur. SAT puanları açısından, SAT puanlarının, Eğitim Harcamaları ve LogPctSAT vb. değişkenlerin en azından bir kısmının bir fonksiyonu olması beklenir. (Eğer böyle bir beklenti olmasaydı, başlangışta bu değişkenlerle ilgili bir veri toplamazdık.) Değişkenler hatasız bir biçimde ölçülebilse ve kestirmeler hata karışmadan yapılabilse dahi, iki veya üç değişkenin *Y*'yi mükemmel şekilde yordaması beklenemez. Tüm yordayıcı değişkenler dikkate alınsa bile, *Y* için yapılan kestirmelere bir miktar hata karışacaktır. Bu hata, **artık varyans** veya **artık hata** olarak adlandırılmakta, $KO_{\text{artık}}$, $KO_{\text{hata}}$ veya $s^2_{0,12345}$ ile gösterilmekte ve

artık varyans

artık hata

$$\frac{\sum (Y - \hat{Y})^2}{N - p - 1}$$

ifadesiyle tanımlanmaktadır. Gösterim 15.1'deki varyans analizi özet tablosunda hata varyansı, hata terimi olarak gösterilmektedir ve değeri 664,646'dır.

*Y* hakkında yapılacak kestirimlerin olabildiğince doğru olması için artık hatanın çalışmada minimum düzeyde tutulması gerekir. Bundan dolayı, artık hata kavramı son derece önemlidir. Daha sonraki bölümde bu kavrama tekrar döneceğiz.

$KO_{\text{artık}}$'ın karekökü *kestirimin standart hatası* olarak adlandırılır. Bölüm 9'da yer verilen standart hata ile aynı anlama sahip olan bu değer artık puanlar $(Y - \hat{Y})$ sütununun standart sapmasıdır. Kestirimin standart hatası Gösterim 15.1'de, varyans analizi tablosundan önce yer alan Model Özeti'nde "Std. Error of Estimate" başlığı altında verilmiştir. Buradaki örnek için kestirimin standart hatası 25,781 olarak hesaplanmıştır.

## 15.6 Dağılım Varsayımları

Buraya kadar değişkenlerin doğası hakkında herhangi bir varsayımda bulunmadık. $b_i$, $\beta_i$ ve *R* istatistikleri (çoklu korelasyon katsayısı) akla uygun ölçümlerdir ve dağılım varsayımlarından bağımsızdır. Bunu söylemekle birlikte, eğer bu ölçümler önemli durumlarda kullanılacaksa bazı varsayımların gerekli olacağını söylemek gerekir. (Bölüm 9'a geri dönüp giriş kısmındaki kısa tartışmalar [s. 525] ile Kesim 9.7 ve Kesim 9.14'ü tekrar okumak

yararlı olabilir [s. 266 ve s. 290]. Bu kesimler doğrusal regresyon modelleri ile iki değişkenli-normal modeller arasındaki farkı açıklamakta ve ilgili varsayımları ele almaktadır.)

**çok değişkenli normal**

Tartışılan istatistiksel testlerin uygulanabilmesi için değişkenlerin doğasına bağlı olarak iki farklı varsayımdan birinin sağlanması gerekir. Eğer $X_1$, $X_2$,… $X_p$ tesadüfi değişkenler olarak düşünülüyorsa; $Y$, $X_1$, $X_2$,… $X_p$ değişkenlerinin birleşik dağılımının **çok değişkenli normal** olduğu varsayımının yapılması gerekir. Buradaki örnekte de yordayıcılar önceden saptanmayıp bulunduğu anda ölçüldüğünden tesadüfi değişkenler olduğu söylenebilir. (Kesim 9.12'de açıklanan iki değişkenli normal dağılımın çoklu değişkenlere yönelik bir uzantısıdır.) Teoride bu varsayım birçok test için gerekli olsa da; çok değişkenli-normal dağılımdan sapmalar tolere edilebilirdir. (Bu bizim için bir şanstır, çünkü Şekil 15.1'e bakıldığında, örnekte veri kümesi çok değişkenli normal-dağılım göstereceğe benzememektedir.) İlk olarak, uygulanan testler oldukça dayanıklıdır. İkinci olarak, uygulamada, $R$'nin 0'dan manidar fark gösterip göstermediğinden çok, büyüklüğü ile ilgilenmekteyiz. Bir başka deyişle, varyans analizindekinin aksine regresyon analizinde $R$'ye ilişkin hipotez testleri ile ilgilenmekteyiz. Örneğin, $R$ = ,10'a eşit bir katsayının, bağımlı değişkene ilişkin varyansın sadece %1'lik bir kısmını açıklayabileceğinden, böyle bir katsayının istatistiksel olarak manidar olup olmadığı önemli değildir.

Eğer $X_1$, $X_2$,…, $X_p$ değişkenleri sabit ise, $Y$'nin şartlı dağılımının (yani $X_i$ değişkenin belirli bir düzeyi için $Y$'nin dağılımının) normal ve bağımsız olduğu varsayımına kolaylıkla ulaşılabilir. Yine burada, normallikten orta düzeyde sapmalar tolere edilebilir. Regresyon analizine ilişkin varsayımların ne ölçüde karşılandığının değerlendirilmesine yardımcı olan regresyonun geçerliği ile ilgili tartışmalar sonraki kesimlerde derinlemesine ele alınmıştır.

Sabit model ve $Y$'nin normalliğine ilişkin varsayımlar Bölüm 16'da ele alınacaktır. Bu bölümde genel olarak tesadüfi değişkenlerle ilgileneceğiz. Çok değişkenli normallik varsayımı gereğinden katı bir varsayımdır, fakat yeterliliktir. Örneğin, pek az kimse $b_j$'nin standart hatasının hesaplanması ile ilgilenir; standart hatanın hesaplanması çok değişkenli-normallik varsayımını gerektirmez. Fakat, $b_j$'nin güven aralıklarının hesaplanması veya test edilmesi, normallik varsayımını gerektirir. Dolayısıyla örneğimizdeki veri kümesine çoklu normallik varsayımı uygulanacaktır.

## 15.7 Çoklu Korelasyon Katsayısı

**çoklu korelasyon katsayısı**

**$R_{0,123\ldots p}$**

Gösterim 15.1'e göre, SAT puanları ile iki yordayıcı değişken (Eğitim Harcamaları ve LogPctSAT) arasındaki çoklu korelasyon ,941 bulunmuştur. **Çoklu korelasyon katsayısı** genel olarak, $R_{\mathbf{0,123..p}}$ ile gösterilmektedir. Bu gösterim ($Y$ veya $X_0$) ölçüt değişkeninin aynı anda 1, 2, 3,..., $p$ yordayıcı değişkenleri ile yordandığını gösterir. Hangi yordayıcı değişkenlerin analize dâhil edileceğine dair bir karışıklık olmadığında, gösterimdeki alt indisler kullanılmamakta ve çoklu korelasyonu göstermek için sadece "$R$" yeterli olmaktadır.

Daha önce de gördüğümüz gibi $R$, $Y$ ölçüt değişkeniyle yordayıcı değişkenlerin en uygun doğrusal kombinasyonu arasındaki korelasyon olarak tanımlanmaktadır. Yani $R$,

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_pX_p$$

olmak üzere, $r_{Y\hat{Y}}$'den başka bir şey değildir.

Buna göre eğer istersek $\hat{Y}$'yi elde etmek için regresyon eşitliği kullanabilir, daha sonra Şekil 15.2'de yaptığımız gibi $Y$ ile $\hat{Y}$ arasındaki korelasyonu hesaplayabiliriz. $R$'nin bu şekilde hesaplanması önerilmemesine rağmen, çoklu korelasyon katsayının aslında neyi temsil ettiğinin anlaşılabilmesi için yararlı olabilir. Uygulamada, $R$ (veya $R^2$) çoklu regresyon analizini gerçekleştirmek amacıyla kullanılan tüm bilgisayar programlarında rapor edilmektedir. Örneğimiz için SAT puanları ile Eğitim Harcamaları ve LogPctSAT değişkenleri arasındaki çoklu korelasyon ,886'dır.

$R$ normal bir korelasyon katsayısıdır. Dolayısıyla, Pearson momentler çarpımı korelasyon katsayısına uygulanan işlemlerin tümü çoklu korelasyon katsayısına da uygulanabilir. (Bu doğrudur, çünkü $R = r_{Y\hat{Y}}$'dür.) Bununla birlikte, $R^2$ bağımlı değişkende yordayıcı değişkenler aracılığıyla açıklanabilen varyans miktarını gösterdiğinden, çoklu korelasyonda, $R$'den çok $R^2$ değeri ile ilgileniriz. Örneğimizde $R^2 = ,941^2 = ,886$ olduğundan, SAT puanlarındaki değişkenliğin %88,6'sının Eğitim Harcamaları ve LogPctSAT değişkeni ile açıklanabildiğini söyleyebiliriz. Eğitim Harcamaları değişkeninin tek başına SAT puanlarındaki değişkenliğin %14,5'ini açıklayabildiği göz önüne alındığında, bu varyans miktarının sadece Eğitim Harcamaları değişkeni ile açıklanabilen varyans miktarından yaklaşık %75 daha fazla (%88,6 - %14,5 = %74,1) olduğu görülmektedir.

Ne yazık ki $R^2$ evrendeki ilgili parametrelerin yansız bir kestirimi değildir $\left(R^{*2}_{.123\ldots p}\right)$. Bu yanlılığın büyüklüğü $N$ ve $p$'ye bağlı olarak değişir. $N = p + 1$ olduğunda, $Y$ ile $X_1$, $X_2$,… $X_p$ değişkenleri arasındaki ilişkinin evrendeki gerçek değerine bakılmaksızın, kestirme mükemmel ve $R = 1$ olur. (İki noktadan geçen doğru herhangi iki noktaya mükemmel bir uyum gösterecektir; taburenin her ayağı bir nokta olmak üzere üçayaklı bir tabure üzerindeki düzlem ile tabure ayakları mükemmel bir uyuma sahiptir vb.) $R^2$'nin nispeten yansız kestiricisi aşağıdaki eşitlikle verilmiştir:

$$\text{R}^{*2}\text{'nin kestiricisi} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Örnekteki veri kümesi için hesaplanacak olursa;

$$\text{R}^{*2}\text{'nin kestiricisi} = 1 - \frac{(1 - ,886)(49)}{47} = ,881$$

bulunur. Bu sonuç Gösterim 15.1'de yer alan SPSS çıktılarındaki düzeltilmiş $\text{R}^2$ değerine eşittir.

$R$'nin tanımından sadece 0 ile 1 arasında değerler alabildiği anlaşılmış olmalıdır. $R$'nin, $R^2$'nin pozitif karekökü olması ve $r_{Y\hat{Y}}$ korelasyonuna karşılık gelmesi, neden sadece pozitif değerler alabildiği sorusuna açıklık getirir, $Y$ ile $\hat{Y}$ arasında negatif bir korelasyon bulunması olasılığı son derece düşüktür. Bu oldukça önemli bir husustur. Çünkü SAT puanlarını sadece Eğitim Harcamaları değişkeni ile yordamak istediğimizde iki değişken arasındaki basit korelasyon katsayısının -,381 olduğu bilinmesine karşın çoklu korelasyon değeri hâlâ ,381 olmaktadır. Burada neler olup bittiğini anladığınız sürece herhangi bir karışıklık olmayacaktır.

$R^2$ değerine kıyasla, düzeltilmiş $R^2$ değeri, evrendeki katsayının karesinin daha az yanlı bir kestirimidir. Bundan dolayı, araştırmalarda daha çok düzeltmiş $R^2$ değerinin rapor edilmesi beklenebilir. Gerçekte, $R^2$ değeri bilgisayar çıktılarında nadiren rapor edilmektedir. Bu kitapta neden öyle olması gerektiğine cevap verilmese de; normalde $R$ veya $R^2$ değerinin rapor edildiğini bilmelisiniz.

## $R^2$ Değerinin Manidarlığının Test Edilmesi

Her bir yordayıcı değişkenin $Y$'nin yordanmasına manidar katkı sağlayıp sağlamadığının belirlenebilmesi için bu değişkenlere ait regresyon katsayısının ($b_j$) test edilmesi gerekir. Bundan dolayı, hangi yordayıcı değişkenlerin $Y$'nin manidar bir kestiricisi olduğu regresyon analizinde cevap aranan sorulardan biridir. Ancak, çoklu regresyon analizinde sorulması gereken ilk soru, "Yordayıcı değişkenlerin tümü bir arada alındığında, $Y$'yi manidar düzeyde kestirip kestirmediğidir?" Genel ilişki manidar değilse, her bir değişkenin manidarlığına bakılmasına gerek kalmayacağından, çoklu regresyon analizinde öncelikli olarak sorulması gereken soru budur.

$X_1$, $X_2$,…, $X_p$ ile $Y$ değişkenleri arasındaki genel ilişkiyi test etmenin en kolay yolu, çoklu korelasyon katsayısının manidarlığının test edilmesidir. Bu ifade; $R^*$ evrendeki korelasyon katsayısını temsil etmek üzere, $H_0$: $R^* = 0$ hipotezinin test edilmesi anlamına gelir. Testin doğası gereği, $R^2$'nin test edilmesi $R$'nin test edilmesine göre daha kolay olsa

da hangisi test edilirse edilsin ulaşılacak sonuç aynıdır. $R^2$'nin test edilmesi, Bölüm 9'da sadece yordayıcı değişkenin olduğu durumlarda $R^2$'nin manidarlığını test etmenin basit bir uzantısıdır. Çoklu regresyon durumunda $p$ yordayıcı için eşitlik aşağıdaki gibidir:

$$F = \frac{(N - p - 1)R^2}{p(1 - R^2)}$$

Örneğimizdeki veri için $N = 50$, $p = 2$ ve $R^2 = ,886$ olduğundan,

$$F = \frac{(50 - 2 - 1)(,886)}{2(,114)} = \frac{47(,886)}{,228} = 182,64^4$$

elde edilir. Bulunan $F$ değeri Gösterim 15.1'deki özet tabloda verilen $F$ değerine eşittir. Hesaplanan $F$ değeri (182,64) 2 ve 47 serbestlik derecesindeki $F$ değerinden büyük olduğundan, $H_0$: $R^* = 0$ hipotezini reddederiz. Dolayısıyla, Eğitim Harcamaları ve LogPctSAT değişkenlerinin SAT puanlarını manidar düzeyde yordadığını söyleyebiliriz. (Analiz çıktılarında bu, $F$ değeri için $p = ,000$ olarak rapor edilmiştir.)

### Örneklem Büyüklükleri

Düzeltilmiş $R$ ve $F$'ye ait yukarıdaki formülden hareketle söyleyebileceğiniz gibi, yapılan korelasyon kestirimi örneklem büyüklüğüne ($N$) ve yordayıcı sayısına ($p$) bağlıdır. Pek çok kişi, ölçüt değişken ile yordayıcılar arasında bir ilişki yoksa $R$'nin 0'a yaklaşması gerektiğini düşünür. Gerçekte, *tesadüfi bir veri* kümesi için beklenen $R$ değeri $p/(N\text{-}1)$'dir.

Buna göre, örneklem büyüklüğü 50 ve yordayıcı değişken sayısı 2 olduğunda ve ölçüt değişken ile yordayıcı değişkenler arasında gerçek bir ilişki bulunmadığında, beklenen değer 0 değil; $R = ,04$ olacaktır. Bu nedenle, nispeten büyük bir örnekleme sahip olunması önemlidir. Her bir yordayıcı değişken için en az 50 katılımcının örnekleme dâhil edilmesi, bu konuda genel bir kural olarak uzun yıllardan beri dikkate alınmaktadır. Bununla birlikte, Harris (1985) bu kuralı destekleyen ampirik bir araştırma bulunmadığına dikkat çekmektedir. Bu kural özellikle uç durumlarda yanıltıcıdır. Çünkü tek bir yordayıcı değişken için 10 katılımcının yeterli olacağı hiç kimse tarafından savunulamaz. Harris $N$ ve $p$ oranı yerine, $N$ ile $p$ arasındaki farka dayanan bir kural ileri sürmüştür. Harris'e göre, $N$ değeri $p$'nin en az 50 fazlasına eşit olmalıdır. Bu kuralı biraz yumuşatarak örneklemdeki katılımcı sayısının yordayıcı değişkenin en az 40 fazlası ($N \geq p + 40$) olması gerektiğini ileri süren araştırmacılar da olmuştur. Söz konusu iki kural doğrudan korelasyon katsayının güvenirliği ile ilişkili iken; Cohen, Cohen, West ve Aiken (2003) bu soruna istatistiksel güç açısından yaklaşmışlardır. Cohen, Cohen, West ve Aiken'a (2003) göre, tek bir yordayıcı değişken olduğunda ,30 olan bir evren korelasyonunda ,80 olan bir istatistiksel güce sahip olabilmek için $N = 124$ olmalıdır. Aynı istatiksel güçte ve yine ,30 olan bir evren korelasyonunda, beş yordayıcı değişken mevcutsa $N = 187$ olması gerekir. Görülebileceği gibi, mâkul bir istatistiksel güce ulaşılabilmesi için büyük örneklemlere ihtiyaç duyulmaktadır. Belki de bu konudaki en iyi kural Darlington'ın (1990) kuralıdır.

## 15.8 Kısmi ve Yarı Kısmi Korelasyon Katsayıları

Kısmi ve yarı kısmi korelasyon katsayıları birbiri ile ilişkili iki korelasyon katsayısı olup her ikisi de, ayrıştırılmış bir veya daha fazla değişkeninin etkisinin kontrol altına alınmasını içerir.

---

4 Diğer yerlerde olduğu gibi, hesaplamış olduğumuz değer, yuvarlama hatasından dolayı burada hesaplanan değere göre farklılıklar gösterebilir. Bilgisayar programları hesaplamalar sırasında, virgülden sonra rapor ettiği basamağın çok daha fazlasını dikkate almaktadır. Burada sunulan eşitlik için virgülden sonra 3 basamak kullanılması, analiz sonucuda 27,118 değerinin rapor edilmesiyle sonuçlanmıştır. Gerçek cevabın 27,184 olduğu dikkate alındığında, hesaplanan değerin bilgisayar programlarında rapor edilen analiz çıktıları ile örtüştüğü söylenebilir.

## Kısmî Korelasyon

**$r_{01.2}$ kısmi korelasyonu**

**$r_{01.2}$ kısmi korelasyon**unun, hem *X* hem de *Y* değişkeni bir veya daha çok değişken açısından kontrol altında alındıktan sonra, bu iki değişken arasındaki korelasyon olduğunu görmüştük. Daha açık bir deyişle kısmi korelasyon, bir veya daha fazla başka değişken aracılığıyla orijinal değişkenin kestiriminden oluşan iki artık kümesi arasındaki korelasyondur.

Üniversitedeki başarı ile gelir arasındaki ilişkiyi belirlemek isteyen bir araştırmacı düşünelim. İki değişken arasındaki ilişkiyi belirlemek için, başarı ve gelir değişkenlerine ait ölçümler elde edildikten sonra korelasyon analizi gerçekleştirilir. Korelasyon analizi sonucunda iki değişken arasındaki ilişkinin manidar bulunduğunu kabul edelim. Araştırmacı bu sonuçlardan hareketle, üniversitedeki öğrencilerin, eğer üniversitede başarılı olamazlarsa iş hayatında yüksek gelir elde edemeyeceklerini söyler. Bu sırada sınıfın arka sıralarındaki zekî bir öğrenci hem üniversitedeki başarının hem de gelir düzeyinin zekâ seviyesi ile ilişkili değişkenler olabileceğini fark eder. Bu öğrenci, zekî insanların hem üniversitede daha başarılı olduğunu hem de ileride iyi bir gelire sahip olacaklarını söyleyerek, üniversitedeki başarı ile gelir arasındaki ilişkinin yanıltıcı olabileceğini savunur.

Bu tartışmayı çözmenin en basit yolu, üniversitedeki başarı ve gelir değişkenlerinin her ikisine de zekâ düzeyine ait düzeltme uygulandıktan sonra, iki değişken arasındaki kısmi korelasyon katsayısının hesaplanmasıdır. Buna göre, gelir değişkeni için zekâ düzeyi kontrol altına alınarak artık değerler elde edilir. Bu artık değerler, gelir değişkeninde zekâ düzeyi ile açıklanamayan varyansı temsil eder. Burada elde edilen değerler, gelir değişkeninin zekâ seviyesinden arındırılarak saflaştırılmış hali şeklinde düşünülebilir. Daha sonra bu işlem, üniversitedeki başarı değişkeni için tekrarlanır. Yani, üniversitedeki başarı değişkeninde de zekâ düzeyinin etkisi kontrol altına alınarak artık değerler elde edilir. Bu artık değerler, üniversitedeki başarının zekâ düzeyi ile açıklanamayan (bağımsız olan) kısmına karşılık gelir. Bu işlemlerin ardından; "Gelir düzeyinde zekâ seviyesi ile açıklanamayan değişkenlikler, üniversitedeki başarının zekâ düzeyinden bağımsız olan değişkenlikleri aracılığıyla yordanabilir mi?" şeklindeki temel araştırma problemi cevaplandırılabilir. Bu araştırma problemini cevaplamak için hesaplanan korelasyon katsayısı zekâ seviyesi (IQ) kontrol altına alındıktan sonra, üniversitedeki başarı ile gelir arasındaki ilişkiyi gösterir.

Bu kısmi korelasyon katsayısı, $r_{01.23...p}$ ile temsil edilir. Bu gösterimde, noktanın solundaki iki indis aralarındaki korelasyona bakılacak değişkenleri gösterir. Noktanın sağ tarafındaki indisler de etkisi kontrol altında tutulacak değişkenleri temsil eder.

## Yarı Kısmî Korelasyon

**yarı kısmi korelasyon $r_{0(1.2)}$**

Hem bu bölümde hem de Bölüm 16'da son derece yararlı olacak bir korelasyon türü de $r_{\mathbf{0(1.2)}}$ şeklinde gösterilen **yarı kısmi korelasyon**dur.[5] Adından da anlaşılacağı gibi, yarı kısmi korelasyon, ölçüt değişken ile ayrıştırılmış bir yordayıcı arasındaki korelasyondur. Diğer bir deyişle, kısmî korelasyonda ($r_{01.2}$) ölçüt ve yordayıcı 1 değişken ayrıştırılırken, yarı kısmi korelasyonda $r_{0(1.2)}$ sadece yordayıcı değişkene düzeltme uygulanır, ölçüt değişkene herhangi bir kontrol altına alma işlemi uygulanmaz. Bu durumda yarı kısmi korelasyon katsayısı, *Y* ile $X_1$'in üzerine kestirilen kısmından kalanıdır ($X_1 - \hat{X}_1 = X_{1\,\text{kalan}}$). Aslında *Y* değişkeni ile $X_1$'in $X_2$'den bağımsız olan parçasıdır. Yarı-kısmî korelasyona iki model açısından farklı bir bakış, birinin diğerinden daha az sayıda yordayıcı ihtiva etmesidir.

$$r^2_{0(1,2)} = R^2_{0,12} - r^2_{0,2}$$

Örneğimizdeki Eğitim Harcamaları değişkenine ait SAT puanları ve LogPctSAT değişkenine göre düzeltilmiş Eğitim Harcamaları arasındaki yarı kısmî korelasyon hesaplanabilir. Bu veri kümesi için;

$$R^2_{\text{SAT.eğitim harcamaları.LogPctSAT}} = ,886 \text{ ve } R^2_{\text{SAT.LogPctSAT}} = ,857$$

olur, böylece

---

[5] Yarı kısmî korelasyon teriminin kullanıldığını gördüğüm tek ders kitabı, lisansüstü eğitimimi bitirdikten hemen sonra gördüğüm McNemar (1969) idi. O zaman bu terimi, terim bulmakta zorlanan bir kimsenin yazdığını düşündüm, Bu terimin SPSS'te de kullanıldığını göreceksiniz.

$$r^2_{\text{SAT.eğitim harcamaları.LogPctSAT}} = r^2_{\text{SAT(Eğitim Harcamaları.LogPctSAT)}} - r^2_{\text{SAT.LogPctSAT}}$$
$$= ,886 - ,857 = ,029$$
$$r_{\text{SAT(Eğitim Harcamaları.LogPctSAT)}} = \sqrt{,029} = ,170$$

elde edilir. $r_{0(1.2)}$ için yukarıdaki formül, çoklu regresyon eşitliği ve korelasyonun neyi temsil ettiğini açıklamak için bir fırsat tanır. Bu formülün yeniden düzenlenmesiyle,

$$R^2_{0.12} = r^2_{02} + r^2_{0(1.2)}$$

eşitliği elde edilir. Bu formüle göre çoklu korelasyonun karesi; ölçüt değişken ile yordayıcı değişkenlerden biri arasındaki korelasyonun karesiyle, ölçüt değişken ve diğer yordayıcı değişkenin ilk değişkenden bağımsız olan kısmı arasındaki korelasyonun karesi toplamına eşittir. Dolayısıyla, $R$ veya $R^2$ yi, değişkenlerinden birinin sağladığı olabildiğince *çok bilgi* ile ikinci ve diğer değişkenlerden sağlanan *gerekli bilginin toplamı* olarak düşünebiliriz. Genel olarak;

$$R^2_{0.123\ldots p} = r^2_{01} + r^2_{0(2.1)} + r^2_{0(3.12)} + \cdots + r^2_{0(p.123\ldots p-1)}$$

olur. Bu formülde $r^2_{0(3.12)}$, yordanan değişken ile 3. yordayıcı değişkenin 1. ve 2. yordayıcı değişkenlerden bağımsız kısmı arasındaki korelasyonun karesidir. Çoklu regresyona böyle bir bakış her bir yordayıcı değişkenin, ölçüt değişkenin kestirilmesindeki rolünü ve Bölüm 16'da belirtilen varyans analizine en küçük kareler yöntemi ile yaklaşma yönünden yararlı olur. Yordayıcı değişkenler birbirinden bağımsız olduğunda, bu formül aşağıdaki şekli alır. Çünkü yordayıcı değişkenlerin birbirinden bağımsız olması, kontrol altına alınması gereken ortak bir varyansın bulunmadığı anlamına gelir. Bu durumda önceki formül aşağıdaki hâle gelir.

$$R^2_{0.123\ldots p} = r^2_{01} + r^2_{02} + r^2_{03} + \cdots + r^2_{0p}$$

LogPctSAT değişkenine göre düzeltilmiş SAT puanları ve yine LogPctSAT değişkenine göre düzeltilmiş Eğitim Harcamaları arasındaki kısmi korelasyonun karesi, bundan sonra açıklanan yöntem yardımıyla ,198 olarak hesaplanmıştır. Bu sonuç, SAT puanlarının *LogPctSAT değişkeni ile açıklanamayan kısmına* ait değişkenliğin %20'sinin Eğitim Harcamaları değişkenin *LogPctSAT değişkeni yardımıyla açıklanamayan kısmıyla* kestirilebildiğini gösterir. Daha açık bir anlatımla, SAT puanlarının LogPctSAT değişkeninden bağımsız kısmına ilişkin varyansın %20'si, Eğitim Harcamaları değişkeninin LogPctSAT değişkeninden bağımsız kısmı ile açıklanabilmektedir. Bu nokta bir sonraki bölümde ele alınacaktır.

Kısmî ve yarı kısmî korelasyon katsayılarının test edilmesine dair ayrı bir manidarlık testine ihtiyacımız yoktur, çünkü böyle bir test için regresyon katsayılarının test edilmesiyle ilgili olanlar içinde yeteri kadar vardır.

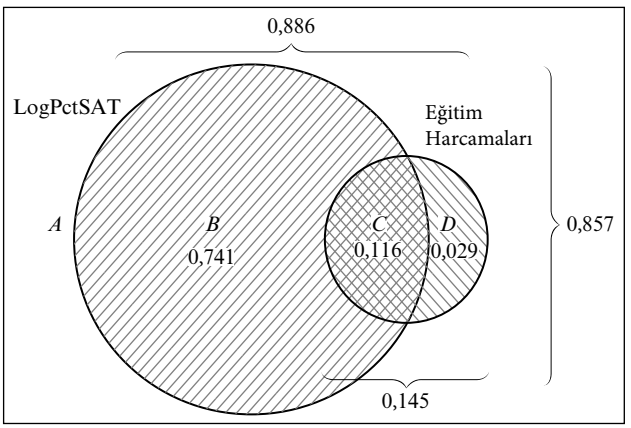
Eğer test manidar ise, ilgili $\beta$ değeri ile kısmî ve yarı kısmî korelasyon katsayıları da manidardır.[6] Bu nedenle, Gösterim 15.1'den Eğitim Harcamaları değişkeni için bu katsayıların tümünün manidar olduğu bilinmektedir. Bununla birlikte, bir katsayının manidarlığından söz edilirken modeldeki diğer değişkenler bağlamında konuşulduğu unutulmamalıdır. Örneğin, Öğretmen Maaşları değişkeni SAT puanlarının yordanmasına dair regresyon modeline dâhil edildiğinde, modele manidar bir katkı sağlamadığı görülmüştü. Bu şekildeki bir sonuç Öğretmen Maaşlarının SAT puanlarının kestirilmesinde herhangi bir rolü olmadığını gösterir. (Gerçekte, SAT puanlarının yordanmasına dair regresyon denklemindeki tek yordayıcı değişken öğretmen maaşı olduğunda, öğretmen maaşları SAT puanlarını şans ile gerçekleşmesi beklenenden daha yüksek bir düzeyde kestirebilmektedir; $R = -,440$, $p < ,001$.) Böyle bir sonuç; Öğretmen Maaşı değişkeninin, regresyon denklemine kendisinden önce dâhil olan diğer değişkenlerin sağladığı bilginin ötesinde herhangi bir bilgi vermediği anlamına gelir.

## Kısmi ve Yarı Kısmi Korelasyonun Alternatif Bir Yorumu

**Venn diyagramı**

Kısmi ve yarı kısmi korelasyonun anlamanın oldukça öğretici olabilecek alternatif bir yolu bulunmaktadır. Bu yöntemde kısmî ve yarı kısmî korelasyonu açıklamak için **Venn diyagramı**ndan yararlanılmaktadır. Şekil 15.4'te gösterilen Venn Diyagramı ölçekli bir şekilde

6 Dikkat edilirse hem kısmi hem de yarı kısmi korelasyon katsayısından söz edilmiştir. Ancak, regresyon analizi için sadece kısmi regresyon katsayısından ($b_j$) bahsedilmiştir. Bu katsayı aynı zamanda, yarı kısmi regresyon katsayısı olarak da adlandırılabilir.




**Şekil 15.4** Kısmi ve yarı kısmi korelasyonu gösteren venn diyagramı

çizilmemiştir. Bu diyagram SAT ve iki yordayıcısı arasındaki ilişkinin çeşitli yönlerini göstermektedir.

Şekil 15.4'teki kutunun SAT puanlarındaki değişkenliğin tamamını temsil ettiğini varsayalım. Bu alanı 1,00 olacak şekilde ayarladım (SAT'a ait açıklanabilecek varyans oranı). LogPctSAT olarak adlandırılan daire, SAT puanlarındaki değişkenliğin LogPctSAT ile açıklanan kısmını gösterir. Diğer bir deyişle, dairenin alanı $r^2_{01} = ,857$'dir. Benzer şekilde, Eğitim Harcamaları ile belirtilen dairenin alanı SAT puanlarındaki değişkenliğin Eğitim Harcamalarıyla açıklanan kısmını belirtir, bu alan $r^2_{02} = ,145$'e eşittir. Son olarak, bu iki dairenin kesişimi, SAT puanlarındaki değişkenliğin LogPctSAT ve Eğitim Harcamalarına ait ortak kısmını gösterir ve ,116'ya eşittir. Her iki dairenin dışındaki fakat kutunun içindeki alan SAT puanlarındaki bu iki değişken yardımıyla açıklanan varyans olup değeri = ,059'dur.

Şekil 15.4'teki kutunun içindeki *B*, *C* ve *D* alanları LogPctSAT *veya* Eğitim Harcamaları değişkenlerinin SAT puanlarındaki açıklayabildiği varyans miktarını temsil eder. (*A* ile gösterilen kısım, her iki değişkenin veya bu değişkenlerin birleşiminin SAT puanlarındaki değişkenliğin açıklayamadığı parçasını yansıtan artık varyans miktarına karşı gelir.) Buna göre, örnekte yer alan iki yordayıcı değişken, SAT puanlarındaki (*Y*) değişkenliğin %88,6'sını açıklayabilmektedir. $B + C + D = ,741 + ,116 + ,029 = ,886$. SAT puanları ile Eğitim Harcamaları değişkenine göre düzeltilmiş LogPctSAT arasındaki *yarı kısmi korelasyonun karesi*, LogPctSAT, SAT ile LogPctSAT'ın Eğitim Harcamalarından bağımsız kısmının (farkının) SAT puanlarında açıklayabildiği varyans miktarıdır. Bu varyans miktarı, Şekil 15.4'te *B* ile gösterilen ve değeri ,741 olan alandır.

$$r^2_{0(1.2)} = R^2_{0.12} - r^2_{01} = ,886 - ,145 = ,741$$

Yarı kısmi korelasyon bu alanın kareköküdür.

$$r^2_{0(1.2)} = \sqrt{,741} = ,861$$

*Kısmi korelasyon karesinin* de benzer bir yorumu vardır. SAT Puanlarında LocPctSAT değişkeninin açıkladığı, fakat Eğitim Harcamaları değişkeninin açıklayamadığı ek yüzde (yarı kısmi korelasyonun karesidir) yerine, Eğitim Harcamaları değişkeninin açıklayabildiği varyanstan sonra LogPctSAT değişkenin açıklayabildiği ek varyans miktarıdır. Örneğin; $r^2_{02} = ,381^2 = ,145$ ve $1 - r^2_{02} = ,855$

$$r^2_{0(1.2)} = \frac{r^2_{0.12}}{1 - r^2_{02}}$$

$$= \frac{,741}{,855} = ,866$$

$$r^2_{0.12} = \sqrt{,866}$$

$$= ,931.$$

Çoklu korelasyonun karesi, kısmi ve yarı kısmi korelasyon aşağıdaki gibi temsil edilebilir.

$r^2_{0(1.2)} = B$ = yarı kısmî korelasyonun karesi

$r^2_{01.2} = \dfrac{B}{A + B} = \dfrac{r^2_{0(1.2)}}{1 - r^2_{02}}$ = kısmî korelasyonun karesi

Ek olarak;

$A = 1 - R^2_{0.12} = Y$'deki (SAT) artık (açıklanamayan) varyans miktarı

$D = r^2_{0(1.2)}$ = diğer yarı kısmi korelasyon

$B + C + D = R^2_{0.12}$ = çoklu korelasyonun karesi

$B + C = r^2_{01} = Y$ (SAT) ve $X_1$ (Eğitim Harcamaları) değişkenleri arasındaki korelasyonun karesi

$C + D = r^2_{02} = Y$ (SAT) ve $X_2$ (LogPctSAT) değişkenleri arasındaki korelasyonun karesi

### Kısmı ve Yarı Kısmi Korelasyon ile Neden İlgileniyoruz?

Neden kısmî ve yarı kısmî korelasyon ile ilgilendiğimiz merak konusu olabilir? Bu bağlamda, kısmî ve yarı kısmî korelasyonun, bildiklerimizin ötesinde ne gibi getirileri olduğu sorusunun cevaplanması gerekir. Kısmî ve yarı kısmî korelasyonu incelemenin çok önemli getirileri vardır. Kısmî ve yarı kısmî korelasyon, sıkıntı yaratabileceği düşünülen değişkenlerin kontrol altına alınabilmesine imkân tanımakta ve bu sayede "*B değişkeninin etkisi kontrol altına alındıktan* sonra *A* ile *Y* değişkenleri arasında ,65'lik bir korelasyon vardır" biçiminde açıklamalar yapabilmemize imkân tanır. Bölümün ilerleyen kısımlarda sunulan bir araştırma, bu konuya ilişkin bir örnek olabilir. Leerkes ve Crockenberg (1999) bir kadının çocukluk döneminde algıladığı anne bakımı ile kendi annelik becerisine ilişkin özgüven veya öz-yeterlilik algısı arasındaki ilişkinin belirlenmesiyle ilgilenmektedirler. Leerkes ve Crockenberg çocukluk döneminde anne bakımına ilişkin algıları yüksek olan kadınların benlik algılarının da yüksek olmasının bu ilişkiyi etkileyip etkilemeyeceğini merak etmektedir. Diğer bir ifadeyle, benlik saygılarındaki farklılıklar kontrol altına alındığında, çocukluk döneminde algılanan anne bakımı ile annelik becerisine ilişkin öz yeterlilik algısı arasında manidar bir ilişki olup olmayacağı problemine cevap bulmaya çalışmaktadırlar. Bu soruyu cevaplayabilmek için benlik saygısının çocukluk döneminde algılanan anne bakımı ile annelik becerisine ilişkin öz-yeterlilik algısı üzerindeki muhtemel etkilerinin kontrol altına alınmasını gerektirir. Kısmî ve yarı kısmî korelasyonlar, bu tür ilişkilerin fark edilebilmesi ve karmaşık ilişkilerin kontrol altına alınabilmesi açısından oldukça etkili bir yöntemdir.

## 15.9 Bastırıcı Değişkenler

Regresyon analizi kapsamında cevaplanması amaçlanan araştırma probleminde, tüm değişkenlerin yordayıcı değişken ile pozitif ilişki içinde olduğunu varsayalım (*C* sabit olmak üzere; eğer *X* ile *Y* arasında negatif ilişki varsa, *C* – *X* ile *Y* arasında pozitif ilişki olacaktır). Böyle bir durumda, tüm regresyon katsayılarının ($\beta_i$ ya da $b_i$) pozitif olması beklenir. Bununla birlikte, bazen bir regresyon katsayısı *manidar düzeyde* negatif olabilir. Böyle bir değişken, eğer manidar ise **bastırıcı değişken** olarak adlandırılır.[7]

**bastırıcı değişken**

Bastırıcı değişken ilk bakışta mantıksız görünebilir. Tanımı gereği, yordayıcı değişken ile ölçüt arasındaki basit korelasyonun pozitif olduğu bilinmektedir. Buna rağmen, regresyon eşitliğinde bu değişkendeki artış $\hat{Y}$ değişkeninde azalmaya neden olmaktadır. Ayrıca, $R^2 = \Sigma\beta_i r_{oi}$ olduğu gösterilebilir. Eğer $r_{oi}$ pozitif ve $\beta_i$ negatifse $\beta_i r_{oi}$ negatif olacaktır. Buna göre, $\beta_i$'ye negatif bir değer verilmesiyle, hatayı minimum düzeyde tutma görevi olan regresyon çözümünün $R^2$'yi düşüreceği görülür. Ancak, önceden bilinenler ile çelişen bu durumun mantıklı bir açıklaması olmalıdır.

---

7 Cohen ve Cohen (1983) iki bastırıcı türünü tartışmışlar ve özellikle araştırmacının öngörülere aykırı sonuçlar ile karşılaşması durumunda yardımcı olabilecek açıklamalara yer vermiştir. Cohen, Cohen, West ve Aiken'ın (2003) kitabının yeni basımında bu tartışmalar bulunmamaktadır. Dolayısıyla, bu konuda bilgi edinmek için ilgili kitabın daha eski basımlarına başvurulmalıdır.

Çok fazla yer kaplayacağı düşüncesiyle bastırıcı değişkenler hakkında kapsamlı tartışmalara girilememektedir, fakat sezgisel olarak mantıklı bir açıklama getirilmesi önemlidir. Bastırıcı değişken hakkında daha ayrıntılı bilgi edinmek için Cohen ve Cohen (1983) ile Darlington (1968) kaynaklarına başvurulabilir. (Özellikle, Cohen ve Cohen'de (1983) yer alan açıklamalar bu konuda yararlı olabilir.) Burada, Darlington'ın (1990) çalışmasından bir örnek sunacağız. Hıza dayalı bir tarih sınavının (kısa süre içerisinde tamamlanması gereken uzun bir sınav) tarih bilgisini ölçmek için kullanıldığını varsayalım. Tarih bilgisi muhtemelen okuma hızından tamamen bağımsız olmasına karşın, söz konusu hız testindeki performans okuma hızından bağımsız olmayacaktır. Bundan dolayı, test puanlarındaki varyansın bir kısmı öğrencilerin bilgisindeki farklılıktan çok, okuma hızlarındaki farklılığı yansıtacaktır. Bu durumda, okuma hızları yüksek olan öğrencileri, *sadece* bu özelliklerinden dolayı cezalandırmış, okuma hızı düşük olanlara da yine *sadece* bu özellikleri nedeniyle yardım etmiş olacağız. Okuma hızı bastırıcı değişken olduğunda, yapılan işlem tam olarak böyle bir sonuca neden olacaktır. Okuma hızı, sınav puanlarındaki hatanın bir kısmını bastırmaktadır.

Darlington, bir değişkenin $Y_r$ ile korelasyonu, $Y$ ile olan korelasyonundan daha yüksek olduğunda, bu değişkenin bastırıcı değişken olarak görev yaptığını belirtmektedir. (Burada, $Y_r$ tarih bilgisinin tarih puanlarından yordanması durumuna ait artık puanları temsil etmektedir.) Tersine, bir değişkenin $Y$ ile korelasyonu $Y_r$ ile olandan daha yüksek olduğunda, bu değişken bastırıcı değişken olarak görev yapmamaktadır. Cohen, Cohen, West ve Aiken (2003), psikolojide bastırıcı ilişkilerin bulunmasının zor olduğunu (en azından istatiksel olarak manidar olanın), biyoloji ve ekonomi bilimlerinde ise bastırıcı ilişkilerin kolaylıkla bulunabildiğini belirtmişlerdir. Bu alanlarda bastırıcı ilişkiler ile sıklıkla karşılaşılmasının nedeni, var olan dengenin korunması mekanizması ile ilgilidir. Mevcut dengenin korunması sürecinde, $X$ değişkenindeki artış $Y$ değişkeninde de artışa neden olur, bu da $Z$ değişkeninde artışa sebebiyet verir. $Z$ değişkenindeki artış ise, $Y$ değişkenindeki azalmayı geri götürür. Bununla birlikte bu mekanizma psikoloji biliminde çok yaygın değildir. Bastırıcı etki ile ilgili sorular oldukça sık sorulmakla birlikte, sorulan soruların büyük bir kısmında bastırıcı etki istatistiksel olarak manidar çıkmamaktadır.

## 15.10 Regresyona Hazırlık

Eğitim harcamaları ve SAT'a katılan öğrenci yüzdesinin logaritmik dönüşümü değişkenlerinden SAT puanlarının yordanmasında önemli bir adımı atladık. Çünkü ilk olarak regresyon analizinin altında yatan bazı kavramların ortaya konulması gerekirdi. Şimdi geriye dönüp bu atladığımız boşluğu doldurabiliriz. Tüm gözlemler ve yordayıcılar modele dâhil edilmeden ve bilgisayar programına regresyon analizini gerçekleştirmesi talimatı verilmeden önce, veri kümesinin dikkatli bir biçimde incelenmesi gerekir. Bu işlem, neredeyse çoklu regresyon analizinin uygulanabildiği bilgisayar programlarının tümünde gerçekleştirilir. Cohen ve arkd. (2003) ile Fox (2002) regresyonun tanılarına ilişkin çok daha kapsamlı ve okunabilir kaynaklardır.

Veril kümesinin incelenmesinde ilk adım, önemli değişkenlerin görsel sunumunu içeren Şekil 15.1'de gerçekleştirilmiştir. Bu şekil incelendiğinde, değişkenlerin çoğunun oldukça dağınık ve SAT'a giren öğrenci yüzdesinin açık bir biçimde iki modlu olduğu görülmektedir. SAT puanlarında da kısmen iki modlu bir durum söz konusudur. SAT puanlarının bu iki modlu yapısı büyük ihtimalle SAT'a giren öğrenci yüzdesinin iki modlu olmasından kaynaklanmaktadır. Kısa süre içinde açığa çıkacak nedenlerden dolayı, SAT'a giren öğrenci yüzdesi (PctSAT) yerine; SAT'a giren öğrenci yüzdesinin doğal logaritması (LogPctSAT) kullanılmıştır. Bu tür bir dönüştürmenin en azından SAT puanları ile her bir eyalette SAT'a katılan öğrenci yüzdesi arasındaki eğrisel ilişkiyi azaltma etkisi vardır. Özellikle SAT'a katılan öğrenci yüzdesi değişkenine logaritmik dönüşüm uygulandıktan sonra, değişkenlerin hiçbirinde uç (extreme outliers) değerler bulunmamaktadır.

**çok değişkenli uç değerler**

Değişkenlere tek tek bakıldığında, fazla uç değerin olmaması, tüm değişkenlerin regresyon analizine dahil edildiği ve hepsinin iyi olduğu anlamına gelmez. Her bir değişken tek tek incelendiğinde uç değerler bulunmasa da, veri kümesinde **çok değişkenli uç değerler** bulunma olasılığı vardır. Her bir değişken ayrı ayrı ele alındığında, bu değişkenler mâkul puanlara sahip olabilir. Ancak, iki veya daha çok değişkenin beklenmedik bir kombinasyonu olabilir. Örneğin; 6 feet boyunda olmak veya 125 pound ağırlığında olmak sıra

dışı bir durum değildir. Fakat, 6 feet boyunda olan kimsenin 125 pound ağırlığında olması sıra dışı olabilir.

Şimdilik, veri kümesinin kabul edilebilir aralığın dışında kalan herhangi bir nokta içermemesi ve dağılımda önemli bir çarpıklık bulunmaması memnuniyet vericidir. Burada yapıldığı gibi, tüm değişkenler kullanılarak ön bir regresyon analizinin uygulanması yararlı bir ikinci adım olacaktır. Burada uygulanan regresyon analizinin amacı ölçüt hakkında tahminde bulunmaktan çok veriyi incelemek olduğundan, yapılan regresyon analizi için "hazırlık" ifadesi kullanılmıştır.

Regresyon hazırlık çalışmasına ilişkin araçların kullanılmasını açıklarken, doğrudan Eğitim Harcamaları çalışmasına ait veri kümesi üzerinden ilerlemek yerine, bu araçların kullanımını göstermek için hazırlanan küçük bir veri kümesinden yararlanacağız. Bu veri kümesi aşağıda ve buna ait grafik Şekil 15.5'te verilmiştir.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **X:** | 1 | 1 | 3 | 3 | 3 | 4 | 5 | 5 | 7 | 6 | 10 | 13 |
| **Y:** | 1 | 2 | 3 | 5 | 7 | 6 | 8 | 10 | 10 | 5 | 4 | 14 |

Üç temel hazırlık istatistiği aşağıda verilmiş ve her biri Şekil 15.5'te gösterilmiştir.

**uzaklık**

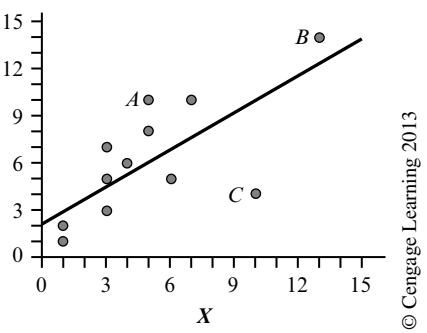
**baskı**

**etki**

1. **Uzaklık (Distance),** bağımlı değişkendeki ($Y$) potansiyel uç değerlerin belirlenebilmesi için oldukça yararlıdır.
2. **Baskı (Leverage), $h_i$,** bağımsız değişkenlerdeki ($X_1, X_2, \ldots, X_p$) potansiyel uç değerlerin belirlenebilmesi için oldukça yararlıdır.
3. **Etki (Influence)**, beklenmedik şekilde etkili gözlemleri belirlemek için bağımlı ve bağımsız değişkenlerdeki uzaklık noktalarını birleştirir. Eğer bir gözlemin varlığı veya yokluğuna bağlı olarak regresyon yüzeyinin konumu belirgin şekilde değişiyorsa, bu gözlemin etkili olduğu sonucuna varılır.

Bağımlı değişkendeki uzaklığın belirlenmesinde kullanılan en yaygın ölçüm artıklardır ($Y_i - \hat{Y}_i$). Bu artıklar, regresyon doğrusu ile herhangi bir nokta arasındaki dikey uzaklığı ifade eder. Şekil 15.5'te gösterilen $A$ ve $C$ noktaları yüksek artık değerlere sahiptir (regresyon çizgisinden uzak noktalar). Tesadüfi hataları temsil eden bu tür noktalar yanlış kaydedilmiş olabilir veya gerçekte bu veri kümesine ait olmayan beklenmedik durumları yansıtabilir. (Tepki süresinin görevle ilgili zihinsel süreçlerin bir fonksiyonu olarak kestirilmesine yönelik bir çalışmada, yaşadığı sinirsel rahatsızlık nedeniyle tepki süresi oldukça yavaş olan bir katılımcının bulunması bu duruma bir örnek olabilir.) Artıklar tüm regresyon analizlerinin standart bir özelliği olup analiz gerçekleştirilirken rutin olarak talep edilmeli ve incelenmelidir.

Baskı (genellikle $h_i$ ile gösterilir), $X_j$ yordayıcı değişkenine ait alışılmamış değerleri gösterir. Tek bir yordayıcı değişken olduğunda, bu değer basit bir biçimde, söz konusu yordayıcıya ait bir puanın bu yordayıcı değişkenin ortalamasından sapmasının bir fonksiyonudur. Şekil 15.5'teki $B$ noktası yüksek baskı noktasına bir örnektedir. Çünkü $X$ yordayıcı değişkeni bu nokta için (13) $\bar{X}$'dan çok uzaktadır. Çoklu regresyonun uygulandığı birçok program, analiz çıktılarında rapor edilmesi için gerekli komut verildiğinde, her bir ölçüm için baskı değerlerini hesaplamakta ve rapor etmektedir. Baskı istatistiğine ilişkin muhtemel değerler; $p$ yordayıcı değişken sayısı olmak üzere, $(p + 1)/N$ ortalamasıyla $1/N$



© Cengage Learning 2013

**Şekil 15.5** Y'nin X'e göre dağılımının grafiği

alt değeri ve 1,0 üst değeri arasında uzanan bir ranja sahiptir. Stevens (1992), özellikle, 3(p + 1)/*n* değerini aşan baskı (leverage) değerlerine yakından bakılmasını önermektedir.

Bir noktanın yüksek uzaklık ve $h_i$ değerlerine sahip olması regresyon üzerinde mutlaka yüksek bir etkisinin olacağı anlamına gelmez. Ancak bu noktaların regresyon analizi üzerinde yüksek bir etkiye sahip olma potansiyeli vardır. Bir noktanın yüksek etkisinin olabilmesi için hem yüksek uzaklık hem de yüksek baskı değerine sahip olması gerekir. Şekil 15.5'te, *B* yüksek sapma değerine ancak kısmen küçük bir uzaklık değerine sahiptir. Diğer taraftan *A* noktasının uzaklık değeri yüksek, sapma değeri ise düşüktür; çünkü *A* noktası *X*'in ortalamasına yakın bir yerdedir. *C*'nin ise hem artık hem de sapma değeri yüksektir. Bu durum, *C*'nin regresyon üzerinde önemli bir etkisinin olduğunu gösterir. Bu etkiye ilişkin en yaygın ölçüm **Cook Uzaklığı'dır (Cook's Distance).** *Cook's D* olarak gösterilen bu ölçüm, *i'yinci* gözlemin veri kümesinden çıkarılarak analizin tekrarlanması durumunda $b_j$ de oluşacak *değişikliğin* kareler toplamının bir fonksiyonudur.

**Cook Uzaklığı (Cook's Distance)**

Gösterim 15.2, Şekil 15.5'teki veriye ait çeşitli hazırlık istatistiklerini içermektedir. Bu istatistikler SAS programı yardımıyla elde edilmiştir. Ancak benzer istatistikler, istatistikte kullanılan diğer programların neredeyse tümü tarafından da rapor edilmektedir.

Hazırlık istatistiklerini teşhis etmek için ilk olarak uzaklık ölçümlerinin sunulduğu ve artık (Resid) olarak adlandırılan sütuna bakılmalıdır. Bu sütun Şekil 15.5'te görülen bilgiyi yansıtmaktadır. Bu sütundaki değerlere bakıldığında 8. ve 11. gözleme ait artıkların büyük olduğu görülmektedir. *Y*'nin değerlerinin sadece 1 ile 14 arasında değiştiği dikkate alındığında, -5,89'luk bir artık değerin önemli olduğu söylenebilir.

Eğer veri kümesi çoklu regresyonun altında yatan varsayımları karşılarsa, *Y*'nin değerinin regresyon doğrusu etrafında normal dağılıma sahip olması beklenir. Diğer bir deyişle, çok büyük veri kümelerinde *X*'in belirli bir değerine karşılık gelen tüm *Y* değerleri normal dağılma sahip olacaktır. Bu değerlerin %5'i, 1,96 düzeltilmiş standart hatadan ötede, regresyon doğrusunun dışında kalacaktır. (Burada "düzeltilmiş" kelimesini kullandım, çünkü standart hatanın büyüklüğü, kısmen *X* değerlerinin *X*'in ortalamasından sapma derecesine bağlı olacaktır. Bu sapmanın derecesi baskı değerleriyle-$h_i$ ölçülmektedir.) Bu bağlamda, bir noktanın regresyon eğrisinden uzakta olup olmadığının sorulması anlamlı olacaktır. Eğer durum böyle ise, bu konuda endişelenmeliyiz. Artıkların büyüklüğüne ilişkin *t* testi, **Studentized artık (Residual)** olarak adlandırılan ve RStudent ile gösterilen bir istatistik ile verilmektedir. Bu test, $(N - p - 1)$ serbestlik derecesindeki standart *t* testi olarak yorumlanabilir. Aşağıda görüldüğü gibi, 11. sıradaki gözlem için RStudent = -3,54'tür. Bu değer, önemli ve manidar bir sapma olduğundan böyle bir değerin araştırmacıyı durdurması gerekir. RStudent bir hipotez testi istatistiğinden çok, artığın büyüklüğüne ilişkin bir gösterge olarak düşünülmelidir. (Burada *N* tane *t* testinin hesaplandığı dikkate alınmalı ve bunun hata oranındaki artışı beraberinde getirdiği göz önünde bulundurulmalıdır.) Fakat manidar

| | OBS | X | Y | PRED | RES1D | RSTUDENT | HAT DIAG H | MSE | COOK'S D |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 1 | 1 | 3,23 | −2,23 | −0,87 | 0,20 | 8,22 | 0,10 |
| | 2 | 1 | 2 | 3,23 | −1,22 | −0,47 | 0,20 | 8,71 | 0,03 |
| | 3 | 3 | 3 | 4,71 | −1,71 | −0,62 | 0,11 | 8,55 | 0,03 |
| | 4 | 3 | 5 | 4,71 | 0,29 | 0,10 | 0,11 | 8,91 | 0,00 |
| | 5 | 3 | 7 | 4,71 | 2,29 | 0,85 | 0,11 | 8,26 | 0,05 |
| | 6 | 4 | 6 | 5,45 | 0,55 | 0,19 | 0,09 | 8,88 | 0,00 |
| | 7 | 5 | 8 | 6,19 | 1,81 | 0,65 | 0,08 | 8,52 | 0,02 |
| "A"- > | 8 | 5 | 10 | 6,19 | 3,81 | 1,49 | 0,08 | 7,16 | 0,09 |
| | 9 | 6 | 5 | 6,93 | −1,93 | −0,69 | 0,09 | 8,46 | 0,02 |
| | 10 | 7 | 10 | 7,77 | 2,33 | 0,86 | 0,11 | 8,24 | 0,05 |
| "C"-> | 11 | 10 | 4 | 9,89 | −5,89 | −3,54 | 0,26 | 3,73 | 1,01 |
| "B"-> | 12 | 13 | 14 | 12 ,11 | 1,89 | 0,98 | 0,54 | 8,06 | 0,55 |

SAS Enstitüsü çıktısından uyarlanmıştır.

**Gösterim 15.2** Şekil 15.5'te sunulan veri kümesi için hazırlık istatistikleri

veya değil, 3,54 değerindeki bir standart hata, dikkate alınması gereken beklenmedik bir değerdir. Bu durumda yapılan kestirme iyi olmayacaktır.

Şimdi, H başlıklı sütunda verilen sapma değerlerine dönelim. Bu sütun incelediğinde, baskı değerlerinin ,00 ile ,20 aralığında değiştiğini görürüz ve ortalama baskı değerini (p + 1)/N = 2/12 = 0,167 buluruz; bu değer beklenen değere yakındır. Bununla birlikte, 11'nci ve 12'nci gözlemlere ait baskı değerlerinin yüksek olduğuna dikkat edilmelidir. Bu iki gözlem, "Stevens'ın verdiği 3(p+1)/n = 3(2)/12) = ,50'den büyük olmamalıdır." temel kuralına göre yüksektir. Tabloya bakıldığında, 11'nci gözleme ait artık değerin de yüksek olduğunu görürüz. Bu nedenle, bu gözleme ait orta büyüklükteki bir baskı değeri regresyon üzerindeki önemli bir etkiye işaret edebilir. 12'nci gözleme ait baskı değeri 11'nci gözleme ait baskı değerinin yaklaşık iki katı kadardır. Ancak, 12'nci gözlemin artık değeri küçüktür ve bu gözlem regresyon doğrusuna yakın bir noktada yer almaktadır. Dolayısıyla bu gözlemin regresyon üzerinde daha küçük bir etkisinin olması muhtemeldir.

Cook *D* değeri, regresyon üzerindeki etkiye ilişkin bir ölçümdür. Bu değer, uzaklık, baskı ($h_i$) ve artıkların kareler ortalaması $KO_{\text{artık}}$ değerlerinin bir fonksiyonu olarak değişir. 11'nci ve 12'nci gözlemler hariç son sütundaki değerlerin büyük bir kısmı oldukça küçüktür. 11'nci gözleme ait Cook *D* değeri 1,00'ı geçmektedir. Cook *D*'nin örneklem dağılımı, tartışmalı bir konudur ve hangi değerlerin büyük kabul edilmesi gerektiğine dair genel bir kural yoktur. Bununla birlikte 1,00'ın üzerindeki değerler beklenmedik değerlerdir.

Şekil 15.5'teki üç noktanın her birinin tabloya da yansıdığını ifade ederek Gösterim 15.2'deki sonuçları özetleyebiliriz. *A* noktası oldukça büyük fakat istatistiksel olarak manidar olmayan bir artık değerine sahiptir. Ancak *A*'nın hem baskı değeri hem de regresyon etkisi küçüktür. *B*'nin baskı değeri yüksektir, ancak Cook *D* değeri düşüktür ve *B* noktasının silinmesi $KO_{\text{artık}}$ değerinde kayda değer bir düşüşe neden olmayacaktır. *C* noktasının artık değeri yüksektir. Ayrıca *C* oldukça büyük bir baskı ve Cook *D* değerine sahiptir. *C* noktasının silinmesi $KO_{\text{artık}}$ değerinde önemli bir düşüş sağlayacaktır. Bu, üzerinde oldukça dikkatle düşünülmesi gereken bir gözlem türüdür. Buna rağmen, rahatsız edici olduğu ve silinmesinin sonuçları daha iyi göstereceği gerekçesiyle bir verinin silinmesine hemen karar verilmemelidir. 11'nci gözlem gibi veri üzerinde dikkatli bir biçimde düşünülmelidir. Bu gözlemin dışarıda tutulması ve bu gözleme farklı muamele edilmesi için mantıklı gerekçeler olabilir. Ya da bu gözlem gerçekten hatalı olabilir. Bu gerçek bir veri kümesi olmadığından bu veri kümesi ile daha fazla bir şey yapılamamaktadır.

Regresyon hazırlığına ilişkin istatistiklerin hesaplanması, bir veri kümesinin grafiğine bakılarak regresyon eşitliğini doğrulamaya kıyasla gerektiğinden fazla bir iş yükü olarak görülebilir. Bununla birlikte, sunulan örnekte sadece bir yordayıcı değişkenin olduğu durumlara bakılmıştır. Çok sayıda yordayıcı değişken olduğunda, veri kümesinin grafiğini çizip regresyon eşitliği üzerinde orantısız bir etkiye sahip olan noktaları görsel bir şekilde ortaya koymanın mâkul bir yolu bulunmamaktadır. Bu durumda, en azından tek değişkenli grafikleri, belki de noktaların beklenmedik dağılımlarını aramak için her bir yordayıcı değişken ile ölçüt değişken dikkate alınarak oluşturulan iki değişkenli grafikleri oluşturmanız ve hazırlık istatistiklerini hesaplamanız gerekir. Bu istatistiklere bakarak özel durumlar üzerinde daha yakın çalışmalar yapılabilir.

Eğitim Harcamaları ile SAT puanlarına ilişkin veri kümesine dönülürse, hazırlık istatistiklerine ilişkin bazı ek hususlar açıklanabilir. Gösterim 15.3, SPSS çıktılarında[8] rapor edilen ancak Gösterim 15.1'de bulunmayan ek istatistikleri içermektedir. Bu değerler, regresyon iletişimi kutusundaki "Kaydet (Save)" seçeneği işaretlenip daha sonra uygun istatistikler seçilerek elde edilebilir.

Yukarıdaki hazırlık istatistiklerindeki gözlemlere baktığımızda, bazılarının yüksek artık ve studentized artık değerlerine sahip olduğunu görürüz. Büyük ölçüde dikkate değer olan tek studentized artık değeri 48'inci eyalet olan Batı Virginia'ya aittir. Ancak Cook *D* değerlerine baktığımızda, eyaletlerden hiçbiri beklenmedik bir değere sahip değildir. Bu veri kümesindeki en büyük Cook *D* değeri 0,1230'dur. Bu sonuçlara göre, regresyon

8 SPSS programı; baskı değerlerini ve buna bağlı olarak Studentized Range istatistiklerini SAS, JMP, SYSTAT, BMDP programlarına göre biraz farklı bir şekilde hesaplamaktadır. Baskı değerleri, 1/N faktörü ile daha düşüktür. Ancak bu durum ortalama baskı değerinin (p+1)/N yerine, p/N'e eşit olması dışında yorumlamada bir farka neden olmamaktadır.

| Observation | Residual | Rstudent | Cook's *D* | Hat Diag ($h_i$) |
|---|---|---|---|---|
| 1 | −4,5159 | −,1798 | ,0006 | ,0312 |
| 2 | −11,7667 | ,4899 | ,0122 | ,1121 |
| 3 | 1,4609 | ,0580 | ,0000 | ,0248 |
| 4 | −51,6151 | −2,0661 | ,0924 | ,0410 |
| 5 | −2,9717 | −,1189 | ,0003 | ,0407 |
| … | … | … | … | … |
| 29 | 54,9324 | 2,1943 | ,0972 | ,0371 |
| 30 | −25,6401 | −1,1063 | ,0968 | ,1718 |
| … | … | … | … | … |
| 48 | −61,5098 | −2,4168 | ,0508 | ,0054 |
| 49 | 20,5930 | ,8368 | ,0227 | ,0688 |
| 50 | −34,5975 | −1,3757 | ,0321 | ,0284 |

SPSS'den uyarlanmıştır.

**Gösterim 15.3** Hazırlık istatistikleri

sonuçları üzerinde orantısız bir etkiye sahip olan bir eyalet bulunmadığı rahatlıkla söylenebilir.

## Tanılama Grafikleri

Eyaletlerin tek tek veya birleşiminin regresyon eşitliği üzerinde orantısız bir etkiye sahip olmadığının görülmesi endişelenecek bir şey olmadığı anlamına gelmez. Veri kümesiyle ilgili başka sorunların olması muhtemeldir. Gerçekten SAT'a katılan öğrenci yüzdesi değişkeninin logaritmik dönüşümü alınarak üstesinden gelinen bir problem vardı.

Regresyon katsayılarına ilişkin testler, artıkların yordayıcı değişken puanlarının tümü için eşit olduğunu varsaymaktadır. Bu varsayım, artıkların varyansının her bir yordayıcı değişkenin düzeylerinde ve son regresyon denklemindeki genel $\hat{Y}$ üzerinde koşullu sabit olduğu anlamına gelir. Normalliği kontrol etmek için daima bakılması gereken iki şey, kestirilen değerlere karşı artıkların grafikleri ve artıkların Q-Q grafikleridir. Şekil 15.6'nın en üst kısmında, regresyon eşitliğindeki Eğitim Harcamaları ve LogPctSAT değişkeninin yerine PctSAT'ın kullanılması durumuna ilişkin sözü edilen iki grafik görülmektedir.

Sol üst taraftaki grafik boyunca çizilen doğru, veri ile uyumlu olacak şekilde düzeltilmiş regresyon çizgisidir. Bu çizginin belirgin biçimde eğrisel olduğuna dikkat ediniz. Artık değerler için hiçbir örüntü olmamalıdır. Ancak bu grafiğe bakıldığında artık değerlerin bir örüntüye sahip olduğu açık bir şekilde görülmektedir. Crawley'ye (2007) göre, geceleri gökyüzünün görüntüsü nasılsa bu grafiğin de öyle olması gerekir. Geceleri yıldızlar tüm gökyüzünde dağınık bir biçimde yer aldığı gibi, noktaların da tüm grafik üzerinde dağınık bir biçimde bulunması gerekir. Burada böyle bir durum yoktur. Sol alt tarafta, regresyon eşitliğinde Eğitim Harcamaları değişkeniyle birlikte LogPctSAT değişkeninin kullanılması durumuna ait grafik görülmektedir. Yani, sol üst taraftaki grafik PctSAT değişkeninin yordayıcı olarak kullanıldığı regresyon eşitliğine ait iken, sol alt taraftaki grafik regresyon eşitliğinde LogPctSAT değişkeninin kullanılması durumuna aittir. LogPctSAT değişkenin yordayıcı olarak kullanıldığı bu grafikte noktalar daha dağınık bir görünüme sahiptir. Şekil 15.6'nın sağ tarafında, her iki artık kümesinin de oldukça normal olduğunu görebilirsiniz ki bu oldukça önemlidir. Cohen ve ark. (2003), Levene tarafından geliştirilen artıkların homojenliği testini açıklamıştır. Artıkların homojenliği testi; temelde Bölüm 7'de bağımsız gruplar *t* testi için varyansların heterojenliği tartışılırken yer verilen Levene testi ile aynıdır ve *X* ekseni boyunca artan değerlere karşılık artık değerindeki artma ya da azalmaya odaklanmaktadır. Cai ve Hayes (2008) tarafından, regresyon heterojenliğine karşı çok daha dayanıklı olan bir regresyon katsayısı testi önerilmiştir. Bu testin örnekteki veri kümesine uygulanmasıyla elde edilen sonuçlar, hem Eğitim Harcamaları hem de LogPctSAT değişkenlerine ait regresyon katsayılarının manidar olduğunu doğrulamaktadır.[9]

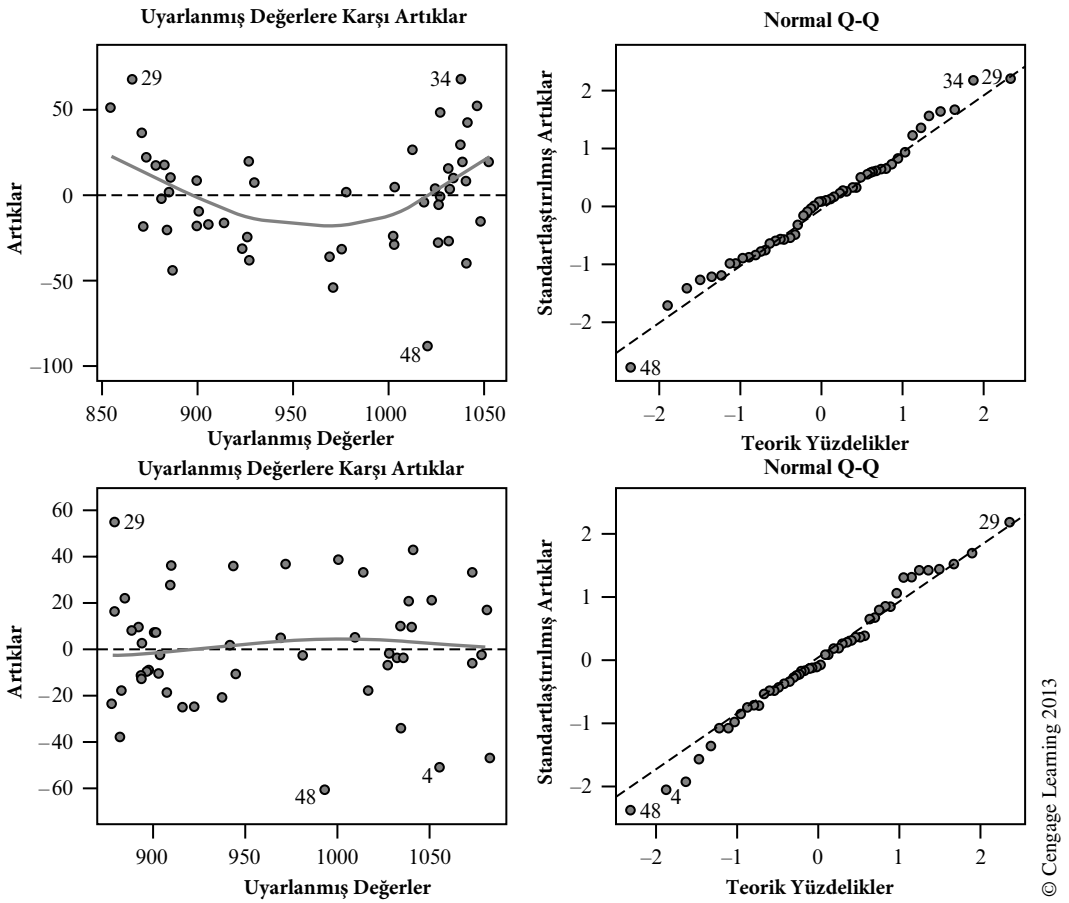
## Modellerin Karşılaştırılması

**iç içe (nested) modeller**

**hiyerarşik modeller**

Bir modeldeki değişkenlerin, ikinci modeldeki değişkenlerin bir alt kümesini temsil ettiği modeller **iç içe (nested)** veya **hiyerarşik** modeller olarak adlandırılır. Örneğin; yordayıcı

9 Cai ve Hayes (2008) bu testleri gerçekleştirmek için SAS macro kullanımını önermiştir. Cai ve Hayes'in (2008) çalışmaları karmaşık olmasına karşın, önerdikleri macro'nun kullanımı kolaydır. Kolaylıkla SAS programına dahil edilebilen Macro uygulaması, Cai ve Hayes'in (2008) çalışmalarında gösterdikleri şekilde kullanılabilir.




© Cengage Learning 2013

**Şekil 15.6** LogPct ile Eğitim Harcamaları değişkeninin yordayıcı olması (üst satır) ve LogPctSAT ile Eğitim Harcamaları Değişkenlerinin yordayıcı olması durumlarına ait artık grafikleri (alt satır)

değişken olarak Eğitim Harcamaları, LogPctSAT, Öğrenci/Öğretmen Oranı ve Öğretmen Maaşı değişkenlerini içeren modelin; Eğitim Harcamaları ve LogPctSAT değişkenlerini içeren Öğrenci/Öğretmen Oranı veya Öğretmen Maaşı değişkenlerini içermeyen modele göre SAT puanlarının daha iyi yordayıp yordamayacağını belirlemek isteyebiliriz. İç-içe desenin olduğu bu tür durumlarda, hangi modelin diğerinden manidar derecede daha iyi olduğunun test edilmesi kolay bir işlemdir. Sadece $R^2$ değerini veya regresyon için kareler toplamını karşılaştırabiliriz.

Örneğin; Eğitim Harcamaları, LogPctSAT, Öğrenci/Öğretmen Oranı ve Öğretmen Maaşı değişkenlerini içeren model ile başladığımızı varsayalım. (Daha önce incelenen modele göre iki tane daha fazla yordayıcı değişken içerdiğinden bu modeli seçtim.) Çoklu $R^2$ değeri ,888 bulunmuş olup varyans analizi özet tablosu aşağıdaki gibidir:

**ANOVA[b]**

| Model | | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 243689,5 | 4 | 60922,385 | 89,539 | ,000[a] |
| | Artık | 30618,141 | 45 | 680,403 | | |
| | Toplam | 274307,7 | 49 | | | |

[a]Kestiriciler: (Sabit), Maaş, Öğrenci/Öğretmen Oranı, Harcama

[b]Bağımlı Değişken: SAT

SPSS Inc. çıktısından adapte edilmiştir.

Daha sonra; Öğrenci/Öğretmen Oranı ve Öğretmen Maaşı değişkenleri modelden çıkarılarak regresyon analizi tekrarlanmıştır. Yordayıcı değişkenler olarak Eğitim Harcamaları ve LogPctSAT değişkenlerinin kullanıldığı bu modele ilişkin $R^2$ değeri ,886 bulunmuş ve bu model için elde edilen ANOVA tablosu aşağıda sunulmuştur.

**ANOVA[b]**

| Model | | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 243069,3 | 2 | 121534,649 | 182,856 | ,000[a] |
| | Artık | 31238,381 | 47 | 664,646 | | |
| | Toplam | 274307,7 | 49 | | | |

[a]Kestiriciler: (Sabit), Maaş, Öğrenci/Öğretmen Oranı, Harcama

[b]Bağımlı Değişken: SAT

SPSS Inc. çıktısından adapte edilmiştir.

Dikkat edilirse, ilk modelde açıklanan varyans oranının ikinci modele göre daha az olduğu görülecektir. Eğer kareler toplamı arasındaki farkları hesaplarsak; 243689,5 – 243069,3 = 620,2 = $KT_{\text{fark}}$ olarak bulunur. Kareler toplamlarına ait bu fark serbestlik derecesine bölünerek kareler ortalamasına dönüştürülebilir; fakat serbestlik derecesi nedir? Basitçe, iki modelin yordayıcı sayıları arasındaki farktır (2). Böylece, $KO_{\text{fark}} = KT_{\text{fark}}/sd$ = 620,2/2 = 310,1 bulunur. Ayrıca, kareler ortalaması arasındaki bu fark, genel modeldeki kareler ortalaması farkına bölünerek test edilebilir. Buna göre,

$$F = \frac{\dfrac{KT_{\text{reg (tam)}} - KT_{\text{reg (indirgenmiş)}}}{sd_{\text{reg (tam)}} - sd_{\text{reg (indirgenmiş)}}}}{KO_{\text{artık (tam)}}} = \frac{\dfrac{(243{,}689.5 - 243{,}069.3)}{2}}{664{,}646} = \frac{310{,}1}{664{,}646} = 0{,}467$$

Bu 2 ve 45 serbestlik derecesinde bir *F* değeridir, manidar değildir. Dolayısıyla, Öğrenci/Öğretmen Oranı ve Öğretmen Maaşı değişkenlerini modele dâhil etmekle SAT puanlarının yordanmasına dair regresyon modelinde iyi bir iş yapmayız.

Eğer; Eğitim Harcamaları, LogPctSAT ve Öğrenci/Öğretmen Oranı değişkenlerinin yordayıcı olarak alındığı bir regresyon modeli oluşturur ve öğretmen başına düşen öğrenci sayısının yer almadığı regresyon modeli ile karşılaştırırsak, *F* değerimiz 0,175 ve bu değerin karekökünü alarak 0,418 elde ederiz. Bu değer modelin genelinde Öğrenci/Öğretmen Oranı değişkenine karşılık gelen *t* değeridir. Diğer bir deyişle, sadece bir değişkeni modelden çıkarmak istediğimizde, bu değişken için modelin genelindeki *t* değerine bakarak değişkeni modelden çıkarmanın manidar olup olmayacağını kolaylıkla belirleyebiliriz.

Fakat iç-içe (nested) model yoksa ne yapabiliriz? Verilen örnek için yordayıcı değişken olarak PctSAT yerine LogPctSAT değişkenini kullanmakla akıllıca bir seçim yapıp yapmadığımızı belirlemek istediğimizde, böyle bir problem karşımıza çıkmaktadır. LogPctSAT değişkeninin yordayıcı olarak kullanıldığı model ile PctSAT değişkeninin yordayıcı olarak alındığı model iç-içe (nested) modeller olmadığından, iki model için $KT_{\text{regresyon}}$ arasındaki fark test edilemez. Bunun yerine, **Akaike Bilgi Kriteri** (ABK) kullanılır. Akaike bilgi kriteri bir oran istatistiğine dayalıdır. (Bu istatistiğe burada yer verilmeyecektir.) SPSS programında Akaike bilgi kriterini hesaplamak için söz dizimine (syntax) başvurulması gerekir, bu işlem oldukça basittir. En az bir istatistiği (örneğin, katsayılar) talep ettiğimizden emin olarak, normalde yaptığımız gibi regresyon analizine başlarız. Ardından normalde yapılandan farklı olarak, doğrudan analizi gerçekleştirmek yerine Paste seçeneğini seçer ve ortaya çıkan söz dizimindeki, istatistik alt komutuna "selection" komutunu ekleriz. Böylece aşağıdaki şekilde bir söz dizimine ulaşırız:

**Akaike Bilgi Kriteri (ABK)**

```
REGRESSION
    /MISSING LISTWISE
    /STATISTICS COEFF OUTS SELECTION
    /CRITERIA=PIN(.05) POUT(.10)
    /NOORIGIN/DEPENDENT CombinedSAT
    /METHOD=ENTER expend PctSat.
```

Eğer bu işlem her iki model için gerçekleştirilirse, PctSAT değişkeninin kullanıldığı modelden aşağıdaki sonuçlar elde edilir.

**Model Özeti**

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Kestirmenin Standart Hatası | Seçim Kriterleri: Akaike Bilgi Kriteri | Amemiya Kestirme Kriteri | Mallows Kestirme Kriteri | Schwarz Bayes Kriteri |
|---|---|---|---|---|---|---|---|---|
| 1 | ,905[a] | ,819 | ,812 | 32,459 | 350,906 | ,204 | 3,000 | 356,642 |

[a]Yordayıcılar: (Sabit), PctSAT ve Eğitim Harcamaları
SPSS programında rapor edilen analiz çıktılarının düzenlenmiş halidir.

LogPctSAT değişkeninin kullanılması durumunda model aşağıdaki sonuçları verir.

**Model Özeti**

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Kestirmenin Standart Hatası | Seçim Kriterleri: Akaike Bilgi Kriteri | Amemiya Kestirme Kriteri | Mallows Kestirme Kriteri | Schwarz Bayes Kriteri |
|---|---|---|---|---|---|---|---|---|
| 1 | ,941[a] | ,886 | ,881 | 25,781 | 327,869 | ,128 | 3,000 | 333,605 |

[a]Yordayıcılar: (Sabit), LogPctSAT ve Eğitim Harcamaları
SPSS programında rapor edilen analiz çıktılarının düzenlenmiş halidir

Akaike bilgi kriterinde, küçük olan modelin daha iyi olduğu kabul edilir. İki tablo incelendiğinde, LogPctSAT değişkenin yordayıcı modele ait Akaike bilgi kriterinin daha küçük olduğu görülmektedir. Bununla birlikte, Akaike bilgi kriteri için 327,869 değeri ile 350,906 değeri arasındaki farkın manidar olup olmadığını gösterecek bir test yoktur. Bundan dolayı, sadece SAT'a katılan öğrenci yüzdesine logaritmik dönüşümün uygulandığı değişkenin daha doğru bir tercih olacağına yönelik buradaki açıklamaları dikkate almak zorundayız.

## 15.11 Regresyon Eşitliğinin Oluşturulması

Daha önce, değişkenler arasındaki ilişkileri anlamak veya bir ölçüt değişkeni yordamak amacıyla hiç regresyon eşitliği yazmamış bir kimsenin karşılaşacağı temel problem, modele dâhil edilecek değişkenlerin seçimidir. Genellikle az sayıda değişken yerine, analize dâhil edilebilecek çok sayıda potansiyel değişken bulunduğundan, bu seçim işlem üzerinde düşünülmesi gereken bir sorun olarak karşımıza çıkar. Ne olacağını görmek için tüm değişkenleri analize alma olasılığı bulunmaktadır. Ancak, sadece nasıl bir sonucun ortaya çıkacağını görmek için tüm değişkenlerin analize alınması pratik olmadığı gibi akıllıca bir uygulama da değildir. Önceki kesimlerde aralarında yüksek korelasyon bulunan değişkenleri belirleme ve bu sayede *Y*'nin yordanmasına ait regresyon modeline dahil edilmesi gereksiz olan değişkenleri saptama konusunda tolerans ve varyans bükülme faktöründen yararlanılabileceği görülmüştü. Bununla birlikte, *Y*'nin yordanmasına ait en uygun regresyon eşitliğini elde edebilmenin başka yolları da bulunmaktadır.

### Seçme Metotları

Büyük bir veri kümesinden en iyi regresyon eşitliğini elde etmenin çok sayıda yolu vardır. Bu kesimde söz konusu yaklaşımlara kısaca yer verilecektir. Ancak öncelikle, bu yaklaşımların tümünün uygun olup olmadığının tartışılması gerekir. Birçok durumda bu yaklaşımların tümü uygun değildir.

Ölçüt hakkında kestirimde bulunmaya yönelik bir araştırma problemi ile ilgilendiğimizi ve büyük bir değişken kümesi ile çok sayıda veri noktasına sahip olduğumuzu varsayalım (bir işi kimin iyi yapacağını belirlemek istiyorsunuz, fakat bu konuda tahminde bulunmanıza yardımcı olabilecek bir teorik bilginiz yoksa). Bu durumda, aşağıdaki modeller yardımcı olabilir. Bununla birlikte, eğer belirli değişkenlerin birtakım çıktılarla ilişkili

olup olmadığına bakarak bazı teorik modelleri test etmeye çalışıyorsanız (örneğin; önemli stres kaynakları, günlük olaylar ve ebeveyn stresi gibi değişkenler yardımıyla yetişkinlerin psikolojik durumlarını tahmin edebilir misiniz?) $R^2$'yi maksimum veya $KO_{artık}$'ı (artıkların kareler ortalaması) minimum düzeyde tutma gibi ölçütler temelinde seçilen bir modelin yararlı olması muhtemel değildir. Regresyon eşitliğinin bu iki ölçüt esas alınarak oluşturulması, mantıksal olarak daha uygun bir model yerine; istatistiksel olarak türetilmiş, sadece kısmî bir uyum gösteren ve belki de manidar olmayan modellere odaklanılmasına yol açabildiğinden, oldukça tehlikeli olabilmektedir. Örneğin; iki psikolojik değişkenden hangisinin regresyon analizine ikinci sırada gireceğini belirlemek için adımsal regresyonun uygulanması, genellikle zayıf teorik bilgiye sahte bir istatiksel zefaret eklemektir. Huberty (1989), değişkenlerin önemine göre sıralanması için adımsal regresyonun kullanımına karşı oldukça katı deliller ileri sürmüştür. Henderson ve Denison (1989), önemli birçok noktayı özetledikleri makalelerinde, adımsal regresyonu akla uygun olmayan bir regresyon yöntemi olarak nitelendirmişlerdir.

Hâlâ, değişken seçim süreçlerinin bazı türlerini kullanarak bir regresyon eşitliği oluşturmak isteyebileceğiniz varsayımından dolayı, aşağıda tüm alt kümeler regresyonu, geriye doğru eleme ve adımsal regresyon şeklinde üç alternatif yaklaşım sunulmuştur. Bu konuyla ilgili okunabilir bir tartışma Draper ve Smith'te (1981, Bölüm 6) bulunabilir.

## Tüm Alt Kümeler Regresyonu

**tüm alt kümeler regresyonu**

Kavramsal düzeyde bu yöntemlerden en basiti, **tüm alt kümeler regresyonu** olarak adlandırılır. Bu yöntemin tüm alt kümeler regresyonu olarak adlandırılmasının nedeni, yordayıcı değişkenlerin olası tüm alt kümelerine bakılarak $R^2$ değerini maksimum yapacak veya hatanın kareler ortalamasını minimum düzeyde tutacak grubun seçilmesi esasına dayanmasıdır. İki veya üç yordayıcı değişken ve biraz sabırla çoklu regresyon analizini hesaplamak için herhangi bir standart bilgisayar programı kullanılarak böyle bir analiz yürütülebilir. Bununla birlikte, çok sayıda değişken olduğunda bu konuyu ele almanın tek yolu, SAS PROC RSQUARE gibi alt kümelerin sayısı ile her bir alt kümede yer alacak en büyük ve en küçük yordayıcı sayısının belirlenmesine imkân tanıyan özel bir program kullanmaktır. (Örneğin, bilgisayar programına "Beş yordayıcı değişken kullanarak en büyük $R^2$ değeriyle birlikte sekiz model verilsin." şeklinde bir talimat verilebilir.)

"En iyi"yi farklı yollarla tanımlayabilirsiniz; bu yollar daima aynı model ile sonuçlanmaz. Modeller, (1) $R^2$'nin büyüklüğü, (2) $KO_{artık}$ değerinin büyüklüğü, (3) Mallow $C_p$ olarak adlandırılan bir istatistik ve (4) "PRESS" olarak adlandırılan bir istatistik aracılığıyla seçilebilir. $R^2$ ile $KO_{artık}$ değerinin büyüklüğü hakkındaki açıklamalara daha önce yer verilmişti. En yüksek $R^2$ değerinin (veya düzeltilmiş $R^2$ değeri) elde edilmesini sağlayacak veya hatayı minimum düzeyde tutacak şekilde yordayıcı değişkenlerin kombinasyonu elde edilmeye çalışılır. Mallow $C_p$ istatistiği, herhangi bir modeldeki hatanın büyüklüğü ile yordayıcı değişkenlerin tümünün yer aldığı genel modeldeki hata terimini karşılaştırır (bkz: Draper ve Smith, 1981, s. 299). Yani, Mallow $C_p$ istatistiği, PRESS istatistiğinde olduğu gibi sadece iç-içe (nested) modeller için geçerlidir. İndirgenmiş modeldeki hata teriminin genel modeldeki hata terimine eşit veya daha yüksek olması gerektiğinden, bu oranın minimum düzeyde tutulması istenir.

PRESS (Predicted RESidual Sum of Squares/Kestirilmiş artık kareler toplamı [KAKT]) istatistiği, $\Sigma(Y_i-\hat{Y}_i)^2$ değerine bakması yönüyle $KO_{artık}$ istatistiğine benzer. Ancak, KAKT istatistiğinde kestirimler, ölçüt dışındaki tüm gözlemleri içeren veri kümesinden yapılır. Modellerin KAKT veya $KO_{artık}$ istatistiği dikkate alınarak sıralanması durumunda ulaşılan sonuç daima aynı olmamakla birlikte, genellikle benzer olur. Etkili (varlık ya da yokluk durumuna bağlı olarak regresyon eşitliğinde büyük farklılığa yol açan) veri noktalarına odaklanmasının daha muhtemel olması (bkz: Draper ve Smith, 1981, s. 325), KAKT istatistiğinin en önemli avantajıdır.

Tüm alt testler regresyon yöntemini gerçekleştirmek için bilgisayar başında geçirilmesi gereken uzun sürenin dışında, şanstan etkilenme potansiyelinin yüksek olması bu yöntemin en önemli dezavantajıdır. Tüm olası modelleri veya en azından olası en iyi modelleri veri kümesine uydurarak, sadece veri kümesine özgü olan veri noktalarıyla en iyi uyumu veren modellerin seçilmesi ciddî bir riske neden olur. Son $R^2$ değerinin, ilgili evren parametresinin yansız bir kestirimi olarak kabul edilmesi mantıklı bir düşünce olmayacaktır.

## Değişken Eleme (Backward Elimination)

değişken eleme

**Değişken eleme** işlemi genel olarak *adımsal regresyon* başlığı altında ele alınan bir yöntemdir. Çünkü değişken eleme işlemi aşamalı bir mantık çerçevesinde çalışır. Hem *adımsal regresyonu* hem de değişken eleme yöntemini gerçekleştirmek için kullanılabilecek bilgisayar programları halihazırda bulunmakla birlikte, bu iki yöntem de standart regresyon işlemleri uygulanarak kolayca yürütülebilir.

Değişken eleme yönteminde, tüm yordayıcı değişkenleri içeren bir model ile işleme başlanır. Hesaplanan modelde, her bir yordayıcı değişkene ilişkin regresyon katsayıları incelenerek veya kısmî ve yarı kısmi korelasyonlara bakılarak modele en az katkı sağlayan değişken modelden çıkarılır (bu değişkenin modele istatistiksel olarak manidar bir katkı sağlamadığı varsayılır). Bu değişkenin çıkarılmasından sonra regresyon analizi tekrarlanır. Yine modele en az katkıyı sağlayan değişken belirlenir ve bu değişken modelden çıkarılarak işleme devam edilir. Alternatif durdurma noktaları mümkün olmasına rağmen, genel olarak bu işlem model içindeki yordayıcı değişkenlerin tamamının modele katkısı manidar olana kadar devam edilir. Örneğin, modeldeki yordayıcı değişken sayısına karşılık $R^2$ veya $KO_{\text{artık}}$ grafikleri çizilip bu grafikteki eğri belirli bir yöne doğru kırılma gösterdiğinde işlem sonlandırılabilir.

Değişken eleme veya adımsal regresyonun gerçekleştirilebildiği bilgisayar programlarının çoğu, "*F* to enter", "*F* to remove", "*p* to enter", "*p* to remove" olarak adlandırılan terimlerin bazı kombinasyonlarını kullanır. Bunlardan sadece birini ele almak için "*p* to remove" seçeneğinin kullanıldığını varsayalım. Eğer $\alpha$ = ,05 manidarlık düzeyine ulaşamayan yordayıcı değişkenlerin modelden çıkarılması planlanıyorsa, "*p* to remove" seçilir ve $\alpha$ = ,05 olarak ayarlanır. "*F* to remove" seçeneği basit bir şekilde, bu *p* değerine karşılık gelen kritik *F* değeri olacaktır.[10] (*F* yerine *t* istatistiğini hesaplayan programlarda uygun değişiklikler basit bir şekilde gerçekleştirilebilir.) Bu durum, gerçekte burada gösterildiğinden çok daha karmaşıktır (bkz: Draper & Smith, 1981, s. 311), fakat pratikte burada anlatılan bir mantık çerçevesinde ilerlemektedir.

Değişken eleme yönteminin en önemli dezavantajı, şanstan etkilenmeye oldukça açık olmasıdır. Değişken eleme yöntemi de çok sayıda değişken ile analize başlandığından, modelden çıkarılacak değişkenler belirlenirken veri kümesindeki değişkenler arasında görülebilecek bastırıcı ilişkiler göz ardı edilebilmektedir. Örneğin, 7'nci ve 8'inci yordayıcı değişkenler arasında bastırıcı bir ilişki varsa, değişken eleme yönteminde bu değişkenlerin her ikisinin modele dâhil edilme olasılığı bulunmaktadır. Eğer bu değişkenler arasındaki ilişki gerçek ise, değişken eleme yönteminde her iki değişkenin de modelde yer alması istenilen bir durumdur. Öte yandan, iki değişken arasındaki ilişki gerçek bir ilişki değilse ölçüt değişkenin açıklanmasına katkı sağlamayacak değişkenler modele dâhil edilmiş olacaktır. Darlington (1990, s. 166), bu durumun hem değişken eleme hem de tüm alt gruplar regresyon yöntemi için söz konusu olabileceğini belirtmiştir. Bastırıcı ilişkilerin gerçek ilişkiler olması oldukça nadirdir. Bundan dolayı, özellikle hipotez testlerine eşlik etmeden bu bastırıcı ilişkilere sistematik olarak bakan yöntemler bu değişkenleri göz ardı etmede en basit yöntemlere göre daha yanıltıcı olabilir.

## Adımsal Regresyon

adımsal regresyon

**Adımsal regresyon** yöntemi geriye doğru eleme yönteminin tersi olarak ifade edilebilir.[11] Bununla birlikte, adımların hiçbirinde diğer tüm değişkenler model içinde yer almadığından ve test işlemi için oldukça çabuk gerçekleştirilebildiğinden konuya değişken eleme yönteminden biraz daha farklı yaklaşılacaktır.

---

10 Draper ve Smith'in (1981) dikkat çektiği gibi, en uygun modeller sınanırken *F* istatistiği normal *F* testi gibi değildir. Dolayısıyla manidarlık değerleri öyleymiş gibi yorumlanmamalıdır. Bu nedenle; *F* ve *p* istatistikleri potansiyel değişkenlerin mantıklı sırasının temelini oluşturmasına rağmen, gerçek olasılık değerlerine çok fazla güvenilmemelidir. McIntyre, Montgomery, Srinivason, ve Weitz (1983) doğrudan bu problemi ele almakta ve testin geniş/serbest doğasını göstermektedir. Draper ve Smith'i izlemek isteyen kişiler için adımsal korelasyon katsayısı testleri için daha uygun olan yönergeler de vermişlerdir.

11 Buradaki terminoloji oldukça kötüdür ancak siz benimle birlikte katlanacaksınız. Geriye doğru eleme, ileriye doğru eleme gibi adımsal bir işlemdir. Bununla birlikte, adımsal regresyon denildiğinde burada anlatılan ve değişkenlerin giderek modele dâhil edildiği süreç akıllara gelmelidir.

Adımsal regresyon;

$$R^2_{0.123\ldots p} = r^2_{01} + r^2_{0(2.1)} + r^2_{p(3.12)} + \cdots$$

gerçeğine dayanır. Eğer birinci yordayıcı değişkeni, ölçüt ile en yüksek korelasyona sahip değişken olarak tanımlarsak, regresyonun ilk adımında sadece birinci yordayıcı değişken modelde yer alacaktır. Daha sonra, $r_{0(i.1)}$, $i = 2...p$ formundaki bütün yarı kısmî korelasyonlar hesaplanır. Yordanan değişken ile en yüksek yarı kısmî korelasyona sahip ve $R^2$ değerinde en büyük artışı sağlayacak ikinci değişken ($X_2$ olduğunu varsayalım) ikinci adımda regresyon modeline girer. Böylece, $X_1$ ve $X_2$ değişkenlerinin yer aldığı regresyon modeli için $Y$ elde edilir. Bunu takiben, eklenen değişkenin modele sağladığı katkının manidar olup olmadığı test edilir. Bunun için doğrudan regresyon katsayının veya yarı kısmî korelasyon katsayısının manidarlığı test edilebileceği gibi $R^2$ değerindeki artışın manidar olup olmadığı da test edilebilir. Hangisi kullanılırsa kullanılsın sonuç aynı olacaktır. $R^2$ değerindeki artışın manidar olup olmadığının test edilmesi daha sonraki işlemlerde avantaj sağlayacağından, burada bu yolu kullanacağız. $f$ sayıda yordayıcı değişkene dayanan bir $R^2$ değeri ile $r$ sayıda yordayıcı değişkene göre hesaplanan bir başka $R^2$ değeri (burada $r$ sayıda yordayıcı $f$ sayıdaki yordayıcının bir alt kümesidir) arasındaki farkın manidarlığını test etmek için aşağıdaki eşitlikten yararlanılır.

$$F_{(f-r,\,N-f-1)} = \frac{(N - f - 1)(R^2_f - R^2_r)}{(f - r)(1 - R^2_f)}$$

$R^2_f$ genel modeldeki $R^2$ değeri = $R^2_{0.12}$ iken; $R^2_r$, değişken sayısının daha az olduğu modeldeki $R^2$ ($R^2_{0.1}$) değeridir. Burada, $f$ genel modeldeki yordayıcı değişken sayısı, $r$ ise değişken sayısının daha az olduğu modeldeki yordayıcı değişken sayısıdır.

Bu işlem, modele dâhil edilmesi $R^2$ değerinde manidar bir artışa yol açmayacak değişkene kadar sürdürülür. Bu süreçteki her bir adımda, yeni bir değişkeni modele dâhil etmeden önce, bir önceki adımda regresyon denklemine eklenen değişkenin sağladığı katkının hâlâ manidar olup olmadığının sorgulanması gerekir. Eğer herhangi bir değişkene ait test kritik $F$ değerinin altında veya $p$ değeri manidarlık düzeyinin üstünde olursa, başka bir değişken eklenmeden önce bu değişkenin modelden çıkarılması gerekir. Bu adımı içermeyen işlemler genellikle **değişken ekleme** (forward selection) olarak adlandırılır.

**değişken ekleme**

Burada sunulan üç seçim yönteminden adımsal regresyon muhtemelen en iyisidir. Hem Draper ve Smith (1981) hem de Darlington (1990) adımsal regresyonu, tahminde bulunmak için aynı evrenden tesadüfi olarak çekilen bir veri kümesi kullanılarak elde edilen en uygun eşitlik ve söz konusu şartlar altında belirli bir veri kümesi için maksimum varyansı kestiren eşitlik arasındaki en uygun uzlaşma olarak önermektedir. Ben burada biraz farklı bir ifade kullanıyorum. Adımsal regresyon için "en uygun yöntem" yerine, zayıf çözümler yerine "kötünün iyisi" demeyi daha uygun buluyorum. Ben son çözüme ulaşmak için herhangi bir mekânik yolun kullanılmasına karşı bir tavsiyede bulunuyorum. Değişkenler hakkında bilinenlerden ve ayrı regresyon analizlerinde görülenlerden faydalanılması gerekir. Büyük ihtimalle modele dâhil edilecek değişkenleri seçmek için en azından kuramsal gerekçeler vardır.

## Çapraz-Geçerlik

**çapraz geçerlik**

Çoklu regresyon çalışmalarının önündeki en önemli engel, bağımsız bir veri kümesinde regresyon eşitliğinin **çapraz-geçerliği**dir. Örneğin, veri kümesi iki veya daha fazla veri kümesine bölünerek ilk veri kümesi için regresyon eşitliği oluşturabilir. Daha sonra, çapraz örneklemde kestirilen $Y$ değerlerini ($\hat{Y}_{\text{çap.geç}}$) elde etmek amacıyla, ilk örneklemde ulaşılan regresyon katsayıları diğer örneklemdeki veri kümesine uygulanır. Ardından, yeni alt örneklemde $Y$ ile $\hat{Y}_{\text{çap.geç.}}$ arasındaki ilişkiye odaklanılır. Eğer regresyon katsayılarının kabul edilebilir düzeyde geçerliği varsa, çapraz geçerlik korelasyonu ($R_{\text{çap.geç.}}$–$Y$ ve $\hat{Y}_{\text{çap.geç.}}$ diğer örnekleme ait regreyon denkleminden kestirilen) yüksek olacaktır. Regresyon katsayılarının geçerliği düşük olduğunda ise, bu korelasyon katsayısı da düşük olur. $R^2_{\text{çap.geç.}}$ değeri neredeyse tüm durumlarda $R^2$'den küçük olur. Çünkü $R^2$, veri kümesi için oluşturulmuş regresyon eşitliğine bağlıdır. Veri kümesindeki her türlü dalgalanmayı (örnekleme hatası içinde) açıklamak için

elinden geleni yapan bir eşitliğe sahip olsak da; bu eşitlik örneklemlerdeki dalgalanmaları yeterince iyi açıklayamadığı zamanlarda şaşırılmamalıdır. $R^2$ ile $R^2_{\text{çap.geç.}}$ arasındaki önemli farklılıklar, regresyon çözümünün kabul edilebilir bir geçerliğinin olmadığının göstergesi olarak yorumlanır.

## Eksik Gözlemler

Regresyon analizini gerçekleştirirken eksik veri bir problem olarak karşımıza çıkar. Bu problemin üstesinden gelebilmek için çok sayıda alternatif yöntem tasarlanmıştır. En yaygın yaklaşım, incelenen değişkenler açısından eksik veriye neden olan gözlemlerin veri kümesinden çıkarılmasıdır. Analizde kullanılacak herhangi bir değişken için eksik veriye neden olan gözlemlerin tamamının çıkarılması esasına dayalı bu işlem **liste bazında silme (listwise deletion)** veya **durum bazında silme (casewise deletion)** olarak adlandırılır. SPSS'te yer alan ancak birçok programda bulunmayan ikinci bir yaklaşım **çiftler bazında silme (pairwise deletion)** olarak isimlendirilmektedir. Eldeki veri ne olursa olsun biz bunu kullanacağız. Örneğin, 13'ncü katılımcı hem $X$ hem de $Y$ değişkenlerine ait veriye sahipse, $r_{XY}$ korelasyonu hesaplanırken bu katılımcıya ait veri kullanılacaktır. Ancak, bu katılımcının $Z$ değişkeni için bir puanı yoksa $r_{XZ}$ veya $r_{YZ}$ korelasyonları hesaplanırken bu katılımcıya ait veri dikkate alınmayacaktır. Çiftler bazında veri silme yöntemi kullanılarak korelasyon matrisi bir kere hesaplandığında, regresyon çözümünün kalanı doğrudan bu matris üzerinden yapılacaktır.

liste bazında silme

durum bazında silme

çiftler bazında silme

Bu iki yöntemin de kendi sorunları vardır. Liste bazında silme yönteminin kullanılması örneklemin küçülmesine yol açar. Ayrıca, veri kümesindeki boşlukların tamamen tesadüfî olmaması halinde (yani veri tesadüfi bir şekilde boş bırakılmamışsa) örneklem seçilmiş olduğu evrenin uygun bir temsilcisi olmayacaktır. Diğer taraftan çiftler bazında silme yönteminin kullanılması, verinin eksik olmaması durumunda elde edilecek korelasyon matrisinden farklı bir korelasyon matrisine ulaşılmasına neden olabilir. Çiftler bazında veri silme yöntemi gerçekte elde edilmesi mümkün olmayan bir korelasyon matrisinin elde edilmesine de yol açabilir. $r_{XY}$ ve $r_{XZ}$ verildiğinde $Y$ ile $Z$ arasındaki korelasyonun belirli değerler arasında bulunması *gerektiği* bilinmektedir, Ancak, korelasyona girecek veri kümesi değişime tabi tutulduğunda, diğer iki korelasyon katsayısı ile uyumsuz olan bir $r_{YZ}$ korelasyon katsayısı elde edilebilir. Böyle tutarsız bir korelasyon matrisinin kullanılması ciddi problemler ile karşılaşılmasına sebebiyet verebilir.

veri eklenmesi

Son yıllarda, eksik veri yerine **veri eklenmesi (imputing)** çalışmalarına büyük önem verilmektedir. Temel olarak eksik değerler, bunlar yerine gelebilecek en uygun tahminler ile yapılır. Bu işlemi gerçekleştirmenin çeşitli yolları vardır. Regresyon ile ekleme, kayıp verinin yerine gelecek değerleri saptamanın en iyi yolu olmasa da görülmesi en kolay olanıdır. Kayıp verinin yerine atanacak değerlerin regresyon ile belirlenmesinde, bir bağımsız değişkeni diğerleri aracılığıyla kestirmek amacıyla eldeki veri kullanılarak regresyon analizi yapılır. Regresyon eşitliği oluşturulduktan sonra, herhangi bir katılımcıya ait eksik verinin bulunmadığı değişkenlere ait puanlar bu eşitliğe yerleştirilir ve ardından katılımcıya ait eksik veri tahmin edilir. Bu şekilde, eksik verinin tümü değiştirilebilmektedir. Sonrasında tamamlanan veri kümesi üzerinden regresyon analizi gerçekleştirilir. Bu özel yaklaşımın kullanılması önerilmemektedir. Ancak, tavsiye edilecek yaklaşımın anlaşılmasını kolaylaştıracağı düşünüldüğünden kullanılması önerilmese de, burada sunulmuştur. Önemli olan nokta, bu işlemin gerçekte sahip olunmayan gözlemlere (eksik veriye) ilişkin mantıklı kestirimlerde bulunmak için uyguladığının görülmesidir. Eksik veri ile ilgili daha kapsamlı bilgi için http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html internet adresine ve Howell'ın (2008b) çalışmasına başvurulabilir.

# 15.12 Her Bir Değişkenin "Önemi"

Çeşitli yordayıcı değişkenlerden hareket ederek bir yordanan değişken hakkında kestirimde bulunmak isteyen bir araştırmacı, aynı zamanda bu yordayıcı değişkenlerden hangisinin $Y$ ölçüt değişkeninin kestirilmesinde en önemli katkıyı sağladığını da belirlemek isteyecektir. Ancak, yordayıcı değişkenlerin karşılıklı olarak bağımsız oldukları durumlar dışında bu sorunun basit bir cevabı yoktur. Önceki kesimlerde görüldüğü gibi; kimi zaman $\beta_j$ (veya $\beta_j^2$) değişkenlerin önemine ilişkin *tek* ölçü olarak alınmaktadır. Bu kabul, $Y$'nin kestirilmesinde her bir yordayıcı değişkenin modele tek katkısının $\beta^2$ değerine eşit olduğu

düşüncesinden yola çıkılarak yapılır. Buna göre, $X_1$ değişkeni $Y$ değişkeni ile bir miktar ortak varyans paylaşmakta ve $\beta_j^2$ gösterimi ile temsil edilen bu ortak varyans miktarının diğer değişkenler ile bir *kesişim noktası* yoktur. $X_1$ yordayıcı değişkeninin diğer yordayıcı değişkenlerle paylaştığı ortak varyans miktarının $Y$ değişkenin yordanmasına sağladığı katkı hakkında söylenebilecek hiçbir şeyin olmaması bu ölçüme ilişkin temel problemdir. Dahası, bağımsız olmayan değişkenlerin bağımsız katkıları ifadesi ile ne anlatılmak istenmektedir?

Darlington (1990) önemin ölçüsü olarak $\beta_j$ değerinin kullanımına karşı bir görüş savunmuştur. $\beta_j$, diğer tüm değişkenler açısından eşit olan ancak $X_j$ değişkeni açısından bir birim farklılık gösteren iki gözlem arasındaki farkı (standart sapma birimi olarak) temsil etmektedir. Ancak burada değişkenlerin yüksek korelasyona sahip olması halinde, bu tür durumların nadiren yaşanacağı dikkate alınmamaktadır.

Yordayıcı değişkenler arasında yüksek korelasyon olduğunda (çoklu doğrusal bağıntı bulunması) değişkenlerin önemine ilişkin bir ölçü olarak $\beta$ ağırlıklarının alınmasının önemli bir dezavantajı, örneklemden örnekleme $R^2$ değeri çok az değişmesine rağmen, $\beta$'nın kararsız değerlere sahip olmasıdır. İki veri kümesi göz önüne alındığında, böyle bir durumda aşağıdaki gibi sonuçla karşılaşılması sürpriz olmaz.

Bir durumda;

$$\hat{Y} = 0{,}50Z_1 + 0{,}25\,Z_2$$

ve diğer durumda

$$\hat{Y} = 0{,}25Z_1 + 0{,}50\,Z_2$$

olabilir. Bu eşitliklerin her ikisinde $R^2$ değerleri yaklaşık aynı olacaktır. Şimdi $Y$ değişkenin yordanmasına her bir değişkenin sağladığı katkı belirlenmek istenirse (verilen bir veri için basit bir şekilde, $Y$ değişkenini kestirmek amacıyla regresyon analizinin kullanılmasının aksine), iki veri kümesi için elde edilen sonuçlar farklı olacaktır. Darlington (1990) bu konuda ilginç bir tartışma sunmakta ve $\beta_i$ değerinin, değişkenlerin önemine ilişkin ölçü olarak oldukça sınırlı bir kullanıma sahip olduğunu ifade etmektedir. Bu konuda daha güçlü bir karşı görüşü Cooley ve Lohnes (1971) ileri sürmüştür. Cooley ve Lohnes (1971), $\beta$ hakkındaki kestirimlerin korelasyon matrisinin bileşenleri hakkındaki tahminlere dayandığına dikkat çekmiştir. Bu matris, her biri örnekleme hatası içeren $p + p(p - 1)/2$ tane korelasyon çiftinden oluştuğundan, Cooley ve Lohnes değişkenlerin uygulamadaki öneminin regresyon katsayılarına bağlanması konusunda son derece dikkatli olunması gerektiğini savunmuşlardır.

Buradaki problemin gösterilmesi kolaydır. Bu kitabın daha önceki basımlarında, 50 üniversite dersinin çeşitli değişkenler açısından puanlandığı bir örnek kullanılmış ve öğretmenin ne kadar iyi bir eğitimci olduğu, sınavların ne derece adil olduğu vb. diğer değişkenlere ait puanların yardımıyla genel puanı kestirmek için regresyon eşitliği hesaplanmıştır. Veri kümesinden elde edilen regresyon eşitliği, standart regresyon katsayılarının kullanıldığı aşağıdaki denklemde verilmiştir.

$$Z_{\hat{Y}} = 0{,}662\text{Öğretmen} + 0{,}106\text{Sınav} + 0{,}325\text{Bilgi} - 0{,}105\text{Not} + 0{,}125\text{Kayıt}$$

Bu regresyon denklemi için $R^2 = ,755$ bulunmuştur. Daha sonra orijinal veri ile aynı kaynaktan seçilen ve yine 50 dersten oluşan ikinci bir örneklem alınmıştır. İkinci örneklem için elde edilen $R^2$ değeri ($R^2 = ,710$), ilk örneklemden hesaplanan $R^2$ değeri ile aşağı yukarı aynıdır. Ancak elde edilen regresyon eşitliği ilk örneklem için oluşturulan regresyon eşitliğinden farklı görünmektedir. İkinci örneklem için standart regresyon katsayılarının kullanıldığı regresyon denklemi aşağıda gösterilmiştir:

$$Z_{\hat{Y}} = 0{,}371\text{Öğretmen} + 0{,}113\text{Sınav} + 0{,}567\text{Bilgi} - 0{,}27\ \text{Not} + 0{,}184\text{Kayıt}$$

İki eşitlik karşılaştırılırsa, bazı $\beta_i$ değerlerinde önemli farklılıklar olduğu açıkça görülmektedir.

Öneme ilişkin kullanılması daha fazla tavsiye edilen bir ölçü, $j$ yordayıcı değişkeni ile ölçüt arasındaki yarı kısmî korelasyonun karesidir (diğer tüm yordayıcı değişkenler $j$ yordayıcısının dışında tutulmaktadır) - yani, $r^2_{0(i.123\ldots p)}$. Darlington (1968) bu ölçümü bir yordayıcı değişkenin yararı olarak ifade etmektedir. Daha önce de görüldüğü gibi, bu yarı kısmi korelasyonun karesi, $i$'yinci yordayıcı değişkeninin modelden çıkarılması sonucunda $R^2$ değerinde oluşan eksilmeyi (ya da bu değişkenin modele eklenmesiyle meydana gelen artış) temsil etmektedir. Eğer temel amaç bir açıklamadan çok bir kestirimde bulunmak ise,

öneme ilişkin en uygun ölçüm bu olacaktır. Bu değerin birçok bilgisayar çıktısından elde edilmesi kolaydır.

$$r^2_{0(i.123\ldots p)} = \frac{F_i(1 - R^2_{0.123\ldots p})}{N - p - 1}$$

Bu formüldeki $F_i$ değeri, tek bir $\beta_i$ (veya $b_i$) katsayısına ilişkin $F$ testi değeridir. (Eğer kullanılan bilgisayar programı katsayılar için $F$ testi yerine $t$ testini rapor ediyorsa; $F = t^2$.) $i = 1,\ldots p$ için $F_i$ hariç diğer tüm terimler sabit olduğundan, değişkenler arasında $F_i$ değerlerine göre yapılan sıralama yarı kısmi korelasyona göre yapılan sıralama ile aynıdır. Dolayısıyla bu değer değişkenleri regresyon eşitliğindeki yararlılıkları açısından sıralamak amacıyla kullanılabilmektedir.

Darlington (1990) değişkenlerin önemi hakkında konuşurken, yarı kısmî korelasyonun karesinin alınmaması şeklinde ilginç bir görüş bildirmiştir. Bununla birlikte, korelasyonların karesinin alınması ya da alınmaması değişkenlerin sıralamasını etkilemeyecektir. (Bir değişkenin mutlak önemi hakkında ikna edici bir tartışma için Darlington'ın kaynağına başvurulabilir.)

Değişkenleri önemine göre sıralamak için yaygın olan ancak kabul görmeyen bir yol, değişkenlerin adımsal regresyona dâhil olma basamaklarına göre sıralanmasıdır. Bu yaklaşımın problemi, değişkenler arasındaki ilişkileri göz ardı etmesidir. Buna göre, dâhil edilecek ilk değişken, sadece ölçüt ile yüksek korelasyonundan dolayı analize girmektedir. İkinci değişken, ilk yordayıcı değişken ile arasındaki ilişki kontrol altına alındıktan sonra ölçütle arasındaki korelasyon esas alınarak analize dahil edilir. Ancak bu ikinci adımda analize giren yordayıcı değişkende ilk adımda analize giren değişken dışındaki yordayıcılar için bir düzeltme uygulanmamaktadır. Analize girecek üçüncü değişken belirlenirken, ilk iki yordayıcı değişken için düzeltme uygulandıktan sonra bu değişken ile yordanan değişken arasındaki korelasyon dikkate alınır ve süreç bu şekilde devam eder. Diğer bir deyişle, her bir yordayıcı değişken farklı bir temelde adımsal regresyona dâhil olur ve buna bağlı olarak değişkenlerin öneminin adımsal regresyona dâhil oluş sırasına göre belirlenmesi fazla mantıklı bir uygulama olmamaktadır. Basit bir örnek ele almak gerekirse, üç yordayıcı değişkenin bulunduğu bir araştırmada, yordayıcı değişkenlerin yordanan değişken ile arasındaki korelasyon katsayılarının sırasıyla ,79; ,78 ve ,32 olduğunu varsayalım. Ayrıca, 1'inci ve 2'inci değişkenler arasındaki korelasyonun ,95; 1 ve 3'üncü değişkenler arasındaki korelasyonun ise ,20 olduğunu kabul edelim. Bu değişkenler regresyon denklemine 1, 3 ve 2 şeklinde bir sırayla girmekte ve regresyon denklemine son giren değişken manidar olmamaktadır. Ancak, 3'üncü değişkenin 2'nci değişkenden daha önemli olduğu ifadesi hangi anlamda kullanılacaktır. Bir eleştirmene böyle bir durumu açıklamaktan nefret ediyorum – gerçekte, böyle bir durumda önem ile ne kastettiğimi söyleme konusunda kendimi baskı altında hissediyorum. Benzer bir nokta, Huberty (1989) tarafından iyi bir biçimde ele alınmıştır. Değişkenlerin önemine ilişkin oldukça başarılı bir tartışma için Harris'e (1985, 79ff) de başvurulabilir.

## 15.13 Yaklaşık Regresyon Katsayılarının Kullanılması

Daha önce regresyon katsayılarının $R$'de önemli değişikliklere yol açmadan bir örneklemden diğerine önemli dalgalanmalar gösterdiğini belirtmiştim. Bu durum, veri kümesinden daha duyarlı kestirimler elde etmek için, regresyon katsayılarının, katsayıların bir temsilcisi olarak bunların kaba kestirimlerinin veriden elde edilecek kesin kestirimlerinin yerine alınmasının tavsiye edilmesine yol açabilir. Örneğin, beş yordayıcı değişkenin bulunduğu bir araştırma probleminde aşağıdaki gibi bir regresyon denklemi elde edilmiş olsun:

$$\hat{Y} = 9{,}2 + 0{,}85X_1 + 2{,}1\,X_2 - 0{,}74\,X_3 + 3{,}6\,X_4 - 2{,}4\,X_5$$

Eğer bu katsayılar yuvarlanırsa ne kadarlık bir kayıp oluşacağının sorgulanması gerekir.

$$\hat{Y} = 10 + 1X_1 + 2\,X_2 - 1X_3 + 4\,X_4 - 2\,X_5$$

Bu soruya verilecek cevap yaşanacak kaybın az olacağıdır. Cohen ve ark. (2003), Dawes ve Corrigan (1974) ile Wainer (1976, 1978) kaynakları bu konuda oldukça yararlı olabilir.

Regresyon katsayılarının yuvarlanmasına ilişkin bu yöntem, düşünüldüğünden çok daha yaygındır. Örneğin, var olan değişik yordayıcıları nicelleştirmek ve genel başarı ortalamasını (GBO) tavsiye mektuplarının iki katı ağırlıklandırmak isteyen bir üniversite kayıt ofisi yetkilisi, gerçek regresyon katsayıları için kaba bir kestirimde bulunuyordur. Benzer şekilde, Minnesota Çok Yönlü Kişilik Envanteri (MÇYKE) [Minnesota Multiphasic Personality Inventory (MMPI)] de katsayıların uygun tam sayılara düşürülmesine dayalıdır. Bu *tanısal göstergelerin (diagnostic signs)* klâsik regresyon eşitliğine göre daha iyi, daha kötü veya eşit sonuçlar üretip üretmediği hâlâ tartışmalı bir konudur. Bu probleme ilişkin eski ama oldukça kapsamlı bir çalışma Goldberg'te (1965) bulunur. Regresyon katsayılarının yuvarlanması işlemi çoklu regresyona olan güveni zayıflatmamalıdır. Çünkü yuvarlanmış katsayılar da yuvarlanmamış katsayılar kadar iyi sonuçlar üretebilmekte hâttâ kimi zaman yuvarlanmış katsayılar yeni veriye uygulandığında daha iyi sonuçlar ortaya çıkabilmektedir. Bu durum ayrıca, regresyon katsayılarındaki küçük farklılıklara çok sıkı bir şekilde bağlanılmaması gerektiğine işaret eder.

## 15.14 Aracı Değişken ve Şartlı Değişken İlişkileri

aracı değişken

şartlı değişken

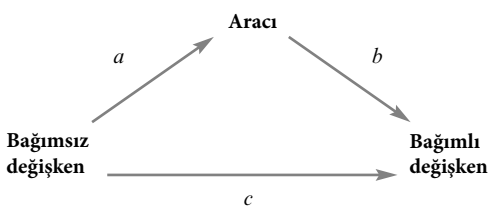
Son 25 yılda, çoklu regresyon ile ilgili psikoloji literatüründe en fazla atıf alan çalışma Baron ve Kenny'nin (1986) **aracı değişken** ve **şartlı değişken** ilişkisi (mediating ve moderating) arasındaki ayrımı kaleme aldıkları araştırmadır. İki değişken arasındaki ilişkinin üçüncü bir değişken tarafından kontrol ediliyor olabileceği düşüncesi, aracı ve şartlı değişkenler arasındaki ayrımın temelini oluşturur. Aracı ve şartlı değişkenler ilişkisi, psikologların üzerinde önemle durdukları konuyu açıklayabilme potansiyeline sahip olduğundan, kitapta bu konu üzerinde fazlaca duracağım. Aracı ve şartlı değişken ilişkilerinin açıklanabilmesi için veri kümesindeki değişkenlere yakından bakılarak iki değişken arasındaki ilişki altında yatan mekanizmaların belirlenmesi gerekecektir.

### Aracılık

aracılık ilişkisi

**Aracılık ilişkisi**nde, iki değişken arasındaki ilişki, başka bazı değişkenlerin aracı etkisiyle oluşmaktadır. Örnek olması için, daha önce sözünü ettiğim bir ilişkiyi ele alacağım. Çocukluk döneminde algılanan anne bakımının yeterlilik hissi ve benlik saygısının gelişmesine katkıda bulunduğu ve yüksek yeterlilik hissi ile benlik saygısının, ileride annelik becerisine ilişkin öz-yeterlilik algısının yüksek olmasını sağladığı ifade edilmişti. Bu örnekte, yeterlilik hissi ve benlik saygısı değişkenleri, çocukluk döneminde algılanan anne bakımı ile annelik becerisine ilişkin öz-yeterlilik algısı arasındaki ilişkiye *aracılık* etmektedir.

Baron ve Kenny'nin (1986), aracılık etkisi ile ilgili tüm hususları ele aldıkları çalışma, aracılık etkisinin anlaşılabilmesi açısından oldukça etkili bir araştırmadır. Baron ve Kenny'ye (1986) göre, aracılık ilişkisinden söz edilebilmesi için çeşitli şartların karşılanması gerekir. Bu şartların ne derece gerekli olduğu tartışmalı bir konudur. Bununla birlikte, aracılık ilişkisi için karşılanması gerekli olduğu ifade edilen bu şartlar ile başlayarak ilerleyeceğiz. Aşağıda, aracılık ilişkisinin şema ile gösterimi sunulmuştur:




© Cengage Learning 2013

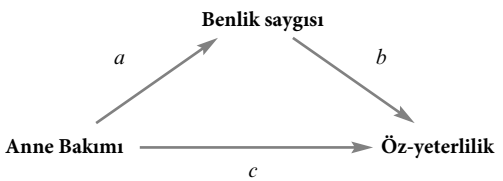
Burada, doğrudan bağımsız değişkenden bağımlı değişkene doğru ilerleyen ve "*c*" ile gösterilen ok açıklanmak istenen baskın ilişkidir. Aracılık ilişkisi, bağımsız değişkenden potansiyel aracı değişkene (*a*) ve bu potansiyel aracı değişkenden de bağımlı değişkene doğru giden (*b*) olmak üzere iki kısımdan oluşmaktadır.

Baron ve Kenny aracılık ilişkisi bulunduğunu iddia edebilmek için, ilk olarak bağımsız değişken ile aracı olduğu düşünülen değişken arasında manidar ilişki bulunduğunun göste-

rilmesi gerektiğini belirtmişlerdir. Bu ifade açık görünebilir, ancak esasında tartışmalı bir konudur. İkinci adım, ilk gerekliliğe benzer nedenlerden dolayı, bağımlı değişken ile aracı değişken arasında ve bağımlı ile bağımsız değişkenler arasında manidar ilişkiler bulunduğunun gösterilmesidir. [Bu gereklilikler mantıklı görünmesine rağmen, iyi bir kanıt bunları gereksiz kılabilir (Hayes, 2009'a bakınız).] MacKinnon, Lockwood, Hoffman West ve Sheets (2002), bu gerekliliklere dayalı bir aracılık ilişkisinin çok düşük bir istatistiksel güç ile sonuçlanabileceğini göstermiştir.

Bu üç şart, üç ilişkinin de (a, b ve c) manidar olmasını gerektirir. Ölçüt değişkeni kestirmek için aracı değişken ile bağımsız değişken aynı anda analize alındığında, bağımsız değişkenin bağımlı değişken üzerindeki yordayıcı rolünün incelendiği önceki analizde iki değişken arasında bulunduğu saptanan ilişkinin önemli ölçüde azalacağını göstermek Baron ve Kenny'nin ilk adımını temsil etmektedir. Diğer bir deyişle, aracı ilişki kontrol altına alındığında bağımsız ve bağımlı değişken arasındaki ilişkinin belirgin ölçüde azalıp azalmadığının belirlenmesi gerekir. Bağımlı ve bağımsız değişken arasındaki ilişkiye ait *c* değerinin 0'a düşmesi aracı değişken ilişkisinin bulunduğuna dair en güçlü kanıttır. Deneyimlerime göre *c* değerinin 0'a kadar düşmesi pek mümkün değildir. Genellikle *c*'nin değeri düşmekte ancak 0'a eşit olmamaktadır (bu değer hâlâ manidar olabilir).

Leerkes ve Crockenberg (1999), çocukların kendi anneleri tarafından nasıl büyütüldüğü ile kendileri çocuk sahibi olduklarında anneliğe ilişkin öz-yeterlilik algıları arasındaki ilişkinin belirlenmesi ile ilgilenmişlerdir. Araştırmayı beş aylık bebekleri olan 92 anne üzerinde yürütmüşlerdir. Araştırmada, çocukluk döneminde algıladığı anne bakımı yüksek olan kişilerin çocuk sahibi olduklarında anneliğe ilişkin öz-yeterlilik algılarının da yüksek olması beklenir. Bununla birlikte, Leerkes ve Crockenberg (1999) benlik saygısının bu ilişkiye aracılık etmiş olabileceğini varsayarak araştırmayı bir adım ileri taşımışlardır. Leerkes ve Crockenberg (1999), çocukluk döneminde algıladığı anne bakımı yüksek olan kişilerin yüksek benlik saygısına sahip olabileceğini ve bu yüksek benlik saygısının anneliğe ilişkin yüksek öz-yeterlilik algısını beraberinde getirebileceğini iddia etmişlerdir. Benzer şekilde, çocukluk döneminde algıladığı anne bakımı düşük olan kişilerin düşük benlik saygısına sahip olacağını ve düşük benlik saygısının anneliğe ilişkin öz-yeterlilik algısında da düşüşe sebep olabileceğini ileri sürmüşlerdir. Çocukluk döneminde algılanan anne bakımı, benlik saygısı ve anneliğe ilişkin öz-yeterlilik algısı arasında bulunduğu ifade edilen bu ilişki aşağıdaki şekilde ifade edilebilir:




© Cengage Learning 2013

Baron ve Kenny'nin savunduğu ön şartların karşılanıp karşılanmadığı, değişkenler arasındaki basit korelasyonlar incelenerek belirlenebilir. SPSS programından yararlanılarak bu korelasyon katsayılarına bakılmış ve elde edilen sonuçlar aşağıda gösterilmiştir:

**Korelasyonlar**

| | | Anne Bakımı | Benlik Saygısı | Öz-Yeterlilik |
|---|---|---|---|---|
| Anne Bakımı | Pearson Korelasyon | 1 | ,403** | ,267* |
| | p | | ,000 | ,010 |
| | N | 92 | 92 | 92 |
| Benlik Saygısı | Pearson Korelasyon | ,403** | 1 | ,377** |
| | p | ,000 | | ,000 |
| | N | 92 | 92 | 92 |
| Öz Yeterlilik | Pearson Korelasyon | ,267* | ,377** | 1 |
| | p | ,010 | ,000 | |
| | N | 92 | 92 | 92 |

**0,01 düzeyinde manidar (2-kuyruklu)
*0,05 düzeyinde manidar (2-kuyruklu)

SPSS çıktısından uyarlanmıştır.

Yukarıdaki tablodan, çocukluk döneminde algılanan anne bakımının hem benlik-saygısı hem de öz-yeterlilik ile ilişkili olduğunu görüyoruz. Ayrıca benlik-saygısı ile öz-yeterlilik arasında da pozitif manidar ilişki bulunmaktadır. Bu ilişkiler Baron ve Kenny tarafından karşılanması önerilen şartları sağlar niteliktedir. Bir sonraki adım, benlik saygısı ve çocukluk döneminde algılanan anne bakımı değişkenleri kullanılarak anneliğe ilişkin öz-yeterlilik algısının kestirilmesidir. Anneliğe ilişkin öz-yeterlilik algısının bağımlı değişken olduğu bu kestirime ait sonuçlar aşağıda görülmektedir.

**Katsayılar[a]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | t | Mnd. | Korelasyonlar | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | Std Hata | Beta | | | Basit | Kısmî | Yarı Kısmî |
| 1 | (Sabit) | 3,266 | ,140 | | 23,243 | ,000 | | | |
| | Anne Bakımı | ,110 | ,042 | ,267 | 2,631 | ,010 | ,267 | ,267 | ,267 |
| 2 | (Sabit) | 2,936 | ,173 | | 16,954 | ,000 | | | |
| | Anne Bakımı | ,057 | ,044 | ,138 | 1,297 | ,198 | ,267 | ,136 | ,126 |
| | Benlik saygısı | ,146 | ,048 | ,321 | 3,022 | ,003 | ,377 | ,305 | ,294 |

SPSS çıktısından uyarlanmıştır.

[a] Bağımlı Değişken: Anneliğe ilişkin öz yeterlilik algısı

Tablodaki ilk modelde, anneliğe ilişkin öz yeterlilik algısını kestirmek amacıyla kullanılan tek yordayıcı değişken çocukluk döneminde algılanan anne bakımıdır. Tabloya bakıldığında, çocukluk dönemimde algılanan anne bakımı ile birlikte benlik-saygısı değişkeni de modele dâhil edildiğinde, benlik-saygısının annelik öz-yeterliliğine ait manidar bir yordayıcı olduğu, anne bakımı değişkeninin ise artık annelik öz-yeterliğinin manidar bir yordayıcısı olmadığı görülmektedir ($t$ = 1,297, $p$ = ,198). Bu sonuç, benlik-saygısı değişkeninin çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı arasındaki ilişkide aracı bir role sahip olduğuna dair bir kanıt olarak yorumlanabilir. Analiz sonuçları ayrıca, yarı kısmî korelasyonun SPSS tarafından "part korelasyon" olarak rapor edildiğini göstermektedir. Çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı arasındaki basit (zero-order) korelasyon ,287 iken benlik saygısı değişkeni kontrol altına alındığında, iki değişken arasındaki yarı kısmî korelasyon ,126 olmaktadır

Bu sonuçlar, Leerkes and Crockenberg tarafından ileri sürülen "Çocukluk dönemimde algılanan anne bakımı ile anneliğe ilişkin öz yeterlilik algısı arasındaki ilişkide, benlik-saygısı aracı bir rol oynamaktadır." hipotezini desteklemektedir. Bu sonuca göre, bireyin çocukluk döneminde algıladığı anne bakımının yüksek olması benlik-saygısının gelişmesine katkıda bulunmakta ve benlik saygısının gelişmiş olması ise çocuk sahibi olduğunda annelik öz-yeterliliğinin de yüksek olmasını sağlamaktadır.

Çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz yeterlilik algısı arasındaki ilişki başlangıçta manidar iken; benlik-saygısı değişkeni modele dahil edildiğinde, bu ilişkinin artık manidar olmadığı görülmektedir. Buna göre, Leerkes ve Crockenberg'ün çalışmasına ait veri Baron ve Kenny tarafından aracılık ilişkisinde karşılanması gerekli olduğu ifade edilen ikinci şartı sağlamaktadır. Ancak, böyle bir sonuç her zaman mümkün olmamaktadır. Esasen, başlangıçta manidar olan bir ilişkinin potansiyel aracı değişkenin analize girmesinden sonra manidar olmaması pek sık karşılaşılan bir durum değildir. Aracı değişkenin modele dâhil edilmesinin başlangıçta manidar olan bir ilişkide düşüşe neden olması çok daha sık karşılaşılan bir sonuçtur. Bu durumda ne yapılması gerektiği hakkında önemli tartışmalar olmuştur. Bununla ilgili iki farklı yaklaşıma değinilecektir. Sobel (1982), Baron ve Kenny tarafından önerilen teste kıyasla daha basit bir test ileri sürmüştür. Sobel (1982) tarafından önerilen bu test aşağıda açıklanmaktadır; fakat bu testin de çeşitli problemleri vardır. İlk olarak, kimse $t$ oranını elde etmek için gerekli olan standart hata değeri hakkında aynı fikirde görünmemektedir. Ayrıca dolaylı ilişkinin (çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı arasında benlik saygısı değişkeninin aracı etkisiyle görülen ilişki) örneklem dağılımının normal olduğunu varsaymakta, ancak bu varsayımın genellikle gerçekleşmediği bilinmektedir. Bu durum kısa bir süre içerisinde görülecektir.

Bağımlı ve bağımsız değişken arasındaki ilişkinin değerinin düşmesine rağmen, hâlâ manidar olduğunda, aracılık ilişkisini test etmenin bir yolu, bağımsız değişkenden aracı

**Tablo 15.4** Aracı değişkenin iki kısmı için standart hatalar ve regresyon katsayıları

| Yol a | | | Yol *b* (Anne bakımı ayrıştırılarak) | | |
|---|---|---|---|---|---|
| Anne Bakımı | → | Benlik- Saygısı | Benlik Saygısı | → | Annelik Öz-yeterliği |
| *b* | | 0,364 | *b* | | 0,146 |
| $\beta$ | | 0,403 | $\beta$ | | 0,321 |
| $S_{\beta a}$ | | 0,096 | $S_{\beta b}$ | | ,106 |
| *t* | | 4,178* | *t* | | 3,033* |

© Cengage Learning 2013

değişkene ve aracı değişkenden bağımlı değişkene giden ilişkinin tamamının manidar olup olmadığına bakılmasıdır. Sobel (1982) tarafından önerilen testte bu işlem yapılmaktadır.

İki regresyon eşitliğine ait önemli istatistikler Tablo 15.4'te görülmektedir. SPSS, Beta için standart hata değerini rapor etmediğinden, bu değerin standart hatasının hesaplanması gerekir. Tabloda verilen *t* değerleri, standartlaştırılmamış regresyon katsayısının (*b*) standart hataya veya standartlaştırılmış regresyon katsayısını bu katsayıya ait standart hataya bölünmesiyle elde edilebilir. Buna göre, aşağıdaki eşitlikler elde edilir:

$$t = \frac{\beta}{s_\beta}; \quad s_\beta = \frac{\beta}{t} = \frac{0,403}{4,178} = 0,096$$

Benzer şekilde, çocukluk döneminde algılanan anne bakımı kontrol altına alındıktan sonra benlik-saygısından annelik öz-yeterliliğine giden yol için aşağıdaki ifadeler bulunur.

$$t = \frac{\beta}{s_\beta}; \quad s_\beta = \frac{\beta}{t} = \frac{0,323}{3,041} = 0,106$$

Elde edilen sonuçlar Tablo 15.4'te görülmektedir.

Daha sonra; Anne bakımı → Benlik-saygısı → Annelik öz-yeterliği yolu için regresyon katsayısı, *a* ile *b* ilgili yol olmak üzere $\beta_a$ x $\beta_b$ = 0,403 × 0,321 = 0,129 bulunur. Bu değerin, benlik saygısının kontrol edilmesiyle, çocukluk döneminde algılanan anne bakımı ve annelik öz yeterliği arasındaki yarı kısmî korelasyona eşit olduğuna dikkat edilmelidir. (Yol *c* doğrudan çocukluk döneminde algılanan anne bakımından benlik saygısına ilerleyen yoldur.) Ek olarak; bu iki parçadan oluşan yol için standart hata,

$$s_{\beta_a\beta_b} = \sqrt{\beta_a^2 s_b^2 + \beta_b^2 s_a^2 - s_a^2 s_b^2}$$

olur. Bu formülde, $\beta_a$ ve $\beta_b$ yollar, $s_a^2$ ve $s_b^2$ de bu yollara ait regresyon katsayılarının standart hataların kareleridir.[12] Burada, iki yolun (path) birleşimine ait standart hata aşağıdaki gibi hesaplanabilir:

$$s_{\beta a \beta b} = \sqrt{\beta_a^2 s_b^2 + \beta_b^2 s_a^2 - s_a^2 s_b^2} = \sqrt{,403^2(,106^2) + ,321^2(,096^2) - (,106^2)(,096^2)}$$
$$= \sqrt{0,0027} = 0,052$$

Yol katsayısının (0,403 x 0,321 = 0,129) ve standart hatasının (0,052) olduğunu biliyoruz, bu değerlerden yararlanılarak *t* değerini aşağıdaki gibi buluruz:

$$t = \frac{b_a b_b}{s_{\beta_a} s_{\beta_b}} = \frac{,129}{,052} = 2,48$$

Sobel (1982) bu oranının asimptotik normal dağılıma sahip olduğunu ifade etmektedir. Büyük örneklemlerde, bu oran 1,96 değerini geçtiğinde yokluk hipotezinin *a* = ,05 manidarlık düzeyinde reddedilmesine neden olur. Küçük örneklemler için muhtemelen *N* – 3 serbestlik derecesinde bir *t* dağılımı olurdu. Verilen örnek için yukarıdaki sonuçlardan da beklendiği gibi yol, istatistiksel olarak manidardır. Bu nedenle, benlik saygısının çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı arasında güçlü bir aracı etkiye sahip olduğu sonucuna varılabilir. Çocukluk döneminde algılanan anne bakı-

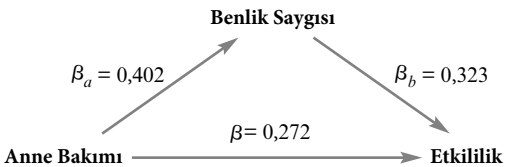
12 Daha önce söz edildiği gibi, bu eşitliklerin tam şekli ile ilgili olarak araştırmacılar arasında anlaşmazlıklar vardır. Ancak, burada sunulan eşitlik Baron ve Kenny tarafından önerilmektedir.

mından anneliğe ilişkin öz-yeterlilik algısına doğrudan giden yola ait regresyon katsayısı (yarı kısmî korelasyon) manidar olmadığından, çocukluk döneminde algılanan anne bakımının annelik öz-yeterliliği üzerindeki asıl etkisinin benlik saygısı değişkeni aracılığıyla oluştuğu görülmektedir.

## Özyükleme (Bootsrapping) Kullanımına İlişkin Alternatif Bir Yaklaşım

Baron ve Kenny tarafından yapılan araştırma, aracılık ilişkisinin anlaşılabilmesi için çok etkili bir kaynaktır. Bununla birlikte, bu araştırmaya çeşitli gerekçelerle birtakım itirazlar olmuştur (MacKinnon, Lockwood, Hoffman, West ve Sheets'e (2002) bakınız). Preacher ve Hayes (2009), Baron ve Kenny'nin aracılık ilişkisi hakkında öne sürdüğü kuralların ve Sobel tarafından ortaya konulan testin göz ardı edilmesine dair ikna edici kanıtlar ileri sürmüş ve özyükleme yaklaşımına başvurularak sadece dolaylı ilişkinin istatistiksel manidarlık testine odaklanılmasını savunmuştur. Baron ve Kenny'nin ileri sürdüğü kurallar görünürde oldukça mantıklıymış izlenimi uyandırsa da, bu kuralların neden gereksiz olduğuna ilişkin açıklamaları görmek için Preacher ve Hayes'in çalışmasına bakılabilir.

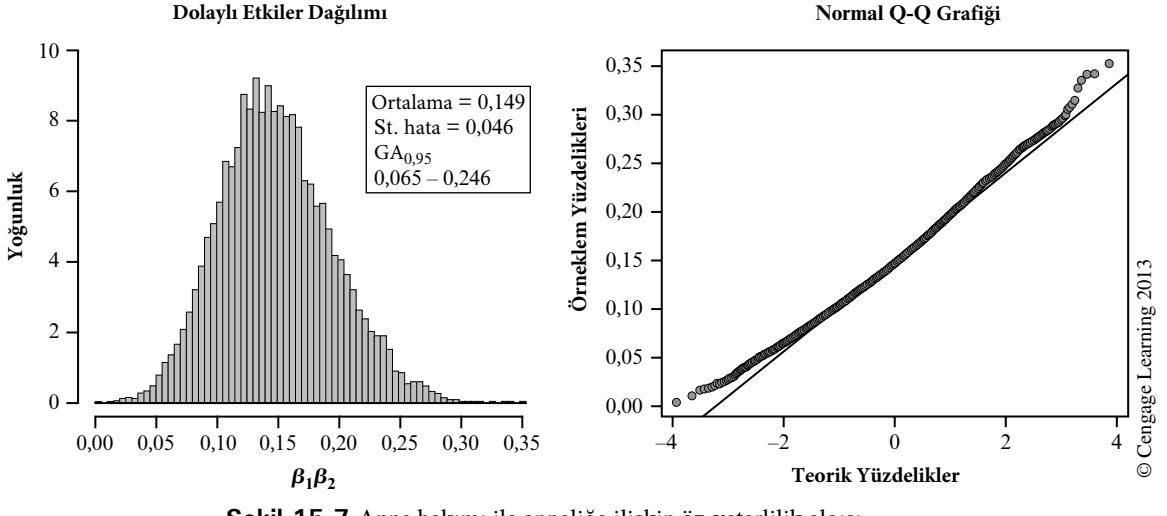
Aşağıda gösterildiği gibi, Leerkes ve Crockenberg tarafından ortaya konulan model ile başlayacağız



© Cengage Learning 2013

Çocukluk döneminde algılanan anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı arasındaki dolaylı yol $\beta_a \times \beta_b$ = 0,402 × 0,323 = 0,130 ifadesiyle tanımlanabilir. Burada cevaplanması gereken soru, bu ilişkinin 0,00'dan manidar düzeyde farklı olup olmadığıdır. Bu soruyu cevaplayabilmenin yollarından biri, 92 gözlemin değiştirilmesiyle her biri orijinal veriden olmak üzere 10.000 seçim yaparak özyükleme örneklemler oluşturmaktır. Her bir seçim için standart regresyon katsayıları ile $\beta_a\beta_b$ hesaplanır. Bu hesaplamalar sonunda elde edilecek 10.000 değerin dağılımı $\beta_a\beta_b$'nin örneklem dağılımına karşılık gelecektir. Ayrıca, $\beta_a\beta_b$'nin standart sapması $\beta_a\beta_b$'ye ait standart hata değeri olacaktır. Daha sonra bu değerler yardımıyla dolaylı ilişkiye ait güven aralığı bulunabilir. Bulunan aralığın %2,5 ile %97,5 yüzdelikleri arasında 0,00 noktasını içermemesi durumunda, dolaylı ilişkinin istatistiksel olarak manidar olduğu sonucuna varılabilir. (Yeniden örneklemeyi gerçekleştirmek için *R* içinde yazılmış bir program bu kitabın internet sitesinden alınabilir.)

Bu sonuçlar Şekil 15.7'nin sol tarafında gösterilmiştir. Şekilden de görüleceği gibi,




**Şekil 15.7** Anne bakımı ile anneliğe ilişkin öz-yeterlilik algısı

örneklem dağılımı kısmen sağa doğru çarpıktır. Şekle bakıldığında, güven aralığı 0,065 ≤ $\beta_a\beta_b$ ≤ 0,246 olarak görülmektedir. Bu aralığın 0,00 noktasını içermemesi dolaylı ilişkinin manidar olduğunu doğrulamaktadır. Şekil 15.4'ün sağ tarafındakki Q-Q grafiği, her bir örneklemdeki 92 gözlem aşırı çarpık olmamakla birlikte, örneklem dağılımının, Hays'in belirttiği gibi çarpık olduğunu göstermektedir.

Literatürde aracılık ilişkisinin test edilmesinde en uygun yaklaşımın hangisi olduğuna dair önemli tartışmalar olmuştur. Standart hatanın alternatif kestirimini kullanarak çevrim-içi bir *t*-testi yaklaşımı için http://people.ku.edu/~preacher/sobel/sobel.htm adresine bakınız. Bu adreste, Preacher ve Hayes (2004) söz konusu problemi cevaplamak için özyükleme yönteminin kullanılmasına izin veren SPSS ve SAS macro programlarını sunmaktadır. Wellington Üniveritesi'nden Paul Jose tarafından hazırlanan ve http://www.victoria.ac.nz/psyc/paul-jose-files/helpcentre/help7_mediation_example.php adresinden ulaşılabilen çalışma aracılık ilişkisi hakkında önemli bilgiler içermektedir. Ek olarak, Jose aracılık ilişkisi için Microsoft Excel'in bir uzantısı olarak tasarlanan, ücretsiz çalışan ve http://www.victoria.ac.nz/psyc/paul-jose-files/medgraph/medgraph adresinden ulaşılabilen bir hesaplayıcı önermektedir, ben bunu çok faydalı buldum. Son olarak, alternatif yaklaşımlar hakkında kapsamlı bir karşılaştırma için MacKinnon, Lockwood, Hoffman, West ve Sheets'e (2002) başvurulması önerilebilir. Aracılık ve şartlı değişken ilişkilerinin tamamı hakkında başarılı bir tartışma için başvurulabilecek bir kaynak olarak Beaujian'e (2008) başvurulabilir.

## Şartlı İlişkiler

**şartlı değişken ilişkileri**

Aracı ilişkiler, bağımsız değişkenin bağımlı değişken üzerindeki etkisi aracılığıyla değişken veya değişkenleri tanımlamaya çalışırken, **şartlı değişken ilişkileri** bağımlı ve bağımsız değişkenler arasındaki ilişkilerin üçüncü bir değişkenin (şartlı değişken) düzeylerinin fonksiyonu olarak farklılaştığı ilişkileri açıklamaya çalışır.

Wagner, Compas ve Howell (1988), daha fazla stres yaşayan bireylerin daha düşük stres yaşayan bireylere göre daha yüksek düzeyde psikolojik rahatsızlık belirtileri yaşayacağı şeklinde bir hipotez kurmuşlardır. Çalışmada bireylerin yaşadıkları stres düzeyini belirlemede günlük sorunlara ilişkin bir ölçme aracı kullanılmıştır. Yani, varyans analizi açısından günlük sorunların temel etkisi incelenmiştir. Bununla birlikte; bireylerin yaşadıkları stresin üstesinden gelmelerine yardımcı olacak yüksek sosyal destek algılamaları durumunda, günlük yaşamlarındaki sorunlar artsa bile yaşadıkları psikolojik rahatsızlık belirtilerindeki artışın daha düşük olması beklenmektedir. Öte yandan, sosyal destek algısı düşük olan bireylerin günlük yaşamlarındaki sorunlar arttıkça yaşadıkları psikolojik rahatsızlık belirtilerinin de çok hızlı bir şekilde yükseleceği öngörülmektedir.

Wagner ve ark. (1988) çalışmalarını üniversiteye başlamadan önce bir uyum (oryantasyon) eğitimine katılan öğrencilerden oluşan bir araştırma grubu üzerinde yürütmüşlerdir. Öğrencilerden yakın zamanlarda yaşadıkları küçük stresli olayların sayısı ile algıladıkları sosyal desteği rapor etmeleri istenmiştir. Daha sonra, bir kontrol listesi yardımıyla öğrencilerin son bir ay içinde yaşadıkları psikolojik rahatsızlık belirtileri sayısı belirlenmeye çalışılmıştır. Veri toplama süreci sonunda 56 katılımcıya ait gözlemlerden oluşan bir veri kümesine ulaşılmıştır. Söz konusu veri kümesi, bu kitabın internet adresinde "hassles.dat" adı ile kayıtlıdır.

İlk adımımız değişkenler arasındaki ilişkilerin incelenmesidir. Değişkenler arasındaki ilişkileri gösteren korelasyon matrisi aşağıdaki gibidir:

**Korelasyonlar**

Pearson Korelasyon

| | Sorunlar | Destek | Semptomlar |
|---|---|---|---|
| Sorunlar | 1,00 | –,167 | ,577** |
| Destek | –,167 | 1,00 | –,134 |
| Semptomlar | ,577** | –,134 | 1,00 |

**$p$ < ,01 (,01 düzeyinde manidardır.)

Beklenildiği gibi, sorunlar ile psikolojik rahatsızlık belirtileri (semptomlar) arasındaki ilişki manidardır ($r$ = ,557). Algılanan sosyal desteğin sorunlar ve psikolojik rahatsızlık belirtileriyle arasındaki ilişkiler ise manidar değildir. Bununla birlikte saptanan bu ilişkiler günlük yaşamda karşılaşılan sorunlar ile psikolojik rahatsızlık belirtileri arasındaki ilişkinin algılanan sosyal destek düzeyine göre farklılaşıp farklılaşmadığı sorusuna cevap olamamaktadır.[13]

Burada cevaplanmaya çalışılan araştırma problemi, varyans analizindeki etkileşimin arkasında yatan probleme benzemektedir. Gerçekte, bu etkileşimi test edebilmenin yolu, günlük yaşamda karşılaşılan sorunlar ile algılanan sosyal desteğin çarpımı olan bir değişken oluşturmaktır. (Burada yapılan işlem, varyans analizine genel doğrusal model yaklaşımının sunulduğu bir sonraki bölümde yapılan işlemlere benzemektedir.) Bununla birlikte, sorunlar ve sosyal destek değişkenlerini sadece çarparsak sonuçlarda iki problemle karşılaşacağız. İlk etapta, sorunlar ve sosyal destek değişkenlerinden biri veya her ikisi bunlardan oluşturulan çarpımla yüksek korelasyon verecektir. Bu yüksek korelasyon veri kümesinde çoklu doğrusallık (multicollinearity) bağıntı sorununu da beraberinde getirecektir. Bu durum, sorunlar ve algılanan sosyal desteğin temel etkisine ilişkin katsayıların manidarlığına dair testi ve bu katsayıların büyüklüğünü önemli ölçüde etkileyecektir. İkinci problem ise, sorunlar ve algılanan sosyal destek değişkenlerinin regresyon analizindeki etkisi incelenirken, değişkenlerden biri diğer değişkenin 0 düzeyinde test edilecektir. Diğer bir deyişle, sorunlara ilişkin test gerçekleştirilirken, bir katılımcının algıladığı sosyal desteğin 0 olması halinde sorunların psikolojik rahatsızlık belirtileri ile ilişkili olup olmadığı test edilir. Benzer şekilde, algılanan sosyal desteğe ilişkin test günlük yaşamda hiç sorun yaşamayan bireyler için gerçekleştirilir. Hem çoklu doğrusal bağıntı problemi hem de bir temel etkinin diğer temel etkinin uç noktasında test edilmesi istenmeyen bir durumdur.

**merkezleme**

Bu iki problemin üstesinden gelebilmek için veriyi **merkez**leyeceğiz. Verinin merkezlenmesi, her bir değişkenin merkezlenmesinin bu değişkene ait ölçümlerden çıkarılarak sapma puanlarının oluşturulması anlamına gelir. Bu işlemden sonra, günlük yaşamda karşılaşılan sorunlar değişkenine ait 0 puanı (merkezlenmiş), bu değişkenin merkezlenmesine eşit puana sahip olan gözlemleri temsil eder. Buna bağlı olarak, verinin merkezlenmesinin ardından algılanan sosyal desteğin temel etkisinin günlük yaşamda karşılaşılan sorunlar değişkenin 0 noktasında test edilmesi esnasında bu değişkenin merkezlenmesine karşılık gelen noktada test edilmesi anlamına gelir. Aynı şekilde, verinin merkezlenmesinin ardından, günlük yaşamda karşılaşılan sorunların temel etkisi, algılanan sosyal desteğin 0 noktasında değil, bu değişkenin merkezlenmişine karşı gelen noktada test edilmesidir. Bu sayede yukarıda ifade edilen ikinci problem çözülmüş olmaktadır. Çünkü temel etkilerden biri diğer temel etkinin mâkul bir noktasında test edilmektedir. Verinin merkezlenmesinin ardından elde edilen korelasyon sonuçlarına bakıldığında, çoklu doğrusal bağıntısının önemli ölçüde azaldığı görülmektedir. Dolayısıyla, verinin mekezlenmesinin yukarıda belirtilen problemlerden ilkinin aşılmasına da katkı sağladığı söylenebilir.

Değişkenlerin merkezlenmesinin ardından, bu merkezlenmiş değişkenlerin çarpımından yeni bir değişken oluşturulur. Sorunlar, algılanan sosyal destek ve psikolojik rahatsızlık belirtileri değişkenlerine ait merkezlemeler sırasıyla 170,1964, 28,9643 ve 90,4286'dır. Bu değişkenlerin merkezlenmesine ve merkezlenmiş değişkenlerin etkileşimine ait eşitlikler aşağıdaki gibidir. Değişken adının önündeki "c" (merkezlenme/centered) harfi bu değişkenin ortalandığını gösterir.

csorun = sorun – 170,1964
cdestek = destek – 28,9463
csordes = csorun × cdestek

Ortalanmış (ve ortalanmamış) veri kümeleri arasındaki korelasyonlar aşağıdaki tabloda gösterilmiştir. Bu tabloya, günlük yaşamda karşılaşılan sorunlar ile algılanan sosyal destek değişkenlerinin merkezlenmemiş hallerinden türetilen değişkenlere ait korelasyon katsayılarını ne kadar yüksek olduğunu göstermek amacıyla dahil ettim. Dolayısıyla aşağıdaki tablo sayesinde, günlük yaşamda karşılaşılan sorunlar ve algılanan sosyal destekten türeti-

13 Bu tartışma, bir değişkene ilişkin regresyon katsayısının diğer değişkenin düzeyleri sabit tutulmasına dair daha önceki açıklamaları hatırlatabilir. Her bir regresyon katsayısının yaklaşık olarak aynı olacağının varsayılması gerekir. Bir başka deyişle, bir değişkenin diğer değişkenlerle etkileşiminin olmadığı ifade edilmektedir.

**Korelasyonlar**

Pearson Korelasyon

| | Sorun | Destek | Semptomlar | sorders | csorun | cdestek | csordes |
|---|---|---|---|---|---|---|---|
| Sorun | 1,000 | –,167 | ,577** | ,910** | 1,000** | –,167 | –,297* |
| Destek | –,167 | 1,000 | –,134 | –,510** | –,167 | 1,000** | ,402** |
| Semptomlar | ,577** | –,134 | 1,000 | ,585** | ,577** | –,134 | –,391** |
| sorders | ,910** | –,510** | ,585** | 1,000 | ,910** | –,510** | –,576** |
| csorun | 1,000** | –,167 | ,577** | ,910** | 1,000 | –,167 | –,297* |
| cdestek | –,167 | 1,000** | –,134 | –,510** | –,167 | 1,000 | ,402** |
| csordes | –,297* | ,402** | –,391** | –,576** | –,297* | ,402** | 1,000 |

© Cengage Learning 2013

**Korelasyon 0,01 düzeyinde manidardır) (2-yönlü).
*Korelasyon 0,05 düzeyinde manidardır) (2-yönlü).

**Tablo 15.5** Sorunlar ve psikolojik rahatsızlık belirtileri arasındaki şartlı değişken ilişkisi için regresyon çözümü

len değişken ile günlük yaşamda karşılaşılan sorunlar arasındaki korelasyonunun, verinin merkezlenmesinden nasıl etkilendiği görülebilmektedir. Tablo incelendiğinde görüleceği gibi, verinin ortalanması, temel etkiler ile etkileşimler arasındaki korelasyonu önemli ölçüde düşürmektedir. Bu bizim amacımızdır. Değişkenlerin merkezlenmesi, aralarındaki korelasyonu değiştirmez – sadece bu değişkenlerin etkileşiminden oluşan yeni değişken arasındaki korelasyonu değiştirir.

**Model Özeti**

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Kestirimin Standart Hatası |
|---|---|---|---|---|
| 1 | ,623[a] | ,388 | ,353 | 16,8932 |

[a]Yordayıcılar: (Sabit), CSORUN, CDESTEK, CSORDES

**ANOVA[b]**

| Model | | Kareler Toplamı | sd | Kareler Ortalaması | F | Mnd. |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 9427,898 | 3 | 3142,633 | 11,012 | ,000[a] |
| | Artık | 14839,816 | 52 | 285,381 | | |
| | Toplam | 24267,714 | 55 | | | |

[a] Yordayıcılar: (Sabit), csorun, cdestek, csordes
[b] Bağımlı Değişken: Semptomlar

**Katsayılar[a]**

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | t | Mnd. |
|---|---|---|---|---|---|---|
| | | B | Std. Hata | Beta | | |
| 1 | (Sabit) | 89,585 | 2,292 | | 39,094 | ,000 |
| | csorun | 8,594E-02 | ,019 | ,506 | 4,473 | ,000 |
| | cdestek | ,146 | ,305 | ,057 | ,479 | ,634 |
| | csordes | -5,06E-03 | ,002 | -,262 | -2,144 | ,037 |

[a] Bağımlı Değişken: Semptomlar
SPSS programında rapor edilen analiz çıktılarından uyarlanmıştır.

Şimdi, etkileşim terimi ortalanmış diğer iki yordayıcı değişken ile birlikte regresyon analizine dâhil edilerek iki yordayıcı değişkenin etkileşimi incelenebilir. Semptom belirtileri, regresyon analizinin bağımlı değişkenidir. Regresyon analizine ilişkin sonuçlar Tablo 15.5'te gösterilmiştir. (Ortalanmış değişkenlerin etkileşimi kullanıldığı sürece, temel etki için ortalanmış veya ortalanmamış değişkenlerden hangisinin kullanıldığı önemli değildir [sabit hariç]. Buradaki örnekte temel etki için ortalanmış veriyi kullandım.)

Regresyon analizine ilişkin çıktılar incelendiğinde $R^2$ değerinin ,388 ve bu değerin manidar olduğu görülmektedir. (Etkileşim terimi olmadan $R^2$ değeri ,334 bulunmuştu - tabloda gösterilmemiştir.) Regresyon katsayılarının sunulduğu tabloda, ortalanmış sorun değişkeni ile etkileşim teriminin manidar olduğu görülmektedir (sırasıyla, $p$ = ,000 ve $p$ = ,037). Fakat algılanan sosyal destek değişkeni manidar değildir. Algılanan sosyal destek değişkenine ait $t$ değeri, bu değişkenin regresyon analizinden çıkarılmasının $R^2$ değerinde manidar bir düşüşe neden olmayacağını gösterse de; etkileşim terimi sosyal destek değişkenini içerdiğinden bu değişkenin regresyon çözümünde yer almasına izin verilmiştir.

Analiz sonuçlarına göre regresyon eşitliğimiz aşağıdaki hale gelir:

$$\hat{Y} = ,086\ \text{csorun} + 0,146\ \text{cdestek} - ,005\ \text{csordes} + 89,585$$

Böylece başlangıç noktasındaki problemi cevapladık (algılanan sosyal destek, günlük yaşamda algılanan sorunlar ile semptom belirtileri arasındaki ilişkiyi hafifletmektedir). Bununla birlikte, bu ilişkilerin grafikle gösterimi etkileşim ilişkisinin yorumlanmasına yardımcı olabilir. Finney, Mitchell, Cronkite ve Moos (1984), Jaccard, Turrisi ve Wan (1990) ile Aiken ve West (1991) kaynaklarında bu yaklaşım ile ilgili oldukça kapsamlı tartışmalar bulunmaktadır. Özellikle, Aiken ve West (1991) şartlı aracılık ilişkisi ile ilgili en temel eserdir. Normand Péladeau, http://www.provalisresearch.com/ internet adresinden ücretsiz olarak erişilebilen Italass[14] adlı bir program sunmuştur. Bu program bilgisayar ekranında etkileşimin grafiğini sunmakta ve bu grafik üzerinde algılanan sosyal destek değişkenin düzeylerini değiştirebileceğiniz bir araç çubuğu sağlamaktadır.

En basit çözüm, csorun ile csemptom (psikolojik rahatsızlık belirtileri) arasındaki ilişkiye algılanan sosyal desteğin sabit bir düzeyi için bakılmasıdır. cdestek değişkeni dağılımı bu değişkene ait puanların -21 ile +19 arasında değiştiğini göstermektedir. Burada; –15, 0 ve +15, cdestek değişkenine ilişkin düşük, orta ve yüksek puanları temsil etmektedir. (Bu değerlerden herhangi birini kullanma zorunluluğu yoktur. İstenilen herhangi bir değer kullanılabilir. Burada, olup biten işlemlerin daha iyi görülebilmesi için uçtaki değerler kullanılmıştır.)

İlk olarak, regresyon katsayıları yerine genel gösterimlerini koyarak regresyon denklemini yeniden yazacağım. Ayrıca, csordes değişkeninin nasıl hesaplandığını dikkate alarak csordes için csorun × cdestek ifadesini kullanacağım. Son olarak, işlemleri kolaylaştırmak için değişkenlerin sırasını yeniden düzenleyeceğim.

$$\hat{Y} = b_1\text{csorun} + b_2\text{cdestek} - b_3\text{csordes} + b_0$$
$$\hat{Y} = b_0 + b_2\text{cdestek} + b_3(\text{csorun} \times \text{cdestek}) + b_1\text{csorun}$$

Bu ifadede gerekli düzenlemeler yapıldığında,

$$\hat{Y} = b_0 + b_2\text{cdestek} + \text{csorun}(b_3\text{cdestek} + b_1)$$

elde edilir. Daha sonra gerçek regresyon katsayıları yerine konulursa aşağıdai denklem bulunur.

$$\hat{Y} = [89,585 + 0,146 \times 15] + \text{csorun}\ (-,005 \times 15 + ,086)$$
$$\hat{Y} = 91,755 + 0,11 \times \text{csorun}$$

Bu, bir doğrusal denklemdir ve algılanan sosyal destek yüksek olduğunda (yani, 15), günlük yaşamda karşılaşılan sorunlar ile $\hat{Y}$ arasındaki ilişkiyi temsil eder.

---

[14] Péladeau'nun SimStat programı ücretli olmasına karşın, Italassi ücretsiz olarak kullanılabilen bir programdır.

Şimdi algılanan sosyal desteğin orta düzeyde veya düşük olduğu durumlarda günlük yaşamda karşılaşılan sorunlar ile $\hat{Y}$ arasındaki ilişkiye karşılık gelen eşitliği belirlemek için sosyal destek yerine 0 ve –15 değerlerini koyalım:

Algılanan sosyal destek 0 olduğunda,

$\hat{Y}$ = 89,585 +,086 × csorun

Sosyal destek = –15 olduğunda,

$\hat{Y}$ = 87,395 + ,161 × csorun

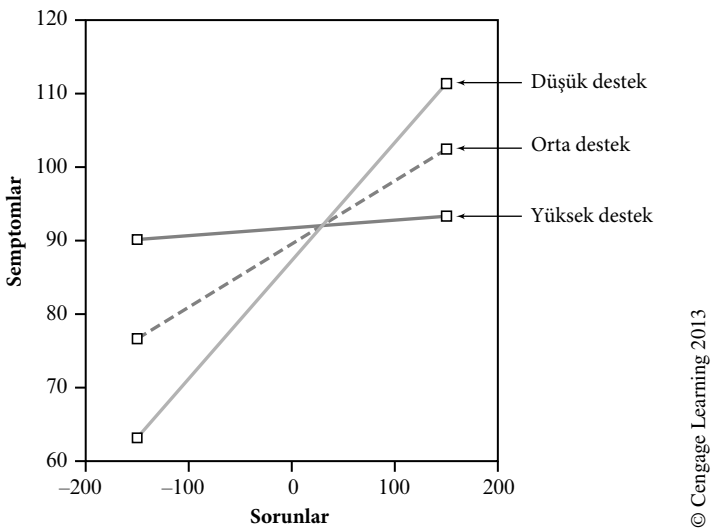
denklemleri elde edilir.

Sorun değişkeninin frekans dağılımına baktığımda, düşük, orta ve yüksek puanların yaklaşık olarak -150, 0 ve 150 olduğunu gördüm. Bu nedenle, cdestek değişkeninin düşük, orta ve yüksek düzeylerinin her biri için sorun değişkenine ait düşük, orta ve yüksek değerler ile semptom belirtileri için kestirilen değerleri hesapladım, sonuçları aşağıdaki tabloda gösterdim. Tablodaki değerleri yukarıdaki üç regresyon eşitliğinden yararlanarak elde ettim. Eşitlikteki csorun yerine -150, 0 ve 150 değerlerini koydum.

| | | Merkezlenmiş Destek | | |
|---|---|---|---|---|
| | | −15 | 0 | 15 |
| Ortalanmış Sorun | –150 | 63,245 | 76,685 | 90,105 |
| | 0 | 87,395 | 89,585 | 91,755 |
| | 150 | 111,545 | 102,485 | 93,405 |

© Cengage Learning 2013

Sosyal destek değişkeninin farklı düzeyleri için kestirilen değerlerin grafikleri ayrı ayrı çizilirse, algılanan sosyal destek yüksek olduğunda günlük yaşamdaki karşılaşılan sorunlardaki artma semptom belirtilerinde daha düşük bir artışa neden olmaktadır. cdestek değişkeni, bu değişkenin orta düzeyi olan 0 noktasına taşındığında (cdestek = 0), günlük yaşamda karşılaşılan sorunlardaki artış semptom belirtilerinde daha yüksek bir artışa neden olmaktadır. Son olarak algılanan sosyal destek düşük olduğunda (cdestek = –15), günlük yaşamda karşılaşılan sorunlardaki artış semptom belirtilerinde çarpıcı bir artışa sebebiyet vermektedir. Algılanan sosyal destek düzeyine bağlı olarak günlük yaşamda karşılaşılan sorunlar ile psikolojik rahatsızlık belirtileri arasındaki ilişki Şekil 15.8'de gösterilmiştir.




© Cengage Learning 2013

**Şekil 15.8** Sosyal desteğin farklı düzeyleri için günlük yaşamda karşılaşılan sorunların bir fonksiyonu olarak psikolojik rahatsızlık belirtilerinin grafiği

Burada örnek olarak verilen araştırma probleminde olduğu gibi, veri analizinde etkileşim teriminin (yani, $X_1 \times X_2$) kullanılması son yıllarda psikolojide yaygın hale gelmiştir. Bununla birlikte, $X_1$ ve $X_2$ değişkenlerinin düzeylerinin kasıtlı olarak değiştirildiği deneysel desenler başta olmak üzere $X_1$ x $X_2$ etkileşimine ait terimin manidar olduğu durumların bulunması oldukça güçtür. McClelland ve Judd (1993) bu problemi araştırmış ve standart saha çalışmaları desenlerinin etkileşimlerini belirlemede gücünün yetersizliğinin nereden geldiğini göstermiştir. Bu, deneysel olmayan araştırmalarda etkileşimin araştırılmasında önemli bir makaledir.

## 15.15 Lojistik Regresyon

lojistik regresyon

Son birkaç yıldır **lojistik regresyon** psikoloji literatüründe popüler bir istatistiksel teknik haline gelmiştir. (Lojistik regresyonun salgın hastalıkları inceleyen bilim dallarındaki ve sağlık alanındaki popülaritesi daha eskiye dayanmaktadır.) Lojistik regresyon, bağımlı değişkenin iki kategorili olduğu veri kümelerine uygulanabilen bir istatistiksel tekniktir.[15] Bir hastalık için uygulanan tedavinin etkililiğini belirlemek amacıyla tedaviden sonra hayatta kalan hastaların 1 ve yaşamını yitirenlerin 0 şeklinde kodlanması lojistik regresyonun sağlık alanında oldukça yaygın şekilde kullanıldığı durumlardan biridir. Benzer bir kodlamaya psikoloji alanında da başvurulmaktadır. Danışanın psikolojik durumunda iyileşme varsa 1, iyileşme yoksa 0 şeklinde bir kodlama uygulanabilir veya başarılı = 1 ve başarısız = 0 şeklinde kodlama yapılabilir. Verilen örneklerdeki gibi, bağımlı değişkenin iki kategorili olduğu durumlarda, lojistik regresyon için bir muhtemel aday daha vardır

diskriminant analizi

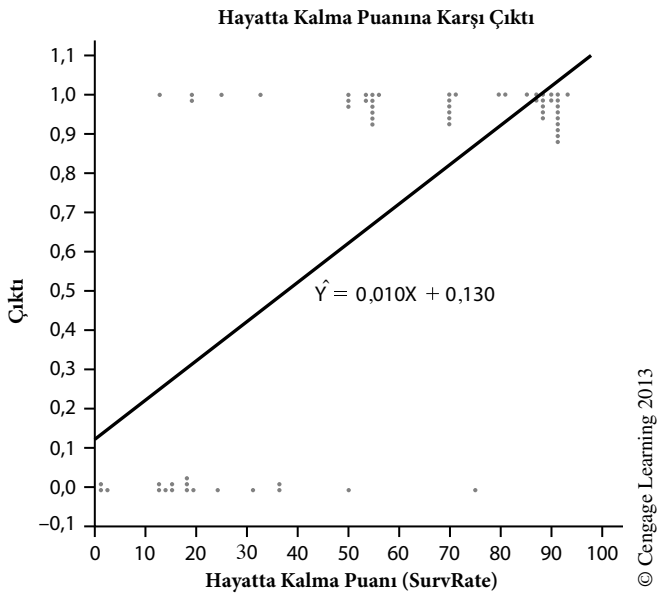
Fakat iki kategorili bir bağımlı değişken olduğunda, analize aday olabilecek en az iki istatistiksel teknik vardır. Bunlardan biri, bir dizi değişkene dayalı iki veya daha fazla grubu birbirinden ayırt etmek için kullanılan bir istatistiksel teknik olan **diskriminant analizi**dir. Lojistik regresyonun diskriminant analizinden daha iyi bir istatistiksel teknik olup olmadığı genellikle akıllara gelen bir sorudur. Daha iyinin nasıl tanımlandığı her zaman açık değildir. Bununla birlikte, diskriminant analizinin regresyon analizine kıyasla iki zayıf yönü vardır. İlk olarak, diskriminant analizinde 0 ile 1 aralığının dışında kalan olasılık değerleri üretilebilmektedir. Bu aralığın dışında kalan bir olasılık değerinin mümkün olmadığı bilinmektedir. İkinci olarak, diskriminant analizi bağımsız değişkenin normalliği gibi gerçekçi olmayan bir varsayıma bağlıdır. Diğer taraftan, lojistik regresyon 0 ile 1 aralığının dışında kalan olasılık değerleri üretmez ve bağımsız değişkenin normalliği gibi bir varsayımın karşılanmasını gerektirmez. Lojistik regresyonda bağımsız değişken kategorili veya sürekli bir değişken olabilir. Buna bağlı olarak, çalışmalarda lojistik regresyon, diskriminant analizine göre daha fazla uygulanır hale gelmektedir.

İkinci alternatif, iki kategorili değişkenin bağımlı değişken olarak kullanılıp standart çoklu regresyon analizinin uygulanmasıdır. Gerçekte birçok durumda sonuçlar benzer olmaktadır. Ancak, genelde lojistik regresyonun tercih edilmesi için gerekçeler vardır. Bu açıklamaya rağmen, basit bir örnek ele alacağım. Örnek olması açısından, bireylerin kanser tedavisine verdikleri cevaplarla ilgili gerçek bir veri kümesini kısmen değiştirerek ele alacağım. Alacağım veri kümesi, yakın zamanda kanser teşhisi konulan insanlardaki davranışsal değişkenler ve stresi konu edinen bir çalışmaya aittir. Çalışmanın amacı doğrultusunda, en azından bir yıldır çalışmada yer alan hastalara bakılacaktır. Tedaviden sonra, iyileşme gözlenen hastalar (Çıktı) 1 ve tedavinin etkili olmadığı veya hayatını kaybeden hastalar 0 ile kodlanacaktır. (Kodlamalar sırasında kullanılan rakamlar tutarlı olduğu sürece hangi rakamların kullanıldığı önemi değildir. Örneğin, 1 ve 2 veya 5 ve 8 gibi bir kodlama da kullanılabilir[16].) 66 katılımcıdan 48'inde tedavi sonrası iyileşme gözlenirken, 16'sında iyileşme gözlenmemiştir. Çalışmaya, hastaya doktoru tarafından hastalığın tanı aşamasında verilen Hayatta Kalma Puanı şeklinde tek bir bağımlı değişken ile başlandığını varsayalım. Hayatta Kalma Puanı 0 ile 100 arasında değişen bir puan olup 5 yıl hayatta kalma olasılı-

---

[15] Lojistik regresyon bağımlı değişkenin üç veya daha fazla kategorisinin olduğu durumlarda da kullanılabilmektedir. Ancak bağımlı değişkendeki kategori sayısının ikiden fazla olduğu durumlara burada yer verilmeyecektir.

[16] Kodlama işlemi sırasında dikkatli olunmalıdır. Çünkü farklı bilgisayar programları aynı kodlara farklı şekilde muamele edebilmektedir. Bilgisayar programlarının bir kısmı yüksek puanı başarı ve düşük puanı başarısızlık olarak kodlarken; diğer bir kısmı için tersi bir durum söz konusudur. Eğer analiz sonucunda beklenenin tam tersi bir sonuç ile karşılaşıldıysa, kullanılan bilgisayar programın iki kategorili veriye nasıl muamele ettiğini görmek için programın kullanma kılavuzuna bakılması yararlı olur.




**Şekil 15.9** Hayatta kalma puanının bir fonksiyonu olarak çıktı

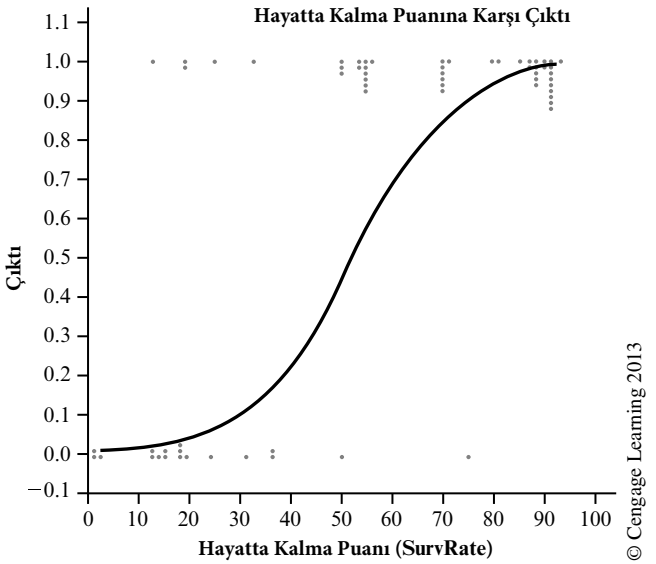
ğını temsil etmektedir. Burada verilen örneğe ait veri kümesi kitabın internet sayfasında Survrate.dat adı ile kayıtlıdır. Dolayısıyla kitaba ait internet sayfası ziyaret edilerek çalışmaya ait veri kümesine erişilebilir.

Hayatta kalma puanı ile Sonuçlar arasındaki ilişkiye bakmanın yollarından biri, Çıktıların *Y* ekseninde gösterildiği bir grafiğinin çizilmesidir. Böyle bir grafik Şekil 15.9'da görülmektedir. [Bu grafikte çakışan noktaların görülebilmesi için hafifçe ayrı gösterdim. Hayatta kalma süresine ait 91 puanı için (SurvRate = 91) çıktının 1 değerinde dizilmesinin nedeni budur.] Beklendiği gibi, grafikte, hayatta kalma puanı yüksek olan kişilerin tedaviden sonra iyileşme oranının daha yüksek olduğu görülmektedir. Şu an için çok sayıda katılımcının bulunduğunu ve bu katılımcılar için hayatta kalma puanının her bir değeri ile ilişkili olarak ortalama puanın (0 ve 1 puanlarının ortalaması) hesaplanabildiğini varsayalım. (Bu ortalamalar hayatta kalma puanının değerlerine bağlı olduğundan, **koşullu ortalamalar** olarak adlandırılır.) Koşullu ortalamalar, hayatta kalma puanının herhangi bir değerinde iyileşme gösteren hasta oranı olur. Eğer standart bir regresyon çizgisi bu veri kümesine uydurulursa, bu doğru hayatta kalma puanının bir fonksiyonu olarak iyileşme olasılığına ait regresyon doğrusu olur. Ancak tahmin edilebileceği gibi, hayatta kalma puanının birçok değeri için *kestirilen* olasılık 0 ile 1 aralığının dışında kalacaktır. (Bir olasılık değerinin bu aralığın dışında kalması ise imkânsızdır.) Bu durum tek başına standart doğrusal regresyonu zayıf bir seçenek haline getirir. İkinci bir problem daha vardır. Eğer hayatta kalma puanının farklı değerleri için Sonuçların *varyansı* hesaplanırsa, bu varyans değerlerinin hem küçük hem de büyük hayatta kalma puanı değerlerinde oldukça küçük olduğu görülecektir. (Çünkü hayatta kalma puanı düşük olan kişilerin neredeyse hepsi 0, hayatta kalma puanı yüksek olan kişilerin ise neredeyse tümü 1 şeklinde kodlanmaktadır.) Ancak, hayatta kalma puanı orta düzeyde olan kişiler 0 ve 1 kodlarının eşit karışımı şeklindedir ve bu durum varyansın yüksek olmasına neden olur. Bu durumun varyansların homojenliği varsayımının ihlâline neden olacağı açıkça görülür. Bununla birlikte normallik hakkında bir şey söylememektedir. Bu nedenlerden dolayı, hayatta kalma puanının tüm değerleri boyunca iyileşme puanlarının yüzdesi 20'nin altına düşmediği veya 80'nin üzerinde çıkmadığı sürece oldukça iyi kestirimler üretse de standart doğrusal regresyon iki kategorili bağımlı değişkenlerde akıllıca bir seçim değildir. Diğer bir problem gerçek ilişkinin doğrusal olmasının muhtemel olmamasıdır. Ölçeğin orta noktasının yakınında hayatta kalma puanındaki farklılıklar ölçeğin uç noktalarındaki farklılıklara kıyasla çıktılarda daha büyük farklılıklara yol açmaktadır.

koşullu ortalamalar

çift kıvrımlı (sigmoidal)

Şekil 15.9'daki doğru veriye iyi bir uyum vermezken; bir S-şeklinde veya çift kıvrımlı **(sigmodel)** bir eğri veri ile daha uyumlu olacaktır. Bu şekildeki bir çizgi hayatta kalma



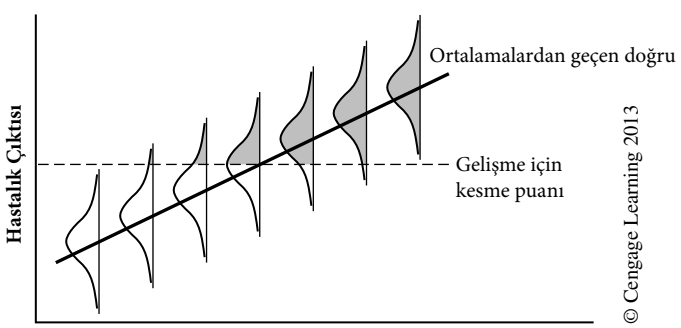

**Şekil 15.10** Çıktının kestirilmesi için daha uygun bir regresyon çizgisi

puanının düşük değerlerinde ilerlendiğinde küçük değişimler gösterirken; orta değerlerinde hızlı ve hayatta kalma puanının yüksek değerlerinde hareket edildiğinde tekrar küçük artışlar gösterecektir. Hiçbir zaman 0 noktasının altına düşmemekte veya 1 değerinin üstüne çıkmamaktadır. Bu grafik Şekil 15.10'da gösterilmiştir.

Dikkat edilirse, bu çizginin sol alt tarafta kalan noktaların tümüne oldukça yakın olduğu görülmektedir. Bu çizgi, hayatta kalma puanı iyileşen ve iyileşmeyen hasta sayısının yaklaşık aynı olduğu değerlerinde hızlı bir biçimde yükselmektedir. Ardından sağ üst taraftaki noktalar kümesine oldukça yaklaşmaktadır. Hayatta kalma puanı değiştikçe nasıl bir iyileşme olasılığı bekleyebileceğinizi düşündüğünüzde bu eğri mantıklı görünecektir.

Çıktıları standart doğrusal regresyone bağlamak için bir yol daha vardır. Önceki kesimlerde regresyon analizi ile ilgili söylenenler düşünüldüğünde, en azından büyük örneklemler için her bir $X$ değerine karşı gelen bir $Y$ değeri vardır. Normallik ve varyansların homojenliği varsayımlarına ilişkin konuşulanlar Şekil 15.9'da şematik olarak görülmektedir. Kişileri iyileşti ya da iyileşmedi şeklinde sınıflandırmak yerine, hastalık sonuçlarını çok daha duyarlı bir şekilde ölçebildiğini varsayalım. (Örneğin, iyileşme durumlarını 100 puanlık bir ölçek üzerinde puanlayabiliriz.) Daha sonra hayatta kalma puanı = 20 olan kişiler için hastalık sonucu puanlarına ait tam bir dağılım elde edilecektir. Aynı durum hayatta kalma puanı = 30, Hayatta Kalma puanı = 40 vb. kişiler için de geçerlidir. Bu dağılımlar Şekil 15.11'de gösterilmiştir.

Bir kişiyi iyileşmiş olarak sınıflandırdığımızda, bu kişinin hastalık puanına ilişkin çıktıların iyileşmiş kategorisinde yer alabilecek kadar yüksek olduğunu söylemiş oluruz.




**Şekil 15.11** Hayatta kalma puanının fonksiyonu olarak hastalık sonuçları

Bu kişiler tamamen iyileşmiş olabilir, tedaviden sonra kendilerini oldukça iyi hissedebilir veya tedaviden sonra çok küçük bir iyileşme gösterse de iyileşmiş kategorisinde yer almak için gerekli ölçütü karşılamış olabilir. Benzer şekilde, bir kişi tedavi sonucunda sabit kalıp hiç etkilenmemiş olabilir, tedavinin ardından kendini daha kötü hissedebilir veya hayatını kaybedebilir. Ancak bu üç durumda da kişi iyileşme ölçütünü karşılayamadığından iyileşmemiş kategorisinde yer alacaktır.

**sonradan sınıflandırılmış veri (censored data)**

**Sonradan sınıflandırılmış veri (censored data)** ile ne anlatılmaktadır? Sonradan sınıflandırılmış demekle elindeki büyük siyah kalemle başkalarının görmesini istemediği şeyleri kapatan küçük bir adamı kastetmiyorum. Belirli bir kesme noktasının üstünde olan puanların başarılı, altında kalanların başarısız olarak sınıflandırıldığı bir durumdan söz ediyorum. Burada ölçülen bir testteki performans, Boston Maratonu için eleme süresi ya da bir hava yolu şirketine ait uçuşların zamanında ya da geç şeklinde sınıflandırılması olabilir. Bu bakış açısıyla lojistik regresyon, doğrusal regresyonun sonradan sınıflandırılmış veriye uygulanmış hali olarak düşünülebilir. Çünkü veri sadece "başarılı" ve "başarısız" şeklinde sınıflandırılmak üzere sınıflandırılmıştır, modelin biraz farklı uydurulması gerekir.

Şekil 15.11'deki grafiğin yatay ekseni kestiği nokta kritik bir değeri temsil eder. Bu çizginin üzerindeki herhangi bir değer gelişmiş olarak sınıflandırılırken çizginin altındaki herhangi bir değerde gelişmemiş olarak sınıflandırılır. Görüleceği gibi, gelişme oranı her bir eğrinin taralı alanıyla temsil edilir. Bu gelişme oranı soldan sağa doğru gidildikçe, başlangıçta yavaş bir seyir izlerken daha sonra hızlanmakta ve ardından tekrar yavaşlamaktadır. Soldan sağa doğru gidildikçe yaşanan bu değişim, Şekil 15.10'da görünen çift kıvrımlı (sigmoidal) eğriyi hatırlatmaktadır, çünkü bu Şekil 15.10'daki eğriye neden olmaktadır. Şekil 15.11'de gördüğünüz regresyon doğrusu hayatta kalma puanına karşı çıktıların *sürekli* ölçümünün doğrusal regresyonudur ve bu çizgi her bir dağılımın ortalamasından geçmektedir. Sürekli bir ölçüm olduğunda, bu çizgi çözülebilir. Ancak sadece iki kategorili hale getirilmiş değerler için elde edilen veri için Şekil 15.10'da görülen iki kıvrımlı bir eğriye ait çözümler daha iyi olmaktadır.

Hipotetik sürekli verinin hayatta kalma puanının doğrusal bir fonksiyonu olduğu görülmesine rağmen, iki kategorili hale getirilmiş veri (veya iyileşme olasılığı) için doğrusal bir fonksiyonu olmadığı görülmektedir. *p* (iyileşme)den olasılık (iyileşme)ye ve logolasılı (iyileşme) oranlarına basit bir dönüştürmeyle hayatta kalma puanının doğrusal fonksiyonu olan bir değişken elde edilecektir. Bundan dolayı, *p* (iyileşme) logolasılık oranına (iyileşme) dönüştürülerek doğrusal fonksiyona geri dönülebilir. Allison (1999), burada yapılan işlemleri anlamaya yardımcı olabilecek iyi bir kaynaktır. Bu kılavuz SAS programı için yazılmış olmasına rağmen, SAS programı kullanılsın veya kullanılmasın, bildiğim, bu konuda en iyi ve yararlı açıklamaların bulunduğu kaynaktır.

Devam etmeden önce, odds ve log odds kavramlarını hatırlatmanın yararlı olacağı düşünülmektedir. Bölüm 6'dan bu yana bu kavramlara yer verilmemiştir. Dabbs ve Morris (1990) askerleri testosteron seviyesine göre yüksek ya da normal ve daha önce bir suça karışma durumuna göre evet ya da hayır şeklinde sınıflandırdıkları bir çalışma yürütmüşlerdir. Bu sınıflandırma işleminde aşağıdaki sonuçlar elde edilmiştir.

| | | Suça Karışma | | |
|---|---|---|---|---|
| | | Evet | Hayır | Toplam |
| **Testesteron** | Normal | 402 | 3614 | 4016 |
| | Yüksek | 101 | 345 | 446 |
| | | 503 | 3959 | 4462 |

© Cengage Learning 2013

Bu veri kümesi için, testosteron seviyesi açısından normal grupta yer alan bir askerin suça karışmış olmasının olasılık değeri (suça karışma frekansı)/(suça karışmama frekansı) olur. (Frekanslar yerine olasılık değerleri kullanıldığında bu oran *p(*suça karışma) / *p(*suça karışmama = *p*(suça karışma) / (1-*p*(suça karışma)) şekline dönüşür.) Normal testeron grubu için suça karışmış olmanın olasılık değeri 402/3614 = ,1001 olur. Normal testosteron grubu için suça karışmamış olmanın olasılık değeri ise bunun tersi olan 3614/402 = 8,990'dır. Bu son istatistik, normal testosterona sahip bir erkeğin suça karışmamış olmasının suça karışmış olmasına göre 9 kat daha muhtemel olduğu şeklinde yorumlanabilir (ya

da suça karışmamış olma olasılığı, suça karışmamış olma olasılığına göre 9 kat düşüktür.) Yüksek testosteron grubuna bakılırsa; suça karışmış olmanın olasılık değeri 101/345 = 0,293 ve suçu karışmamış olmanın olasılık değeri 345/101 = 3,416'dır. Her iki grupta da erkeklerin suça karışmamış olma olasılığı suça karışmış olma olasılığına göre daha düşüktür. Ancak birçok insanın suça karışmamış olması bekleneceğinden, bu sonuç çok bir şey söylememektedir. Testosteron seviyesi normal olan kişilerden yüksek olan kişilere doğru gidildikçe suça karışmış olmaya ilişkin olasılık değeri, 0,111'den 0,293'e yükselerek yaklaşık üç katına çıkmaktadır. Bu olasılık değerlerinin oranı alınırsa, 0,293/0,111 = 2,64 olur. Bu olasılık oranı; yüksek testosterona sahip bir kişinin normal testosterona sahip birine göre suçu karışma olasılığının 2,64 katı olduğu anlamına gelir. Bu oldukça etkileyici bir istatistiktir.

Bir süre için olasılık oranını bir kenara bırakıp sadece olasılık değerine bakalım. Hafızamızı yoklarsak, kanser hastalarına ilişkin araştırmada, hayatta kalmanın olasılık değerine odaklanıldığı hatırlanacaktır. (Hayatta kalma puanının iki farklı düzeyinin her biri için hayatta kalmanın olasılık değeri ile hayatta kalmamanın olasılık değerlerinin oranı alınarak, olasılık oranına dönüştürülebilir.)

Meme kanserinde hayatta kalma olasılığını kestirmek için olasılık değerinin *doğal logaritması*[17] üzerinde çalışılmaktadır. Elde edilen sonuç, hayatta kalmanın log olasılık değeri olarak adlandırılır. Testosteron seviyesi ile suça karışma arasındaki ilişkinin incelendiği çalışma dikkate alındığında, yüksek testosterona sahip bir erkek için suça karışmış olmanın log odss değeri,

$$\text{log olasılık} = \log_e(\text{olasılık}) = \ln(\text{olasılık}) = \ln(0{,}293) = -0{,}228$$

olur. Log olasılık değeri, 1'den büyük olasılık değerleri için pozitif; 1'den küçük olasılık değerleri için negatiftir (olasılık değeri 0 olduğunda ise log olasılık tanımsızdır). Bazen log olasılık değerinin **lojit** olarak adlandırıldığı görülmekte ve log olasılık dönüştürmesi için **lojit dönüştürme** ifadesi kullanılmaktadır.

**lojit**
**lojit dönrüştürme**

Kanser hastalığına ilişkin çalışmaya dönüldüğünde, Hayatta kalma puanına dayalı olarak Sonuçların tahmini ile başlanacaktır. $p$ = iyileşme görülme olasılığı ve 1-$p$ iyileşme görülmeme olasılığı olmak üzere, aşağıdaki eşitlik çözülecektir.

$$\log(p/1 - p) = \text{log olasılık} = b_0 + b_1 \text{ Hayatta kalma (SurvRate)}$$

Eşitlikteki $b_1$, Hayatta kalma (SurvRate) değişkenindeki bir birimlik artışa karşılık *log olasılık* değerindeki bir birimlik artışı gösterir. Burada verinin nasıl kodlandığına dikkat edilmesi oldukça önemlidir. Sonuç değişkeni için "1 = iyileşme" ve "2 = değişikliğin olmaması veya kötüye gitme" şeklinde kodlanmıştır. Hayatta kalma puanı için yüksek puanlar hastalığın sonucuna ait daha iyi bir kestirimi temsil etmektedir. Dolayısıyla, Hayatta kalma değişkenin daha iyi sonuçlar ile ilişkili olarak pozitif bir katsayıya sahip olması beklenir. Ancak SPSS'te böyle bir durum olmayacaktır. SPSS, Sonuçlar için 1 ve 2 şeklindeki kodlamayı "0 ve 1" şekline dönüştürüp daha sonra 0'a (daha iyi olma) ilişkin kestirimde bulunmaya çalışacaktır. Dolayısıyla katsayısı negatif olacaktır. (SAS, 1'i tahmin etmeye çalışır. Dolayısıyla SAS'ta elde edilen değer pozitif olacaktır. Yani, SAS ve SPSS'te rapor edilen katsayılar mutlak değerce eşit olmakla birlikte, SPSS'te negatif; SAS'ta ise pozitiftir.)

Basit doğrusal regresyonda, kalem kâğıt yardımıyla $b_0$ ve $b_1$ değerleri hesaplanabilmekte ve en küçük kareler yöntemi kullanılarak regresyon eşitliği çözülebilmektedir. Lojistik regresyonda ise işlemler basit bir biçimde yürütülemez. Hayatta kalma değişkeni için veri 0 ve 1 şeklinde olup iyileşmenin koşullu bir oranı değildir. Lojistik regresyonda en küçük kareler yöntemi yerine, en yüksek olabilirlik (maximum likelihood) yöntemi kullanılmakta ve regresyon katsayıları **tekrarlamalı** (iteratively) olarak çözümlenmektedir. Bilgisayar programı; $b_0$ ve $b_1$ için bazı başlangıç değerleri ile analize başlamakta, tahmin edilen log olasılık değerlerinin veri ile ne kadar uyum gösterdiğine bakılmakta, katsayılara düzeltme uygulanmakta, katsayıların veri ile uyumu tekrar incelenmekte ve katsayılara uygulanacak düzeltmenin veri ile uyumu arttırmayacağı noktaya kadar bu işleme devam edilmektedir. Bu süreç karmaşık hesaplamalar içerdiğinden kâğıt kalem kullanarak yapılabilecek bir işlem değildir.

**tekrarlamalı (iteratively)**

---

17 *X*'in doğal logaritması *e* tabanındaki logaritmasıdır. *e*'nin yaklaşık değeri 2,71828 olup *e*'nin değeri sonlu sayıda rakam kullanılarak yazılamamaktadır.

Ayrıca basit doğrusal regresyonda, ilişkinin ve her bir yordayıcı değişkenin regresyon modeline katkısının manidarlığını test etmek için $F$ ve $t$ istatistiklerinden yararlanılmaktadır. Lojistik regresyonda da bu amaca hizmet eden bir istatistik mevcuttur. Bununla birlikte, lojistik regresyondaki manidarlık testinde $F$ veta $t$ testi yerine $\chi^2$ testi kullanılır.

Gösterim 15.14'te, Hayatta kalma süresi (SurvRate) değişkeninin, Sonuçların kestirilmesine yönelik tek yordayıcı olarak kullanılması durumuna dair analiz sonuçları görülmektedir. Basit bir örnek ile başlangıç yapabilmek için tek bir yordayıcı değişken kullanılmıştır. Kısa süre içerisinde, yordayıcı değişken sayısının daha fazla olduğu örneklere de yer verilecektir. Lojistik regresyon ile ilgili temel hususlar yordayıcı değişken sayısından bağımsız olduğundan, yordayıcı değişken sayısının arttırılması, üzerinde tartışılacak değişken sayısının dışında herhangi bir değişikliğe neden olmayacaktır.

Gösterim 15.4'teki istatistiklerin tamamının tartışılması, temel hususlardan uzaklaşılmasına sebep olacağından, bunların tümü tartışılmayacaktır. Gösterim 15.4'teki istatistikler ile ilgili daha derinlemesine bilgi edinmek için Darlington (1990), Hosmer ve Lemeshow (1989) ile Lunneborg (1994) kaynaklarına başvurulabilir. Burada, temel problemlerin ve yaklaşımların açıklanması amaçlanmaktadır.

Analiz çıktılarının ilk kısmı, çoklu regresyon analizi çıktılarına benzemektedir. Bu ilk kısımda, modelin (tüm yordayıcı değişkenler bir arada ele alındığında), bağımlı değişkeni şans ile gerçekleşmesi beklenen düzeyden daha iyi kestirip kestiremediği test edilmektedir. Ancak bu işlemde çoklu regresyonda $F$, lojistik regresyonda ise $\chi^2$ testi kullanılmaktadır.

Number of selected cases: 66
Number rejected because of missing data: 0
Number of cases included in the analysis: 66

Dependent Variable Encoding:

| Original Value | Internal Value |
|---|---|
| 1.00 | 0 |
| 2.00 | 1 |

Dependent Variable. OUTCOME Cancer Outcome

Beginning Block Number 0. Initial Log Likelihood Function

−2 Log Likelihood 77.345746

* Constant is included in the model.

Beginning Block Number 1. Method: Enter

Variable(s) Entered on Step Number
1. SURVRATE Survival Rating by Physician

| | |
|---|---|
| −2 Log Likelihood | 37.323 |
| Goodness of Fit | 57.235 |
| Cox & Snell – R^2 | .455 |
| Nagelkerke – R^2 | .659 |

| | Chi-Square | df | Significance |
|---|---|---|---|
| Model | 40.022 | 1 | .0000 |
| Block | 40.022 | 1 | .0000 |
| Step | 40.022 | 1 | .0000 |

---------- Variables in the Equation ----------

| Variable | B | S.E. | Wald | df | Sig | R | Exp(B) |
|---|---|---|---|---|---|---|---|
| SURVRATE | −.0812 | .0193 | 17.7558 | 1 | .0000 | −.4513 | .9220 |
| Constant | 2.6836 | .8113 | 10.9408 | 1 | .0009 | | |

SPSS çıktısından uyarlanmıştır.

**Gösterim 15.4** Kanser hastalığında hayatta kalma araştırması için lojistik analizi

Analiz, Block Number 0 ve –2 log Likelihood (Olabilirlik) satırı ile başlar. Bu noktada modelde yordayıcı değişken bulunmamaktadır ve -2 log *(olabilirlik)* değeri 77,345746'dır. Bu değer, verideki genel değişkenliğin bir ölçüsü olup varyans analizindeki $KT_{toplam}$ değerinin benzeri olarak düşünülebilir. –2log *L*, yordayıcı değişkenin bulunmadığı modelin veri ile ne kadar uyumlu olduğuna ilişkin bir $\chi^2$ testi olarak düşünülebilir. Bu $\chi^2$ değeri 77,3457'dir. Beklenildiği gibi, yordayıcı değişken içermeyen bu modelde uyum düşüktür. (Eğer uyum mükemmel olsaydı, $\chi$2 değeri 0,00 olurdu.)

SPSS'te bir sonraki blokta, Hayatta kalma süresi (SurvRate) tek yordayıcı değişken olarak analize eklenmektedir. Bu ekleme sonrasında 37,323 şeklinde bir başka "– 2 log (olabilirlik)" değeri elde edilmektedir. Bu değer, Hayatta kalma süresi (SurvRate) değişkeninin yordayıcı olarak analize girmesinden sonra kalan değişkenliği gösterir. Hayatta kalma süresi (SurvRate) değişkeni analize dâhil olmadan ve dahil olduktan sonra elde edilen –2log olabilirlik değerleri arasındaki fark (77,345 – 37,323 = 40,022) hayatta kalma puanı değişkenine bağlı olarak $\chi^2$ değerinde gözlenen azalmayı temsil eder. Tek bir yordayıcı değişken analize eklendiğinden, bu $\chi^2$ değerine ilişkin serbestlik derecesi 1 olarak değerlendirilir. Görüleceği gibi, manidarlık düzeyi ,000'dır. Buna göre, hayatta kalma puanı değişkeninin modele manidar katkı sağladığı söylenebilir. (Model, Blok ve Adım [step] olarak adlandırılan satırların bulunduğuna dikkat edilmelidir. Bir seferde tüm yordayıcı değişkenler (1) analize alındığında üçü aynı olmaktadır.)

Tablonun sonraki bölümü her bir yordayıcı değişkeni ve bu yordayıcı değişkenin testini içermektedir. (Burada sadece bir yordayıcı değişken bulunmaktadır - SurvRate.) Bu bölüme bakıldığında, en uygun lojistik regresyon eşitliği aşağıdaki gibi yazılabilir:

Log olasılık = –0,082 Hayatta kalma süresi (SurvRate) + 2,6836

Eşitlikte Hayatta kalma süresi (SurvRate) değişkenine ait katsayının negatif olması, doktor tarafından verilen hayatta kalma puanı arttıkça log odds'un düştüğünü gösterir. Bu bulgu, SPSS programının, bir hastanın durumunun kötüye gitme veya hayatını kaybetme olasılığını kestirmeye çalıştığını gösterir. Buna göre, doktor tarafından verilen hayatta kalma puanı arttıkça hastanın durumunun kötüye gitme olasılığının azalması beklenmektedir.

Wald istatistiği sonuçlarına göre, Wald's $\chi^2$ = 17,7558, *sd* = 1 ve $p <$ ,0001 olduğundan Hayatta kalma süresi (SurvRate) değişkeninin manidar bir yordayıcı olduğu söylenebilir. Dikkat edilirse Hayatta kalma süresi (SurvRate) değişkenine ait modelin tümündeki $\chi^2$ testi, yani -2log L ile Wald's $\chi^2$ testinin uyumsuz olduğu görülecektir. Hayatta kalma süresi (SurvRate) değişkeni tüm model olduğundan, her iki testin aynı şeyi söylemesi gerektiği düşünülebilir. Bu durum, tek bir yordayıcı değişken içeren standart doğrusal regresyonda, regresyon modelinin manidarlığını gösteren *F* testinin, *t* testinin karesine eşit olması ile aynıdır. Lojistik regresyonda söz konusu iki test arasındaki uyumsuzluk, iki testin $\chi^2$'nin farklı kestirimlerine dayanmasından kaynaklanır. Wald ölçütüne ilişkin çeşitli sorular gündeme gelmiş ve Hosmer ile Lemeshow (1989) bunun yerine olabilirlik oran testine [–2log *L* (olabilirlik)] güvenilmesini savunmuştur.

Lojistik regresyon eşitliğine bakıldığında, Hayatta kalma süresi (SurvRate) değişkenine ait katsayının -0,0812 olduğu görülmektedir. Bu değer; Hayatta kalma süresi (SurvRate) değişkenindeki bir birimlik artışın log olasılık değerinde 0,0812'lik bir kötüye gidişe yol açacağı şeklinde yorumlanabilir. Birçok araştırmacının log olasılık yerine; olasılık değeri ile çalışmak isteyeceğini düşünmekteyim. Bu dönüştürme kolay bir işlem olup basit bir şekilde katsayısının üssü alınabilir. Eğer, 9,99 dolardan daha fazla para ödediğiniz bir hesap makineniz varsa, bu hesap makinesinde büyük ihtimalle $e^x$ butonu olacaktır. -0,0812 değeri hesap makinesine girilip bu tuşa basıldığında, 0,9220 değeri elde edilecektir. Bu değer, Hayatta kalma süresi (SurvRate) değişkeni 1 birim arttırılırsa kötüye gitmeye ait olasılığın 0,920 ile çarpılacağına işaret eder. Basit bir örnek bunun ne anlama geldiğini açıklamaya yardımcı olacaktır.

Hayatta kalma puanı 40 olan bir hastayı düşünelim. Bu hastaya ait log olasılık değeri;

Log olasılık = –0,082(40) + 2,6836 = –0,5643

olacaktır. Eğer $e^{-0,5643}$ hesaplanırsa, 0,569 bulunur. Bu değer, bu hastanın kötüye gitmesinin iyileşmesine göre 0,569 kez daha muhtemel olduğu anlamına gelir.[18] Şimdi, hayatta kalma puanı 41 olan başka bir hastayı ele alalım. Bu hasta için tahmin edilen log olasılık değeri aşağıdaki gibi olacxaktır:

---

18 Eğer odds değerinden hoşlanmıyorsanız odss değerini olasılık değerine dönüştürebilirsiniz. Çünkü odds = $p/(1-p)$ ve $p$ = odds/(1 + odds) şeklindedir.

Log olasılık = -0,082 (41) + 2,6836 = –0,6455

Daha sonra, $e^{-0,6455}$ hesaplanırsa, 0,524 elde edilir. Dolayısıyla ikinci hastanın log olasılık değeri ilk hastanın log olasılık değerine göre; -0,6455 – (-0,5643) = -,0812 daha düşük ve $e^{-0,0812}$ = 0,9220 daha büyüktür (0,569 × 0,922 = 0,524). 0,922 çok büyük bir sayı olarak görünmeyebilir. Buna rağmen, Hayatta kalma puanı 1 birim artan bir kanser hastasının kötüye gitmesi olasılığının %7,8 daha düşük olduğu göz önüne alındığında, böyle bir sonucun kesinlikle küçümsenemeyecek bir değer olduğu ifade edilebilir.

Hayatta kalma süresi (SurvRate) değişkeninin etkisi, log olasılık yerine olasılık değeri açısından görülmek istenirse, bir hesap makinesi kullanarak üs alma işlemi uygulamak yeterlidir. Bununla birlikte, SPSS zaten bu işlemi yapıp sonuçlarda rapor ettiğinden, böyle bir zahmete girilmesine gerek yoktur. "Exp (B)" olarak adlandırılan son sütun *b*'nin üssünün alınmış değerini ($e^{-0,0812}$ = 0,9220) içermektedir

Hayatta kalma puanı, kanser hastalarının iyileşme gösterme durumlarını yordayan anlamlı bir yordayıcı olmakla birlikte her şeyi açıklayamamaktadır. Epping, Jordan, Compas ve Howell (1994), bazı davranışsal değişkenlerin, bireylerin kanser ile başa çıkmasında etkili olup olmadığının belirlenmesiyle ilgilenmişlerdir. Müdaheleci düşünceler (Intrusive thoughts-Intrus) bu kapsamda incelenen ilk değişkendir. (Kanser hastalığını düşünmekten kendilerini alamayan bu kişiler hastalık ile ilgili düşünceleri zihinlerinden uzaklaştıramazlar.) İlgilenilen bir diğer değişken Kaçınma (Avoid) davranışıdır. Kaçınma davranışı kanseri düşünmek istememe ve sorunla uğraşmaktan kaçınmayı yansıtır [Intrus ve Avoid değişkenleri "Olaylar Etkisi Ölçeği'nden" hesaplanan değişkenlerdir (Horowitz, Wilner ve Alvarez, 1979).]

Gösterim 15.5'te Hayatta kalma süresi (SurvRate), Intrus ve Avoid yordayıcı değişkenleri aracılığıyla Sonuçların kestirilmesine ait bulgular sunulmuştur. Bu bulgular incelendiğinde, genel modelin, şans ile gerçekleşmesi, beklenenden daha yüksek bir uyum verdiği görülmektedir. Hiçbir yordayıcı değişken olmadığında, - 2 log olabilirlik = 77,346 olmaktadır. Üç yordayıcı değişken modele eklendiğinde, -2 log olabilirlik değeri 31,650'ye düşmektedir. Bu iki değer arasında 77,346 – 31,650 = 45,696 fark vardır. Üç yordayıcı değişken bulunduğundan, *sd* = 3 olup bu fark istatistiksel olarak manidardır. Hayatta kalma süresi (SurvRate) yordayıcı değişkenin tek başına manidar bir yordayıcı olduğu bilindiğinden modelin genelinin manidar olması beklenir. Tablonun alt satırı incelendiğinde, Wald ki-kare testinin Hayatta kalma süresi (SurvRate) ve Kaçınma değişkenleri için manidar, Inturs değişkeni için ise manidar olmadığı görülmektedir. Bu sonuç, yüksek düzeyde kaçınma davranışını sergileyen bireylerin daha az kaçınma davranışı sergileyen bireyler kadar iyileşme göstermedikleri anlamına gelir (Wald Ki Kare = 4,3310, $p$ = ,0374).[19] Daha özel olarak; Kaçınma değişkeni için regresyon katsayısı 0,1618'dir. Bu katsayı; diğer iki yordayıcı değişken sabit tutulmak üzere Kaçınma değişkenindeki bir birimlik artışın hastanın durumunun kötüleşmesine ilişkin log olasılık değerinde 0,1618 birimlik bir artışa yol açtığı şeklinde yorumlanabilir. Bu değerin üssü alındığında, $e^{0,1618}$ = 1,1756 bulunur. Buna göre, Kaçınma değişkenindeki bir birimlik artış hastanın durumunun kötüye gitmesine ilişkin olasılık değerini 1,1756 katı kadar arttırmaktadır.

Intrus yordayıcı değişkenine ait Wald ki kare testi manidar değildir (Wald $\chi 2$ = 0,5281, $p$ = ,4674). Dolayısıyla bu değişkenin modele manidar bir katkı sağlamadığı söylenebilir. Buna bağlı olarak, Sonuçta bağımlı değişkenin yordanmasına manidar katkı sağlamayan bu yordayıcı değişken modelden çıkarılmalıdır. Aslında bunu yapmak için oldukça iyi bir gerekçe vardır. Hatırlanacağı gibi, tek bir yordayıcı değişken olduğunda, -2 log $L$ olabilirlik ile verilen $\chi^2$ değeri 40,022 elde edilmişti. İki yordayıcı değişkenin daha eklendiği modelde ise $\chi^2$ değeri 45,696 olarak hesaplanmıştır. Burada hoş olan, bu $\chi^2$ değerleri arasındaki farkın, iki modelin yordayıcı değişken sayıları arasındaki farka karşılık gelen serbestlik

19 Hosmer ve Lemeshow'un (1989) Wald ki kare istatistiğinin geçerliği ile ilgili kaygılarına uygun olarak, bu teste ilişkin sonuçlara biraz temkinli yaklaşılmıştır. Wald istatistiği sıkı bir test olmakla birlikte, bu etkiye ilişkin yüksek güven muhtemelen yanlış değildir. Kısa süre içerisinde bu ifadeye ilişkin birtakım doğrulamalar görülecektir.

Dependent Variable Encoding:

| Original Value | Internal Value |
|---|---|
| 1.00 | 0 |
| 2.00 | 1 |

Dependent Variable. OUTCOME Cancer Outcome

Beginning Block Number 0. Initial Log Likelihood Function

−2 Log Likelihood 77.345746

* Constant is included in the model.

Beginning Block Number 1. Method: Enter

Variable(s) Entered on Step Number
1. SURVRATE Survival Rating by Physician
INTRUS
AVOID

| | |
|---|---|
| −2 Log Likelihood | 31.650 |
| Goodness of Fit | 35.350 |
| Cox & Snell – R^2 | .500 |
| Nagelkerke – R^2 | .724 |

| | Chi-Square | df | Significance |
|---|---|---|---|
| Model | 45.696 | 3 | .0000 |
| Block | 45.696 | 3 | .0000 |
| Step | 45.696 | 3 | .0000 |

---------- Variables in the Equation ----------

| Variable | B | S.E. | Wald | df | Sig | R | Exp(B) |
|---|---|---|---|---|---|---|---|
| SURVRATE | −.0817 | .0211 | 14.9502 | 1 | .0001 | −.4092 | .9215 |
| INTRUS | −.0589 | .0811 | .5281 | 1 | .4674 | .0000 | .9428 |
| AVOID | .1618 | .0777 | 4.3310 | 1 | .0374 | .1736 | 1.1756 |
| Constant | 1.6109 | 1.1780 | 1.8700 | 1 | .1715 | | |

SPSS, Inc.'dan uyarlanmıştır.

**Gösterim 15.5** SurvRate, Intrus ve Avoid değişkenlerinin bir fonksiyonu olarak Sonuçlar

derecesinde $\chi^2$ dağılımı göstermesidir. Bu iki modeli, 45,696 – 40,022 = 5,674 işlemi sonucunu 3 – 1 = 2 *sd*'den $\chi^2$ ile test edebileceğimizi gösterir. Bu 2 serbestlik derecesinde kritik $\chi^2$ değeri 5,99'dur. Bu nedenle, eklenen iki yordayıcı değişkenin, modeli manidar düzeyde geliştirmediği söylenebilir. İki model arasındaki fark, yararlı olmayan herhangi iki yordayıcının modele eklenmesi durumunda oluşması beklenen farktan daha yüksek değildir. Ancak, Hayatta kalma süresi (SurvRate) değişkeninin yanı sıra Avoid değişkeninin de manidar olduğu bilinmektedir. Dolayısıyla, yanlış gidenin ne olduğu sorusu akıllara gelmektedir.

Avoid değişkenin eklenmesi modele katkı sağlamıştır. Ancak Kaçınma ile birlikte Intrus değişkenin de modele eklenmesi Kaçınma değişkeninin modele sağladığı katkıyı istatistiksel açıdan anlamsız hale getirmiştir. Diğer bir deyişle, Intrus değişkeninin modele girmesi Avoid değişkeninin modele katkısına seyreltici bir etki yapmıştır. Eğer amaç, Intrus değişkenini içeren modeli test etmekten çok Sonuçları kestirmekse, sadece Kaçınma değişkenin modele eklenmesi durumunda daha iyi bir kestirim yapılabilir. Intrus değişkeni analizden çıkarılıp hayatta kalma süresi (SurvRate) ve Kaçınma değişkenleri ile analiz uygulanırsa aşağıdaki eşitlik elde edilir:

Log olasılık = –0,0823 Hayatta kalma + 0,1325 Kaçınma + 1,1961

Bu eşitlikte Hayatta kalma süresi (SurvRate) ve Kaçınma değişkenlerinin modele katkısı manidardır. Her iki yordayıcının da manidar olması, Kaçınma değişkeninin analize eklenmesinin sadece bir yordayıcının yer aldığı analize kıyasla modeli manidar düzeyde geliştirdiğine işaret eder. Nitekim Wald Ki Kare testinin manidarlığı, Kaçınma değişkeninin, modelin gelişmesine hizmet eden bir yordayıcı olduğunu doğrulamaktadır.

Veri kümesinin doğasından dolayı buradaki örnek sadece sürekli yordayıcı değişkenler içermektedir. Bununla birlikte, iki kategorili yordayıcı değişkenlerin kullanılmasına yönelik bir engel yoktur. Gerçekte de lojistik regresyonun kullanıldığı birçok durumda veri kümesi iki kategorili yapıdadır. İki kategorili yordayıcı değişken kullanılmasının iyi tarafı, yordayıcı değişkendeki bir birimlik değişimin bir kategoriden diğerine kaymayı göstermesidir. Örneğin, cinsiyet değişkeninin bir yordayıcı olarak kullanılması durumunda, Erkek = 1 ve Kadın = 2 olarak kodlanır. Cinsiyet değişkenindeki bir birimlik artış, erkek kategorisinden kadın kategorisine kayma anlamına gelir. Cinsiyet değişkeni için üstel katsayılar, kadın ve erkekler arasındaki oran farkını temsil eder. Cinsiyet değişkeninin kanser hastalığına ilişkin araştırmada bir yordayıcı olarak kullanıldığını ve katsayısının 0,40 olduğunu varsayalım.[20] Bu katsayıya ait $e^{0,40}$ değeri 1,49'dur. Bu değere göre, diğer tüm değişkenler sabit tutulduğunda, kadınlara ilişkin oran erkeklere ilişkin orandan yaklaşık 1,5 kat daha fazla iyileşme sağlamaktadır. "Düzenli spor yapan kişiler spor yapmayan kişilere oranla kalp rahatsızlıklarına %44 gibi bir oranla daha az yakalanmaktadır" ifadesiyle sıklıkla karşılaşılmaktadır. Bu tür ifadeler genellikle, burada tartışılan türdeki araştırma bulgularına dayalıdır.

Bu kısa girişte, lojistik regresyon analizine daha çok bir başlangıç yapabildim. Lojistik regresyon hakkında söylenebilecek çok daha fazla şey vardır. 0 ve 1 veya olasılık değerleri hakkında düşünüldüğünde oran (odds) ve log (odds) oran dönüşümü, yaşanan önemli güçlükler arasındadır. Bu bölümdeki temel amacım, bu güçlüğün aşılmasına yardımcı olmak ve iki kategorili bir bağımlı değişken olduğunda diskriminant analizi ve doğrusal regresyon yerine lojistik regresyon kullanılmasının tercih edilmesine dair kanıtlar sunmaktır. Lojistik regresyon hakkında diğer söylenebilecekler, teknik özellikleri hakkında olup bunlar birçok çalışmada mevcuttur. Özellikle, Allison (1999), Hosmer ve Lemeshow (1989) ile Kleinbaum ve Klein (2002) bu teknik özellikler hakkında bilgi edinmek için başvurulabilecek kaynaklar arasındadır.

## Anahtar Terimler

Regresyon katsayıları (15.1)
Artıklar (15.1)
Tolerans
Varyans bükülme faktörü (15.2)
Eş doğrusallık (15.2)
Çoklu doğrusallık (15.2)
Tekil (15.2)
Standartlaştırılmış regresyon katsayıları (15.2)
Artık varyans (15.5)
Artık hata (15.5)
Çok değişkenli normallik (15.6)
Çoklu korelasyon katsayısı ($R_{0.123...p}$) (15.7)
Kısmi korelasyon ($r_{0.12}$) (15.8)
Yarı kısmi korelasyon ($r_{0.12}$) (15.8)
Venn diyagramı (15.8)
Bastırıcı değişken (15.9
Çok değişkenli uç değerler (15.10)
Uzaklık (15.10)
Baskı (Leverage) ($h_i$) (15.10)
Etki (15.10)
Cook uzaklığı (15.10)
Studentized artık (15.10)
Tolerans (15.10)
Çapraz korelasyon (15.10)
Tekil (15.10)
İç-içe modeller (15.10)
Hiyerarşik modeller (15.10)
Akaike Bilgi Kriteri (ABK) (15.10)
Tüm alt gruplar regresyonu (15.11)
Değişken eleme (15.11)
Adımsal regresyon (15.11)
Değişken ekleme (15.11)
Çapraz doğrulama (15.11)
Liste bazında silme (15.11)
Durum bazında silme (15.11)
Çiftler bazında silme (15.11)
Ek veri konulması (Imputing) (15.11)
Mediating ilişkiler (15.14)
Moderating ilişkiler (15.14)
Ortalanma (15.14)
Lojistik regresyon (15.15
Diskriminant analizi (15.15)
Koşullu ortalamalar (15.15)
Çift kıvrımlı (Sigmodial) ortalamalar (15.15)
Çift kıvrımlı eğriler (15.15)
Sonradan sınıflandırılmış (Censored) veri (15.15)
Lojit (15.15)
Lojit dönüşüm (15.15)
Tekrarlı (İteratively) (15.15)

20 Bu çalışma meme konserine ilişkin bir araştırma olduğundan, cinsiyet değişkeni burada mantıklı bir yordayıcı değildir. Ancak, akciğer kanseri gibi bir kanser türü üzerine yapılan bir çalışmada cinsiyet değişkeni mantıklı bir yordayıcı olabilir.

## Alıştırmalar

Not: Yerden kazanmak ve hesaplamaları kolaylaştırmak amacıyla burada sunulan örneklerin büyük çoğunluğu küçük bir veri kümesine dayalıdır. Çoklu regresyon analizinin gerçek uygulamaları için örneklem büyüklükleri burada kullanılanlara kıyasla çok daha büyük olmalıdır.

**15.1** Bir psikolog şehirlerde algılanan yaşam kalitesi üzerine bir çalışma yapmaktadır. Çalışmaya çok sayıda şehir dâhil edilmiştir ($N$ = 150). Çalışma kapsamında, ortalama sıcaklık (Sıcak), 1000 dolar orta gelir (Gelir), sosyal hizmetlere ilişkin kişi başı harcama (Shizmet) ve nüfus yoğunluğu (Nüfus) değişkenleri yordayıcı olarak işlem görmüştür. Regresyon eşitliği aşağıdaki gibidir:

[$\hat{Y}$ = 5,37 – 0,01 Sıcak + 0,05 Gelir + 0,003 Shizmet - 0,01Nüfus]

a. Regresyon eşitliğini katsayılar açısından yorumlayınız.

b. Ortalama sıcaklığın 55 derece, 12,000 doların orta gelir, sosyal hizmetlere ilişkin kişi başı harcamanın 500 $ ve nüfus yoğunluğunun apartman (blok) başına 200 kişi olduğu bir şehir bulunduğunu varsayalım. Bu şehirde algılanan yaşam kalitesine ilişkin kestirim ne olur?

c. Sosyal hizmetlere ilişkin kişi başı harcamanın 100$ olması dışında diğer tüm özellikler açısından aynı değerlere sahip olan farklı bir şehir için algılanan yaşam kalitesi kestirimi ne olur?

**15.2** Alıştırma 15.1 için

$\beta$ = [ -0,438 0,762 0,081 -0,132]

olduğunu varsayarak sonuçları yorumlayınız.

**15.3** Alıştırma 15.2'deki $\beta$ değerlerine karşılık gelen standart hatalar aşağıdaki gibi olduğuna göre;

[0,397 0,252 0,052 0,025]

Regresyon denkleminden bir değişken çıkarılmak istenirse, eşitlikten çıkması en muhtemel değişken hangisi olur?

**15.4** Büyük bir şirket, çalışanlarının iş tatminlerinin kestirimi ile ilgilenmektedir. Bu kapsamda 15 çalışandan iş tatmini, sorumluluk düzeyi, denetimindeki personelin sayısı (DPS), çalışma ortamına ilişkin puan ve hizmet yılı değişkenlerine ilişkin veri toplanmıştır. Veri toplama işleminin ardından, aşağıdaki gibi bir veriye ulaşılmıştır.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| İş Tatmini | 2 | 2 | 3 | 3 | 5 | 5 | 6 | 6 | 6 | 7 | 8 | 8 | 8 | 9 | 9 |
| Sorumluluk | 4 | 2 | 3 | 6 | 2 | 8 | 4 | 5 | 8 | 8 | 9 | 6 | 3 | 7 | 9 |
| DPS | 5 | 3 | 4 | 7 | 4 | 8 | 6 | 5 | 9 | 8 | 9 | 3 | 6 | 9 | 9 |
| Ortam | 1 | 1 | 7 | 3 | 5 | 8 | 5 | 5 | 6 | 4 | 7 | 2 | 8 | 7 | 9 |
| Hizmet Yılı | 5 | 7 | 5 | 3 | 3 | 6 | 3 | 2 | 7 | 3 | 5 | 5 | 8 | 8 | 1 |

Gösterim 15.6, analiz çıktılarının kısaltılmış halidir.

a. Beş yordayıcı değişkeninin tümünü kullanarak regresyon eşitliğini yazınız.

b. $\beta_i$ değerleri nelerdir?

**15.5** Alıştırma 15.4 için

a. Yordanan değişken ile en yüksek yarı kısmi korelasyona sahip değişken hangisidir?

b. Genel $F$ değeri manidar olmamasına rağmen, çalışma ortamı değişkeni ile $Y$ manidar ilişki ($r$ = ,58) içerisindedir. Böyle bir sonuç nasıl mümkün olabilir?

**15.6** Alıştırma 15.4'teki veri kümesi için düzeltilmiş $R^2$ değerini hesaplayınız.

**15.7** Diğer tüm koşullar eşit olmak üzere, iki değişken arasındaki korelasyon azaldıkça bu değişkenlerin üçüncü bir değişkeni yordama gücü artar. Bu durumu, yarı kısmi korelasyon açısından açıklayınız.

**15.8** Diğer tüm koşullar eşit olmak üzere, herhangi bir regresyon katsayısının farklı örneklemler boyunca değişmezliği, kısmen bu değişken ile diğer yordayıcı değişkenler arasındaki ilişkinin bir fonksiyonudur. Bu durumu açıklayınız.

DEPENDENT VARIABLE . . . . . 1 SATIF
TOLERANCE . . . . . 0.0100

ALL DATA CONSIDERED AS A SINGLE GROUP

MULTIPLE R 0.6974 STD. ERROR OF EST. 2.0572
MULTIPLE R-SQUARE 0.4864

ANALYSIS OF VARIANCE

| | SUM OF SQUARES | DF | MEAN SQUARE | F RATIO | P(TAIL) |
|---|---|---|---|---|---|
| REGRESSION | 40.078 | 4 | 10.020 | 2.367 | 0.12267 |
| RESIDUAL | 42.322 | 10 | 4.232 | | |

| VARIABLE | | COEFFICIENT | STD. ERROR | STD. REG COEFF | T | P(2 TAIL) | TOLERANCE |
|---|---|---|---|---|---|---|---|
| INTERCEPT | | 1.66926 | | | | | |
| RESPON | 2 | 0.60516 | 0.428 | 0.624 | 1.414 | 0.188 | 0.263940 |
| NUMSUP | 3 | −0.33399 | 0.537 | −0.311 | −0.622 | 0.548 | 0.205947 |
| ENVIR | 4 | 0.48552 | 0.276 | 0.514 | 1.758 | 0.109 | 0.600837 |
| YRS | 5 | 0.07023 | 0.262 | 0.063 | 0.268 | 0.794 | 0.919492 |

A SPSS, Inc.'dan uyarlanmışır.

**Gösterim 15.6** Alıştırma 15.4'teki regresyon analizi için çıktı

**15.9** Gösterim 15.6'daki Tolerans sütunu 15.7 ve 15.8 numaralı alıştırmaların cevaplanabilmesine ne gibi bir katkı sağlamaktadır?

**15.10** Alıştırma 15.4'teki veri kümesini kullanarak $\hat{Y}$ değerini elde ediniz ve $R_{0.1234} = r_{Y\hat{Y}}$ olduğunu gösteriniz.

**15.11** Alıştırma 15.10'da hesaplanan $Y$ ve $\hat{Y}$ değerlerini kullanarak;
$KO_{\text{artık}} = \Sigma(Y - \hat{Y})^2 / (N - p - 1)$ olduğunu gösteriniz.

**15.12** Aşağıdaki tesadüfi veri kümesini kullanarak bazı gözlemler veri kümesinden çıkarıldığında çoklu korelasyonda nasıl bir değişim meydana geldiğini gösteriniz (örneğin; ilk olarak 15 gözlem için çoklu korelasyonu hesaplayınız, daha sonra 10, 6, 5, 4 vb. gözlem için çoklu korelasyonu bulunuz).

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 5 | 0 | 5 | 9 | 4 | 8 | 3 | 7 | 0 | 4 | 7 | 1 | 4 | 7 | 9 |
| $X_1$ | 3 | 8 | 1 | 5 | 8 | 2 | 4 | 7 | 9 | 1 | 3 | 5 | 6 | 8 | 9 |
| $X_2$ | 7 | 6 | 4 | 3 | 1 | 9 | 7 | 5 | 3 | 1 | 8 | 6 | 0 | 3 | 7 |
| $X_3$ | 1 | 7 | 4 | 1 | 8 | 8 | 6 | 8 | 3 | 6 | 1 | 9 | 7 | 7 | 7 |
| $X_4$ | 3 | 6 | 0 | 5 | 1 | 3 | 5 | 9 | 1 | 1 | 7 | 4 | 2 | 0 | 9 |

**15.13** Alıştırma 15.12'deki veri kümesi için düzeltilmiş $R^2$ değerini hesaplayınız.

**15.14** Alıştırma 15.4'teki ilk üç değişkene bakınız.

a. Sorumluluk değişkenini kontrol altında tutup iş tatminini yordanan değişken ve denetimdeki personel sayısını yordayıcı değişken olarak alıp yarı kısmi korelasyon ile kısmi korelasyonun karesini hesaplamak için bir bilgisayar programı kullanınız.
b. Bu iki katsayıyı göstermek için bir Venn Şeması oluşturunuz.

**15.15** Alıştırma 15.4'teki ilk üç değişkene bakınız.

a. Şekil 15.1 ile karşılaştırılabilen bir şekil çiziniz.
b. Bu değişkenler için regresyon çözümünü elde edip, çizdiğiniz şekil ile ilişkilendiriniz.

**15.16** Vermont Eyaleti Sağlık Planlama açısından 10 bölgeye ayrılmıştır (bu bölgeler kabaca eyalet içerisindeki illere karşılık gelmektedir). Aşağıdaki veri kümesi 2500 gramın altında doğan bebeklerin yüzdesini ($Y$), 17 yaş ve altındaki bayanların doğurganlık oranını ($X_1$), 17 yaşından küçük veya 35 yaşından büyük bayanlar için riskli doğurganlık oranını ($X_2$), 12 yıldan daha az eğitim alan annelerin yüzdesini ($X_3$), evlenmemiş annelere ait doğum yüzdesini ($X_4$) ve üç aylık süreye kadar tıbbi bakım almayan annelerin yüzdesini ($X_5$) göstermektedir.

| $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| 6.1 | 22.8 | 43.0 | 23.8 | 9.2 | 6 |
| 7.1 | 28.7 | 55.3 | 24.8 | 12.0 | 10 |
| 7.4 | 29.7 | 48.5 | 23.9 | 10.4 | 5 |
| 6.3 | 18.3 | 38.8 | 16.6 | 9.8 | 4 |
| 6.5 | 21.1 | 46.2 | 19.6 | 9.8 | 5 |
| 5.7 | 21.2 | 39.9 | 21.4 | 7.7 | 6 |
| 6.6 | 22.2 | 43.1 | 20.7 | 10.9 | 7 |
| 8.1 | 22.3 | 48.5 | 21.8 | 9.5 | 5 |
| 6.3 | 21.8 | 40.0 | 20.6 | 11.6 | 7 |
| 6.9 | 31.2 | 56.7 | 25.2 | 11.6 | 9 |

© Cengage Learning 2013

Gösterim 15.7'de adımsal regresyon sunulmuştur. (Yerden tasarruf etmek amacıyla sadece ilk üç adım gösterilmiştir. Bu alıştırmanın amaçları doğrultusunda, kaygıya neden olacak istatistiksel manidarlığın yokluğuna izin verilmemiştir, endişelenmeyiniz.)

a. Ardışık adımlar için $R$ değerleri nelerdir?

b. Kısmi korelasyonun tanımından (Venn diyagramı açısından) 2. adımdaki $R^2$ değerinin, 1. adımdaki $R^2$ değeri ile 1. adımda listelenen kısmi korelasyonların -"EŞİTLİKTE YER ALMAYAN DEĞİŞKENLER"- bir fonksiyonu olduğunu gösteriniz.

STEP NO. 1

VARIABLE ENTERED 3 X2
MULTIPLE R 0.6215
MULTIPLE R-SQUARE 0.3862
ADJUSTED R-SQUARE 0.3095
STD. ERROR OF EST. 0.5797

ANALYSIS OF VARIANCE

| | SUM OF SQUARES | DF | MEAN SQUARE | F RATIO |
|---|---|---|---|---|
| REGRESSION | 1.6917006 | 1 | 1.691701 | 5.03 |
| RESIDUAL | 2.6882995 | 8 | 0.3360374 | |

VARIABLES IN EQUATION

| VARIABLE | | COEFFICIENT | STD. ERROR OF COEFF | STD. REG COEFF | TOLERANCE | F TO REMOVE | LEVEL |
|---|---|---|---|---|---|---|---|
| (Y-INTERCEPT | | 3.529 ) | | | | | |
| X2 | 3 | 0.069 | 0.031 | 0.621 | 1.00000 | 5.03 | 1 |

VARIABLES NOT IN EQUATION

| VARIABLE | | PARTIAL CORR. | TOLERANCE | F TO ENTER | LEVEL |
|---|---|---|---|---|---|
| X1 | 2 | –0.19730 | 0.25831 | 0.28 | 1 |
| X3 | 4 | –0.25039 | 0.43280 | 0.47 | 1 |
| X4 | 5 | 0.00688 | 0.69838 | 0.00 | 1 |
| X5 | 6 | –0.59063 | 0.58000 | 3.75 | 1 |

SPSS'den uyarlanmıştır.

STEP NO. 2

VARIABLE ENTERED 6 X5
MULTIPLE R 0.7748
MULTIPLE R-SQUARE 0.6003
ADJUSTED R-SQUARE 0.4862
STD. ERROR OF EST. 0.5001

*(devam ediyor)*

**Gösterim 15.7** Düşük doğum ağırlığı çalışması veri kümesi için adımsal regresyon

ANALYSIS OF VARIANCE

| | SUM OF SQUARES | DF | MEAN SQUARE | F RATIO |
|---|---|---|---|---|
| REGRESSION | 2.6294919 | 2 | 1.314746 | 5.26 |
| RESIDUAL | 1.7505082 | 7 | 0.2500726 | |

VARIABLES IN EQUATION

| VARIABLE | COEFFICIENT | STD. ERROR OF COEFF | STD. REG COEFF | TOLERANCE | F TO REMOVE | LEVEL |
|---|---|---|---|---|---|---|
| (Y-INTERCEPT | 2.949 ) | | | | | |
| X2 | 3 0.113 | 0.035 | 1.015 | 0.58000 | 10.47 | 1 |
| X5 | 6 –0.223 | 0.115 | –0.608 | 0.58000 | 3.75 | 1 |

VARIABLES NOT IN EQUATION

| VARIABLE | | PARTIAL CORR. | TOLERANCE | F TO ENTER | LEVEL |
|---|---|---|---|---|---|
| X1 | 2 | –0.09613 | 0.24739 | 0.06 | 1 |
| X3 | 4 | –0.05399 | 0.37826 | 0.02 | 1 |
| X4 | 5 | 0.41559 | 0.53416 | 1.25 | 1 |

STEP NO. 3

| | |
|---|---|
| VARIABLE ENTERED | 5 X4 |
| MULTIPLE R | 0.8181 |
| MULTIPLE R-SQUARE | 0.6694 |
| ADJUSTED R-SQUARE | 0.5041 |
| STD. ERROR OF EST. | 0.4913 |

ANALYSIS OF VARIANCE

| | SUM OF SQUARES | DF | MEAN SQUARE | F RATIO |
|---|---|---|---|---|
| REGRESSION | 2.9318295 | 3 | 0.9772765 | 4.05 |
| RESIDUAL | 1.4481706 | 6 | 0.2413618 | |

VARIABLES IN EQUATION

| VARIABLE | COEFFICIENT | STD. ERROR OF COEFF | STD. REG COEFF | TOLERANCE | F TO REMOVE | LEVEL |
|---|---|---|---|---|---|---|
| (Y-INTERCEPT | 1.830 ) | | | | | |
| X2 | 3 0.104 | 0.035 | 0.942 | 0.55484 | 8.93 | 1 |
| X4 | 5 0.190 | 0.170 | 0.359 | 0.53416 | 1.25 | 1 |
| X5 | 6 –0.294 | 0.130 | –0.799 | 0.44362 | 5.14 | 1 |

VARIABLES NOT IN EQUATION

| VARIABLE | | PARTIAL CORR. | TOLERANCE | F TO ENTER | LEVEL |
|---|---|---|---|---|---|
| X1 | 2 | –0.14937 | 0.24520 | 0.11 | 1 |
| X3 | 4 | 0.14753 | 0.31072 | 0.11 | 1 |

**Gösterim 15.7** *(Devamı)*

**15.17** Alıştırma 15.16'da, Vermont Sağlık Ofisi'ni ilgilendirdiği kadarıyla $R^*$ değerine hangi anlam yüklenebilir?

**15.18** Alıştırma 15.16'da beş yordayıcı değişken için hesaplanan düzeltilmiş $R^2$ değeri, üç yordayıcı değişken için hesaplanan düzeltilmiş $R^2$ değerinden daha düşüktür. Bu durumun nedenlerini açıklayınız.

**15.19** Alıştırma 15.16'daki veri kümesi beşinci yordayıcı değişken ile ölçüt değişken arasındaki korelasyon düşük ($r$ = ,05) olmasına rağmen, bu değişkenin regresyon analizinde manidar bir yordayıcı role sahip olmasının nedenini açıklayınız.

**15.20** Alıştırma 15.16'daki veri kümesi için $\hat{Y} = 1X_2 - 1X_4 - 3X_5$ eşitliğini hesaplayınız. Bu regresyon eşitliğinin en uygun regresyon eşitliği ile ne kadar iyi uyum gösterdiğini belirleyiniz. Regresyon eşitliğinin neden bu şekilde olması gerektiğini açıklayınız.

**15.21** Alıştırma 15.16'daki veri kümesi için "Üçüncü aydan önce tıbbî bakım alamayan anne sayısının azaltılması, düşük doğum ağırlıklı bebek oranını azaltmak için iyi bir yol olabilir." şeklinde bir çıkarımda bulunmanın ne derece güvenilir olacağını tartışınız.

**15.22** Baskı (Leverage), uzaklık ve etki değerlerini gösteren ve 10 gözlemden oluşan bir veri kümesi oluşturunuz. Standart bir regresyon programı kullanarak bu istatistiklere ilişkin ölçümleri elde ediniz.

**15.23** *X*'in büyük değerleriyle ilişkili olan *Y* değerlerine ait varyansın, *X*'in küçük değerleriyle ilişkili olan *Y* değerlerine ait varyanstan daha yüksek olduğu bir veri kümesi oluşturunuz. Daha sonra regresyon analizini gerçekleştirerek apsis ekseni üzerinde *X* ve ordinat ekseni üzerinde artık değerlerin bulunduğu bir grafik çiziniz. Bu grafikte nasıl bir örüntü ortaya çıktığını açıklayınız.

## Bilgisayar Alıştırmaları

**15.24** Depresyonun mevcut düzeyleri ile diğer değişkenler arasındaki ilişkileri incelemek için Mireault'un (1990) çalışmasına ait veri kümesini kullanınız. Bununla ilgili bilgi ekte mevcut olup söz konusu veri dosyası bu kitabın internet adresinde Mireault.dat adı ile kayıtlıdır. Depresyonun (DepressT)(1); başka kayıplara karşı algılanan savunmasızlık düzeyi (PVLoss(SavKayıp))(2), algılanan sosyal destek düzeyi ((SuppTotl)Destek)(3) ve çocukluk döneminde ebeveyninin kaybedildiği yaş (AgeAtLoss(Yaş)) değişkenlerinin bir fonksiyonu olarak düşünülmesi makul bir öneri olabilir. Burada özetlenen modeli değerlendirmek için bir istatistik paket programı kullanınız. (Sadece çocukluk döneminde ebeveynlerinden herhangi biri vefat eden katılımcılar 1. grupta bulunduğundan, analizler sadece bu grup ile sınırlandırılmalıdır.)

**15.25** Bir araştırmacı, algılanan savunmasızlığa ilişkin toplam puana (PVTotal(SavToplam)), PVLoss(SavKayıp) değişkeninin yanı sıra olası tüm değişkenleri dâhil etmek istemektedir. (Toplam puan; kazalar, hastalık ve yaşam tarzı ile ilgili problemlere karşı savunmasızlığı içermektedir.)

a. Alıştırma 15.24'teki değişkenlere PVT Total (SavToplam) değişkenini ekleyerek regresyon analizini uygulayınız.

b. PVTotal(SavToplam) değişkeninin regresyon analizine dâhil edilmesi $R^2$ değerinde ne gibi bir değişiklik meydana getirmiştir? Bu değişkenin regresyon analizine dahil edilmesinin PVLoss(SavKayıp) değişkeni için regresyon katsayılarının standart hataları üzerinde bir etkisi var mıdır? Eğer kullandığınız program tolerans ve varyans şişkinlik faktörünü (PİV) de rapor ederse, PVTotal(SavToplam) değişkeninin analize dâhil edilmesi bu istatistikler üzerinde nasıl bir etki yaratır?

c. PVTotal(SavToplam) değişkeninin analize eklenmesiyle ilgili nasıl bir sonuca varabilirsiniz?

**15.26** Alıştırma 15.24'te depresyonun; savunmasızlık, sosyal destek ve ebeveyn kaybının yaşandığı yaş değişkenlerinin bir fonksiyonu olarak tanımlandığı bir model önerilmiştir. Savunmasızlığın sosyal destek ve ebeveyn kaybının yaşandığı yaş değişkenleri ile açıklanması alternatif bir bakış açısı olabilir. (Eğer ebeveyn kaybı genç yaşta yaşandıysa ve algılanan sosyal destek düzeyi düşükse, gelecekteki kayıplara karşı kişinin kendini daha savunmasız hissetmesi muhtemel bir durumdur.)

a. Yukarıdaki açıklamalar doğrultusunda regresyon analizine ilişkin probleminizi yeniden düzenleyerek uygun analizleri gerçekleştiriniz. [PVLoss (SavKayıp), SuppTotl (Destek ve Yaş)] değişkenlerini kullanınız.

b. Analiz sonuçlarını yorumlayınız.

**15.27** Alıştırma 15.24 ve 15.26'da incelenen ilişkileri göstermek için bir diyagram çiziniz. Tahmin edilen ilişkileri göstermek için oklar kullanınız ve okların yanına standart regresyon katsayılarını yazınız. (Basit bir yol analizinin uygulanması gerekir.)

**15.28** Alıştırma 15.27'deki diyagrama dikkat edilirse, algılanan sosyal desteğin (Destek) depresyon (Depres) üzerinde hem doğrudan hem de dolaylı etkisinin olduğu görülür. Doğrudan etki, doğrudan sosyal destekten depresyona doğru ilerleyen ok ile gösterilmiştir. Sosyal desteğin SavKayıp değişkenini etkilemesi ve SavKayıp değişkeninin de Depres üzerinde etkili olması ise, algılanan sosyal desteğin depresyon üzerindeki dolaylı etkisidir. Bu doğrudan ve dolaylı etkileri yarı kısmî korelasyon katsayıları açısından açıklayınız.

**15.29** Regresyon geçerlik istatistiklerinin rapor edilmesi talimatını vererek, Alıştırma 15.24'teki analizleri tekrarlayınız.

a. Analiz sonucunda rapor edilen bu istatistiklere göre veri dosyası hakkında ne söylenebilir?

b. En yüksek etki değerine (genellikle *Cook D* değeri esas alınır) sahip olan katılımcıya ait verinin dosyadan silinmesinin veri kümesi üzerinde ne tür bir etki oluşturacağını belirleyiniz.

**15.30** Ölçümlerin güvenirliğinin regresyon çıktıları üzerindeki etkisinin incelenmesi yararlı olacaktır. Alıştırma 15.24'te, SavToplam değişkenine ilişkin ölçümler yeterli derecede güvenilirdir. Bununla birlikte, değişkenlere ait ölçümlerin güvenirliğinin regresyon çıktıları üzerindeki etkisini gösterebilmek için SavToplam değişkenine bir miktar tesadüfi hata eklenerek güvenirliği daha düşük olan yeni bir veri kümesi oluşturulabilir.

a. SavToplam değişkeni için UnrelLos adıyla güvenilir olmayan PVLoss(UnrelSav = PVLoss(SavKayıp) + 7,5 × random" ifadesiyle yeni bir değişken oluşturulur [Buradaki "random (tesadüfî)" tesadüfi sayı üreten bir fonksiyonudur, birçok istatistik programında bulunan "random number funcion". Kullanılan istatistik programında bu fonksiyonu yerine getirmek amacıyla başvurulan komutun tam karşılığını belirlemek için programa ait kullanma kılavuzuna bakılmalıdır. Burada, tesadüfi sayı fonksiyonunun $N(0, 1)$ evreninden örnekleneceği varsayılarak 7,5 çarpımını kullandım. 7,5 çarpımı, UnrelLos(UnrelSav) değişkeninin standart sapmasını %50 arttıracaktır (varyans toplam kuralına bakınız). Diğer sabit değerlerle oynamak isteyebilirsiniz.]

b. Şimdi PVLoss(SavKayıp) yerine PVLoss(UnrelSav) değişkenini koyarak regresyon analizini tekrarlayınız.

c. Bu yeni değişkenin depresyon yordayıcı değişkeninin başka kayıplara karşı algılanan savunmasızlık düzeyi ile tahmin edilmesi üzerinde nasıl bir etkisi vardır? Regresyon katsayıları, bu katsayıların standart hatası, regresyon katsayılarının manidarlığına ilişkin manidarlık testi ve regresyon eşitliğindeki diğer değişkenler için ne gibi değişiklikler olmuştur?

**15.31** Harass.dat adlı veri kümesi, Brooks ve Perot (1991) tarafından cinsel taciz ile ilgili yapılan araştırmaya ilişkin sonuçları doğrulamak için 343 gözlemin kısmen değiştirilmesiyle oluşturulmuş bir veri doyasıdır. Katılımcıların cinsel tacize maruz kalıp kalmama durumu, araştırmanın bağımlı değişkenidir. Araştırmadaki bağımsız değişkenler sırayla; yaş, medeni durum (1 = evli, 2 = bekâr), feminist görüş, davranış sıklığı, davranışın saldırganlığı ve davranışın rapor edilip edilmediğidir (0 = hayır, 1 = evet). Her bir değişken için yüksek değerler, ilgili özelliğin çokluğunu temsil etmektedir. Bir lojistik regresyon programı kullanarak bağımsız değişkenler temelinde bir katılımcının cinsel taciz yaşama olasılığını inceleyiniz.

**15.32** Alıştırma 15.31'i tekrarlayınız. Ancak, bu sefer sadece iki kategorili bir değişken olan medenî durum değişkenini yordayıcı olarak kullanınız. Cinsel taciz yaşamış ya da yaşamamış olma durumu ile medenî durum değişkeni için bir olasılık tablosu oluşturunuz, olasılık oranlarını hesaplayınız ve bu oranları lojistik regresyonun sonuçları ile karşılaştırınız. (Sonuç manidar çıkmayacaktır ama bu önemli değildir.)

**15.33** Davranışın sıklığı, davranışın rapor edilmesi olasılığıyla ilişkili bulunmamıştır. Bu beklenmedik bir sonuç olup bu sonucun olası nedenleri hakkında ne söylenebilir?

**15.34** Malcarne, Compas, Epping ve Howell (1995), kanser tanısı konulmasının hemen ardından ve izleyen dört ay içinde 126 kanser hastasını izlemişlerdir. İlk görüşmelerde (Time 1/ Zaman 1), araştırmacılar hastanın mevcut üzüntü düzeyine (Disstress1/Üzüntü1), kanseri sahip oldukları kişilik türüne isnat etme derecelerine (BlemPer/Kişilik), kanseri atfettikleri davranışlara (sigara içme, yüksek yağlı beslenme alışkanlığı gibi) ilişkin veri toplamışlardır. Dört aylık takip sonunda (Zaman2), araştırmacılar tekrar hastaların üzüntü düzeyine (Disstress2 (Üzüntü2)) ait veri toplamışlardır. (Çalışmada, burada ilgilenilmeyen başka değişkenlere ait veri de toplamışlardır, burada onlarla ilgilenmeyeceğiz.) Çalışmaya ait veri kümesine Malcarne.dat adı ile bu kitabın internet sayfasından erişilebilmektedir.

a. Disstress2 (Üzüntü2) değişkeni, Disstress1/Üzüntü1 ve Kişilik değişkenleri ile yordanırsa nasıl bir sonuç ortaya çıkar?

b. Disstress1 (Üzüntü1) değişkenin a seçeneğine dâhil edilmesinin nedeni nedir?

**15.35** Alıştırma 15.34'te, Kişilik değişkeninin ikinci zamandaki üzüntü (Disstress2/Üzüntü2) ile ilişkili olduğu bir veri kümesi vardır. Ancak Kişilik değişkeni, Üzüntü1 değişkeni ile birlikte analize alındığında Üzüntü2'nin yordanmasına manidar bir katkı sağlamamaktadır. Bu durumun kaynağı olan faktörler nelerdir?

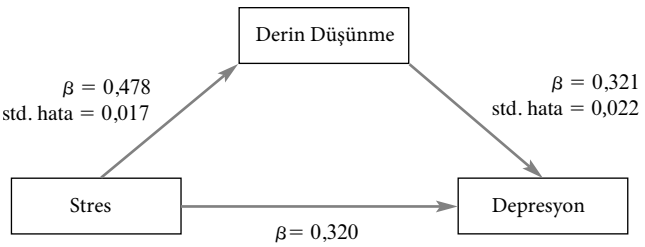
**15.36** Adımsal regresyonun (geniş bir şekilde tanımlanmış) birçok istatistikçinin tercih etmediği bir yöntem olmasının sebepleri nelerdir?

**15.37** İki yordayıcı değişken arasındaki etkileşimin etkisinin nasıl görülebileceğini örneklendirmek için basit değişkenler içeren sade bir örnek oluşturunuz.

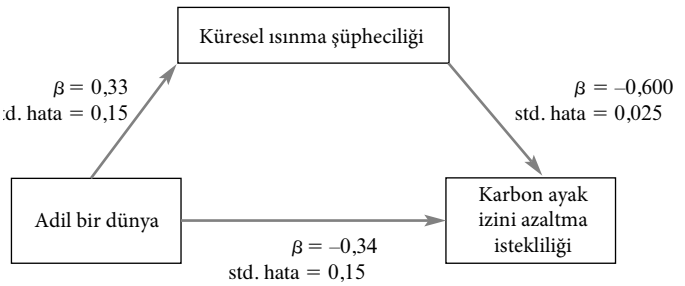
**15.38** Alıştırma 15.37'de oluşturduğunuz veriyi kullanarak, yordayıcı değişkenlerin "merkezlenmesinin (centering)" etkisini gösteriniz.

**15.39** Bildiğiniz gibi, regresyon katsayısı diğer tüm değişkenleri sabit tutarak bir değişkenin etkisi hakkında bilgi verir. Model içinde bir etkileşim terimi olması halinde, regresyon analizine ilişkin bu açıklama hakkında ne söylenebilir?

**15.40** Paul Jose'nin aracılık ilişkisi açıklanırken değinilen bir internet sitesi vardır. Paul Jose, stresin derin düşünceler değişkeninin aracı rolüyle depresyona neden oluğu şeklinde bir probleme odaklanmıştır. (Diğer bir deyişle, stresli bazı insanlar derin düşünceler içerisine girmekte ve bunun bir sonucu olarak da depresyon yaşamaktadırlar.) Bu ilişki için gerçek veri üzerinden üretilen yol diyagramı aşağıda görülmektedir. Stresten depresyona giden yol için verilen beta, depresyonun stres ve derin düşünceler değişkenleri ile yordanmasına yönelik çoklu regresyondur. Depresyonun, sadece stres değişkeni tarafından yordanmasına yönelik regresyon analizi için beta değeri 0,471'dir. Sobes'un testini kullanarak doğrudan stresten depresyona giden yol için hesaplanan katsayıdaki düşüşü test ediniz. (Cevaplar tam anlamıyla eşit olmayacaksa da; çalışmanızı Jose'nin internet adresine (http://www.victoria.ac.nz/psyc/paul-jose files/helpcentre/help7_mediation_example.php) bakarak kontrol edebilirsiniz.)




**5.41** Feinberg ve Willer (2011) adil dünya inancının (just world belief) insanların küresel ısınma kavramını kabul etmedeki isteklilikleri üzerine etkisini araştırmıştır. Çalışma kapsamında katılımcılara adil dünya inancını savunmaları ya da bu inançtan uzaklaşmaları şeklinde ikilemler sunulmuştur. Daha sonra yine aynı katılımcılara, karbon ayak izini (çevreye verilen zararın ölçümü amacıyla, belirli bir zaman içinde bir kişi, kurum veya kuruluş tarafından doğaya salınan karbon miktarı) azaltmak için küresel ısınma konusundaki şüphecilikleri ile ilgili bir anket uygulanmıştır Çalışmada, küresel ısınma şüpheciliği, adil dünya inancı ve karbon ayak izini azaltma konusundaki isteklilik değişkenleri için aşağıdaki gibi bir yol (path) diyagramı elde edilmiştir. (Not: Şüphecilikten istekliliğe giden yol için hesaplanan katsayılar adil dünya inancının etkisi açısından kontrol altına alınmıştır.) Hem adil dünya inancı hem de küresel ısınma şüpheciliği yordayıcı olarak kullanıldığında, adil dünya inancı ve karbon ayak izini azaltma konusundaki isteklilik arasındaki standart regresyon katsayısı, 0,13 değerinde bir standart hata ile -0,16'ya düşmektedir. Adil dünya inancı yüksek olan bireylerin şüphecilikleri adil dünya inancı düşük olan bireylerin şüpheciliklerine kıyasla daha yüksektir. Araştırmacılar korkuya dayalı söylemlerin mesajın etkililiğini azalttığını iddia etmektedir. Alıştırma 15.40'da yaptığınız gibi regresyon katsayılarındaki azaltmayı test ediniz ve uygun çıkarımlarda bulununuz.




© Cengage Learning 2013

**15.42** Bu bölümde, Guber'ın eğitim harcamaları çalışması üzerinde uzun süre durulmuş ve SAT'a katılan öğrenci yüzdesi değişkeni kontrol altına alındığında, eğitim harcamaları değişkeninin manidar bir yordayıcı olmadığı belirlenmiştir. Bununla birlikte, SAT, ülkedeki eğitimin kalitesini tartışmak için doğru bir gösterge değildir. Belki, ACT, okullarda verilen eğitimin niteliğine ilişkin daha doğru bir gösterge olabilir. Bu soruyu cevaplayabilmek için bu kitabın internet sayfasında Tab15-1.dat adı ile kayıtlı olan veri kümesini SPSS paket programından yararlanarak analiz ediniz. Her bir değişkeni ve artıkları inceleyerek cevabınızı tamamlayınız.



Bölüm 16

# Genel Doğrusal Modeller Olarak Varyans Analizi ve Kovaryans Analizi

## Hedefler

Varyans analizinin çoklu regresyonun özel bir durumu olarak görülebileceğinin gösterilmesi; eşit olmayan örneklem büyüklüklerine uygulanan işlemlerin tanıtılması ve kovaryans analizinin sunulması.

## Konular


16.1 Genel Doğrusal Model
16.2 Tek-Yönlü Varyans Analizi
16.3 Faktöriyel Desenler
16.4 Eşit Olmayan Örneklem Büyüklükleri ile Varyans Analizi
16.5 Tek Yönlü Kovaryans Analizi
16.6 Kovaryans Analizinde Etki Büyüklüğünün Hesaplanması
16.7 Kovaryans Analizi Sonuçlarının Yorumlanması
16.8 Kovaryans Analizi Sonuçlarının Raporlaştırılması
16.9 Faktöriyel Kovaryans Analizi
16.10 Çoklu Ortak Değişkenlerin Kullanılması
16.11 Alternatif Deneysel Desenler

PEK ÇOK ARAŞTIRMACI, çoklu regresyon ile varyans analizini farklı problemlere cevap bulmak amacıyla kullanılan, bütünüyle iki ayrı istatistiksel teknik olarak düşünmektedir. Gerçekte durum böyle değildir. En başta her ikisinde de aynı türden sorular sorulur, ikinci olarak farklı ifadelerle olsa da aynı türden cevaplar alınır. Varyans analizi, üç grubun ($T_1$, $T_2$ *ve* $T_3$) farklı ortalamalara ($\bar{X}_i$) sahip olduğunu açıklar; çoklu regresyon analizi ise, ortalamaların ($\bar{Y}_i$) işlemlerle ($T_1$, $T_2$ *ve* $T_3$) ilgili olduğunu ifade eder. Bu nedenle, çoklu regresyon analizi ve varyans analizi aynı kapıya çıkar. Ayrıca, varyans analizi ortalamalar arasındaki farklar üzerine $F$ istatistiği üretir. Regresyon analizi $R$'nin manidarlığı üzerine $F$ istatistiği üretir. Kısa süre içerisinde görüleceği gibi, varyans analizi ile çoklu regresyon analizinden elde edilen $F$ istatistikleri birbirine eşittir.

## 16.1 Genel Doğrusal Model

**genel doğrusal model**

Çoklu regresyon ve varyans analizinde cevap aranan araştırma problemleri benzer türde olduğundan, her iki istatistik aynı tekniğe dayalıdır. Aslında, varyans analizi, çoklu doğrusal regresyon modelinin özel formüllere dayalı özel bir halidir ve **genel doğrusal model** olarak adlandırılır. Varyans analizinin kendi eşitlikleriyle özel formüllere sahip olması bir şans olarak görülebilir. Belirli koşullar karşılandığında (varyans analizinde olduğu gibi), biraz zahmetli olan çoklu-regresyon hesaplamaları nispeten daha basit birkaç denkleme indirgenebilir. Eğer böyle olmasaydı, büyük ihtimalle varyans analizi olarak adlandırılan ayrı bir işlem kümesi olmayacaktı.

Yukarıdaki son ifade, sadece istatistiksel tekniklerin uygulanmasıyla ilgilenen öğrenciler için, bu bölüm de dahil olmak üzere, savunma amacıyla konulmuştur. Daha basit bir şekilde uygulanabilecek bildiğiniz işlemler varken, neden zahmetli işlemler ile uğraşıldığı sorusu sorulabilir. Akademik merak dışında, bu tür bir soruya verilecek çeşitli pratik (uygulamalı) cevaplar vardır. İlk olarak, bu yaklaşım eşit olmayan örneklem büyüklüklerinin ele alınmasına yönelik nispeten basit bir yolu temsil etmekte ve bu yaklaşımın anlaşılması, istatistiksel yazılımlardaki çeşitli seçenekler hakkında akıllıca kararlar verilmesine yardımcı olmaktadır. İkinci olarak, kovaryans analizinin anlaşılması ve uygulanmasına ilişkin cazip ve basit bir yol sunmaktadır - daha geleneksel yöntemler ile incelendiğinde, kullanışsız bir tekniktir. Sonuncusu ve belki de en önemlisi, istatistiksel tekniklerin o yönde geliştiği bir belirti gösterir. Güçlü ve hızlı bilgisayarların büyük ölçüde artan kullanımıyla daha önceleri pratik olmayan işlemler olarak görülen geleneksel istatistiklerin birçoğu, daha kolay bir biçimde gerçekleştirilebilir hale gelmiştir. Tekrarlı ölçümler varyans analizine karma model yaklaşımı düşündüğümüzde böyle bir örnek gördük. Yapısal eşitlik modeli ve faktör analizi diğer örnekler olarak gösterilebilir. Varyans analizi ve genel doğrusal modeller arasındaki ilişkiler (çoklu regresyonda sunulduğu gibi) ile basit varyans analizi problemleri için veri kümelerinin çoklu regresyon analizi çerçevesinde nasıl biçimlendirileceğinin anlaşılmaması, yakın gelecekte kendinizi daha az bildiğiniz hakkında daha fazla teknik kullanılırken bulmasına yol açacaktır. Bu, $t$, $\chi^2$, $F$ vb. tekniklerin yok olacağı anlamına gelmez. Ancak başka teknikler eklenecek ve eklenen yeni teknikler veri kümelerine farklı şekilde yaklaşılmasını sağlayacak yeni yollar açacaktır. Yakın zamanda, Yapısal Eşitlik Modeli'nin kullanılmasında görülen artış bu noktada bir olgudur. Çünkü bu modelin içerdiklerinin çoğu regresyon analizi hakkında bilinenler üzerine inşa edilmiş olup modelin altında yatan mantık ile ilgili öğrenilmesi gereken fazladan şey yoktur.

Son 25 yılda, bu genel konu hakkında oldukça mükemmel ve okunaklı çalışmalar yazılmıştır. Bununla birlikte, konu ile ilgili en açık sunum hâlâ Cohen'dır. (1968) Overall ve Spiegel (1969) tarafından yapılan çalışma da incelenmeye değerdir. Bu çalışmaların her ikisine de *Psychological Bulletin*'dan ulaşılabilir. Bu konudaki diğer başarılı tartışma-

lar için Overall (1972), Judd ve McClelland (1989) ile Cohen, Cohen, West, and Aiken (2003) kaynaklarına başvurulabilir. Ayrıca, Cramer ve Appelbaum (1980) ile Howell ve McConaughy (1982) altta yatan modelin seçimi ve izlenecek işlemler hakkında zıt görüşler sunmaktadır.

Bu bölümü okumak için her ikisi de akla uygun olan iki yol vardır. Bunlardan biri, genel kavramlara bakmak ve gerçek hesaplama teknikleri üzerine hafifçe gitmektir. Bu, öğrencilerimden genellikle izlemelerini islediğim yoldur. Öğrencilerimden istatiksel analizin arkasındaki mantığı anlamalarını isterim ve yapmak zorunda kaldıklarında tüm adımları gerçekleştirebileceklerini (kitap önlerindeyken) hissetmeleri sağlanmalı ancak teknik materyalleri ezberlemeleri talep edilmemelidir. Diğer yandan bazı öğretim elemanları, öğrencilerinden materyalleri daha derin bir düzeyde kavramasını isteyebilir. Bunu yapmak için bazı gerekçeler vardır. Ancak hâlâ bölüm hakkındaki kavramlara ilişkin genel ve derinlemesine bir anlayışa sahip olabilmek için ilk olarak bölümün okunması daha uygun bir yol olur. Daha büyük bir uzmanlık elde edebilmek için, bilgisayar başında oturup bir hesap makinesiyle çok daha fazla problem çözülmesi oldukça önemlidir.

## Doğrusal Model

İlk olarak, ($Y$) yordanan değişkeninin üç yordayıcı ($X_1$, $X_2$ ve $X_3$) değişkenleriyle kestirilmesine dair geleneksel çoklu regresyon problemini düşünelim. Alışık olduğumuz bu model

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + e_i$$

veya, *vektör* gösterimi kullnılarak

$$\mathbf{y} = \mathbf{b}_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3 + \mathbf{e}$$

ifadeleriyle yazılabilir. Bu eşitlikte, $\mathbf{y}$, $\mathbf{x}_1$, $\mathbf{x}_2$ ve $\mathbf{x}_3$ ($n \times 1$) biçiminde verinin vektörleri (sütunlar), $\mathbf{e}$ hata vektörü ($n \times 1$) biçiminde ve $\boldsymbol{b}_0$ ($n \times 1$) biçiminde bileşeni sabit olan bir vektördür. Bu eşitlik $\mathbf{X}$, yordayıcıların $n \times (p + 1)$ matrisi, birinci sütun 1'lerin sütunu ve $\mathbf{b}$ regresyon katsayılarının[1] $(p + 1) \times 1$ biçimindeki vektörü olup aşağıdaki gibi kısaltılabilir:

$$\mathbf{y} + \mathbf{Xb} + \mathbf{e}$$

$\boldsymbol{X}$ yordayıcı değişkenlerin doğrusal kombinasyonunun toplamı olduğundan bu model doğrusal model olarak adlandırılır - hiçbir şey 1'den daha büyük bir güce yükseltilemez.

Şimdi tek-yönlü varyans analizi için geleneksel modeli düşünelim:

$$Y_{ij} = \mu + \tau_j + e_{ij}$$

Bu eşitlikteki $\tau_j$ sembolü; araştırılan problemdeki herhangi bir işleme ilişkin herhangi bir katılımcı için $\tau_j$ değeri olan $\tau_1$, $\tau_2$, $\tau_3 \ldots \tau_p$ değerlerini yazmanın kısa yoludur. Bu model ile geleneksel regresyon modeli arasındaki ilişkiyi görmek için tasarım matrisi kavramının tanıtılması gerekir. Tasarım matrisi, sadece varyans analizinde değil çok değişik durumlarda kullanıldığından bu matrisin anlaşılması önemlidir.

## Tasarım Matrisi

**tasarım matrisi**

**Tasarım matrisi**, grup üyeliğini temsil eden *kodlanmış*, *kukla* veya *sayaç* değişkeni matrisidir. Tasarım matrisinin *tam formu* ($X$), ortalamaları temsil eden $p + 1$ sayıda sütun ve $p$ işlem (grup) etkisini içerir. $\mu$ tüm gözlemlerin bir parçası olduğundan, $\mu$ için bir gözlem daima 1

1 Bu bölümde sunulan birkaç eşitlik matris biçiminde ortaya konulmasına rağmen, sunulan konuların anlaşılması için matris cebir kurallarının bilinmesine gerek yoktur.

ile kodlanır. Diğer tüm sütunlarda, eğer bir katılımcı ilgili sütunun bir üyesi ise "1", değilse "0" ile kodlanır. Buna göre, her biri 2 gözlem içeren 3 grup için tam tasarım matrisi aşağıdaki gibi olur:

$$\sum = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccc} \textbf{\textit{Den.}} & \boldsymbol{\mu} & \boldsymbol{A_1} & \boldsymbol{A_2} & \boldsymbol{A_3} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Dikkat edilirse, 1'nci ve 2'nci gözlem ($A_1$ grubunda yer alan) $\mu$ ve $A_1$ sütunlarında "1", $A_2$ ve $A_3$ sütunlarında "0" ile kodlanmıştır. Çünkü bu grupların elemanı değildir. Benzer şekilde, 3'üncü ve 4'üncü gözlemler, μ ve $A_2$ sütunlarında "1" ile kodlanırken; $A_{2,}$, $A_1$ ve $A_3$ sütunlarında "0" ile kodlanmıştır.

Şimdi, [$\mu$ $\tau_1$ $\tau_2$ $\tau_3$] olarak işlem (grup) etkisine ait vektör $\tau$'yu tanımlayacağız. **X** tasarım matrisi olarak alındığında, varyans analizi modeli matris terimleri cinsinden aşağıdaki gibi yazılabilir:

$$\mathbf{y} = \mathbf{X}\tau + \mathbf{e}$$

Bu eşitlik geleneksel regresyon eşitliği ile aynı formda görülebilir. Regresyon eşitliğindeki her bir $b$ değerinin bağımsız değişkenin etkisi olması gibi, $\tau$ bileşeni her bir kukla grup değişkeninin etkisine karşılık gelir. Bu eşitliğin genişletilmesiyle;

$$\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ e_{12} \\ e_{22} \\ e_{13} \\ e_{23} \end{bmatrix}$$

$$\boldsymbol{y} = \quad \boldsymbol{X} \quad \times \quad \boldsymbol{\tau} \quad + \quad \boldsymbol{e}$$

olur, bu eşitliğe matris çarpım kuralı uygulandığında, aşağıdaki eşitliklere ulaşılır:

$$\begin{aligned} Y_{11} &= \mu + \tau_1 + e_{11} \\ Y_{21} &= \mu + \tau_1 + e_{21} \\ Y_{12} &= \mu + \tau_2 + e_{12} \\ Y_{22} &= \mu + \tau_2 + e_{22} \\ Y_{13} &= \mu + \tau_3 + e_{13} \\ Y_{23} &= \mu + \tau_3 + e_{23} \end{aligned}$$

Böylece her bir katılımcının her bir cevabı için bir model elde edilir. Buna göre, Grup 2'deki ikinci katılımcı için $Y_{22} = \mu + \tau_{22} + e_{22}$ ve İşlem $j$'deki $i$ numaralı katılımcı için $Y_{ij} = \mu + \tau_j + e_{ij}$ ifadesiyle genel bir varyans analizi modeli elde edilmiş olur.

Burada dikkat edilmesi gereken temel husus, tasarım matrisinin varyans analizini çoklu-regresyon analizi çerçevesinde görmeye imkân tanımasıdır. Yani; tasarım matrisi

$Y_{ij} = \mu + \tau_j + e_{ij}$ eşitliğinden $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ eşitliğine geçişe imkân tanır.

Ayrıca **b** bileşeni, $\mu$, $\tau_{1,}$ $\tau_{2,}$…. $\tau_{k}$ değerleridir. Bir diğer ifadeyle, ilgilendiğiniz gerçek işlem etkileridir.

Kullanageldiğimiz tasarım matrisi, üstesinden gelmemiz gereken bazı teknik sınırlılıkları içeir. Bundan dolayı tasarım matrisini, oldukça farklı görünen fakat aynı bilgileri taşıyan bir şekle dönüştüreceğiz. İlk olarak, eğer bir katılımcının $A_1$ ya da $A_2$'de olmadığı söylenirse, söylenmemiş olsa da bu katılımcının $A_3$'te yer aldığı anlaşılabildiğinden, bu sütun gereksizdir. Diğer bir deyişle, *işlemler* için sadece 2 serbestlik derecesi vardır. Bundan dolayı, $A_3$ başlıklı sütunu çıkararak işlem etkisine ait sadece $a$ - 1 sütun içeren bir matris elde kalır. (Çıkarılacak sütun, üç sütundan herhangi biri olabilir. Bununla birlikte, genelde son sütun çıkarılır.) Eğer, korelasyon matrisinin hesaplanması yoluyla çoklu regresyon denkleminin elde edilebildiği herhangi bir bilgisayar programı kullanılacaksa, ikinci bir değişiklik daha gerekecektir. $\mu$ başlıklı sütunun varyansı yoktur ve dolayısıyla korelasyon matrisine giremez - 0'a bölme gibi bir girişime neden olur. Dolayısıyla bu sütunun da çıkarılması gerekir. Son çözüm bu çıkarmadan etkilenmeyeceğinden, bu gerçek bir kayıp değildir. Aslında, çıkarılan bu sütun yazılım tarafından gizlice analize alınmaktadır.

Faktöriyel desenler için daha sonra savunulacak yöntemi kullanarak arzu edilen sıfır hipotezini test etmeye izin vermesi için basit bir biçimde başka bir değişiklik daha yapılacaktır. Çünkü, üçüncü (veya *a*'yıncı*)* işlemin düzeyleri ile ilgili sütun göz ardı edildiğinden, değiştirilen tasarım matrisi için verilen çözümler $\bar{X}$ yerine $\bar{X}_3$ ile ilgili işlem etkisine ait kestirimleri üretir. Diğer bir deyişle; $b_1$ ($\bar{X}_1 - \bar{X}.$) yerine ($\bar{X}_1 - \bar{X}_3$) haline gelir. Eğer istenilen bu ise iyi bir sonuçtur. Fakat işlem etkilerinin, daha çok genel ortalamadan sapma şeklinde görülmesini tercih ederim. Bu sadece daha düzenli bir görünüme sahiptir. Bu yüzden, **X**'in her bir sütununun ortalamasını ($\bar{X}$) 0'a eşit yapmak için tasarım matrisini değiştireceğiz. Bu yeni sistem altında, eğer herhangi bir katılımcı Grup $A_i$'nin bir üyesi ise $A_i$ sütununda "1" olarak kodlanır, eğer *a'* yıncı (son) sütunun bir üyesi ise "–1" ile kodlanır ve bu iki koşula da uymuyorsa "0" ile kodlanır. (Bu kısıtlama, sabit model varyans analizi için $\Sigma\tau_j = 0$ şeklindeki gerekliliğe karşılık gelir.) Bu değişiklikler bizi,

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ matrisinden } \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ matrisine } \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ buradan } \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix}$$

matrisine götürür. Bunlar, başlangıçtaki noktaya göre X'in son biçiminde büyük değişiklikler olarak görülmesine rağmen, aslında gerekli tüm bilgileri taşır. Sadece gereksiz bilgiler dışarıda tutulmakta, sabit terim çıkarılmakta ve sonra işlemin etkilerinin $\bar{X}$'dan sapması olarak verilmesine neden olur.

## 16.2 Tek-Yönlü Varyans Analizi

Bu noktada basit bir örnek sunulmuştur. Tablo 16.1'deki veri, her biri dört işleme tabi tutulan üç deneğe aittir. Tablo 16.1b, ilgili varyans analizi için etki değerinin ($\hat{\tau}_i$) (Bölüm 11'de tartışılmıştır) de yer aldığı özet tabloyu içermektedir. Tablo 16.1c kestirilen işlem etkisini ($\hat{\tau}_j$) içermekte olup burada, $\hat{\tau}_i = \hat{\mu}_i - \hat{\mu}$'dur. Sabit-model varyans analizi $\Sigma\tau_i = 0$ sınırlamasını uygulamaya koyduğundan, $\tau_4$ değeri $\tau_1$, $\tau_2$ ve $\tau_3$ aracılığıyla otomatik olarak tanımlanmaktadır ($\tau_4 = 0 - \Sigma\tau_j$).

**Tablo 16.1** Eşit örneklem büyüklüklerinde basit tek yönlü-desenler için örnek hesaplamalar

**(a) Veri**

| İşlem 1 | İşlem 2 | İşlem 3 | İşlem 4 |
|---|---|---|---|
| 8 | 5 | 3 | 6 |
| 9 | 7 | 4 | 4 |
| 7 | 3 | 1 | 9 |
| 8 | 5 | 2,667 | 6,333 |
| | | | $\overline{X} =$ 5,500 |

**(b) Özet Tablo**

| Kaynak | *sd* | *KD* | *KO* | *F* | $\eta^2$ |
|---|---|---|---|---|---|
| İşlemler | 3 | 45,667 | 15,222 | 4,46 | ,626 |
| Hata | 8 | 27,333 | 3,417 | | |
| Toplam | 11 | 73,000 | | | |

**(c) Kestirilen İşlem Etkileri**

$$\hat{\tau}_1 = \overline{X}_1 - \overline{X}_{..} = 8.0 - 5.5 = 2.5$$

$$\hat{\tau}_2 = \overline{X}_2 - \overline{X}_{..} = 5.0 - 5.5 = -0.5$$

$$\hat{\tau}_3 = \overline{X}_3 - \overline{X}_{..} = 2.67 - 5.5 = -2.83$$

© Cengage Learning 2013

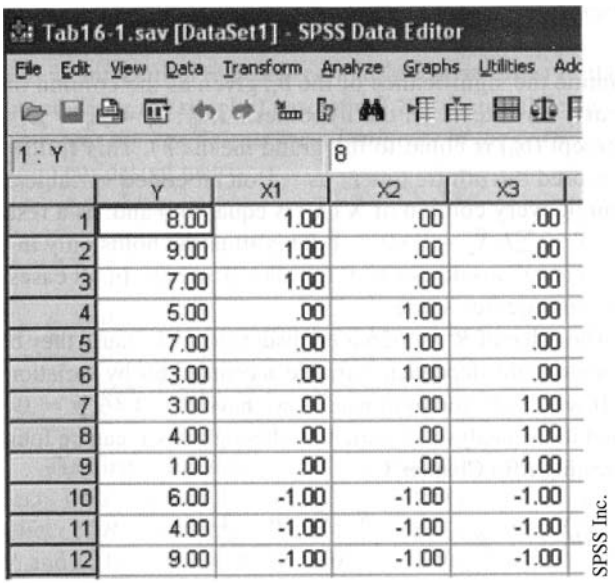
Şimdi, bu verinin istatistiksel durumuna en küçük-kareler çoklu doğrusal regresyon yöntemiyle yaklaşalım. Tablo 16.1'deki ham veride yordanan değişken (*Y*) olarak alınacaktır. Yordayıcı değişkenler için *formun* tasarım matrisi kullanılacaktır.

$$\boldsymbol{X} = \begin{array}{c} \\ \text{İşlem 1} \\ \text{İşlem 1} \\ \text{İşlem 1} \\ \text{İşlem 1} \end{array} \begin{array}{c} \begin{array}{ccc} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{array}$$

Burada, tasarım matrisinin herhangi bir satırındaki bileşenler *o işleme ait gruptaki tüm katılımcılara* uygulanmak için alınır. **X** tasarım matrisi yordayıcıların matrisi olarak kullanılarak yapılan çoklu-regresyon çözümü Gösterim 16.1'de verilmiştir. Burada bağımlı değişken (**Y**) veri matrisinin ilk sütunudur. Sonraki üç sütun birlikte **X** matrisini oluşturur. Bu çözümü elde etmek için SPSS paket programı kullanılmıştır. Ancak başka bir programdan da yararlanılabilir. (Tartışmayı kolaylaştırmak için analiz çıktılarında bazı küçük değişiklikler yapılmıştır.)

X değişkenleri arasındaki korelasyon örüntüleri Gösterim 16.1'de yer almaktadır. Sabit köşegen korelasyonlarıyla bu tür örüntüler, çeşitli gruplarda eşit sayıda katılımcının olduğu her durumda oluşur. (Farklı gruplarda eşit sayıda katılımcının olmadığı faktöriyel desenlerde, sabit bir köşegenin olmadığı gerçeği bu tür durumlarda uygulanacak istatistiksel işlemlerin zorlaşmasına neden olur.)

Dikkat edilirse, regresyon katsayıları bir sütun olarak yazılmaktadır. Bu sütun *vektör* olarak adlandırılabilir, vektör **b** veya varyans analizi terimlerinde vektör $\tau$. Görüleceği gibi burada, $b_1 = 2{,}50$'dir ve bu değer Tablo 16.1'de İşlem 1'in kestirilen etkisiyle aynıdır. Diğer bir deyişle, $b_1 = \tau_1$. Bu eşitlik $b_2$ ve $b_3$ için de geçerlidir. Bu gerçek mutlaka X ve




| | Y | X1 | X2 | X3 |
|---|---|---|---|---|
| 1 | 8.00 | 1.00 | .00 | .00 |
| 2 | 9.00 | 1.00 | .00 | .00 |
| 3 | 7.00 | 1.00 | .00 | .00 |
| 4 | 5.00 | .00 | 1.00 | .00 |
| 5 | 7.00 | .00 | 1.00 | .00 |
| 6 | 3.00 | .00 | 1.00 | .00 |
| 7 | 3.00 | .00 | .00 | 1.00 |
| 8 | 4.00 | .00 | .00 | 1.00 |
| 9 | 1.00 | .00 | .00 | 1.00 |
| 10 | 6.00 | -1.00 | -1.00 | -1.00 |
| 11 | 4.00 | -1.00 | -1.00 | -1.00 |
| 12 | 9.00 | -1.00 | -1.00 | -1.00 |

Korelasyonlar

| | | Y | X1 | X2 | X3 |
|---|---|---|---|---|---|
| Pearson Korelasyon | Y | 1,00 | ,239 | –,191 | –,526 |
| | X1 | ,239 | 1,000 | ,500 | ,500 |
| | X2 | –,191 | ,500 | 1,000 | ,500 |
| | X3 | –,526 | ,500 | ,500 | 1,000 |

Model Özeti

| Model | R | $R^2$ | Düzeltilmiş $R^2$ | Tahminin Standart Hatası |
|---|---|---|---|---|
| 1 | ,791 (a) | ,626 | ,485 | 1,848 |

[a] Yordayıcılar: (Sabit), X3, X2, X1

Katsayılar[a]

| Model | | Standartlaştırılmamış Katsayılar | | Standartlaştırılmış Katsayılar | t | Mnd. |
|---|---|---|---|---|---|---|
| | | B | Std. Hata | Beta | | |
| 1 | (Sabit) | 5,500 | ,534 | | 10,307 | ,000 |
| | X1 | 2,500 | ,924 | ,717 | 2,705 | ,027 |
| | X2 | –,500 | ,924 | –,143 | –,541 | ,603 |
| | X3 | –2,833 | ,924 | –,812 | –3,066 | ,015 |

[a] Bağımlı Değişken: Y

ANOVA[b]

| Model | | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|---|
| 1 | Regresyon | 45,667 | 3 | 15,222 | 4,455 | ,040 |
| | Artık | 27,333 | 8 | 3,417 | | |
| | Toplam | 73,000 | 11 | | | |

[a] Yordayıcılar: (Sabit), X3, X2, X1

[b] Bağımlı Değişken: Y

**Gösterim 16.1** Tablo 16.1'deki veri kümesi için regresyon analizi SPSS çıktıları

$\tau$ tanımını izler. Ayrıca, *t*-oranları sütunundaki $b_i$ değerlerinin manidarlığını incelersek, aynı zamanda ($H_0 : \tau_j = \mu_i - \mu = 0$) hipotezini test etmiş oluruz. Burada sabitin ($b_0$) genel ortalamaya ($\bar{Y}$) eşit olduğu da gözen kaçırılmamalıdır. Bu, tüm kodlanmış değişkenlerdeki, *a*'yıncı işlemin –1 olarak puanlandığı gerçeğini doğrudan izler. (–1) kodu kullanılarak, **X**'in her bir sütununun ortalaması ($\bar{X}_j$) 0'a eşit ve sonuç olarak $\Sigma b_1 \bar{X}_j = 0$ olur. Dolayısıyla, $b_0 = \bar{Y} - \Sigma b_1 \bar{X}_j = \bar{Y} - 0 = \bar{Y}$. Bu eşitlik asadece eşit örneklem büyülüklerinde sağlanır. Aksi durumlarda tüm *i*'ler için $\bar{X}$ 0'a eşit olmamaktadır. Bununla birlikte tüm durumlarda, $b_0$ en küçük kareler anlamında $\mu$ değerinin en iyi kestirimidir.

$R^2$ = ,626 olup bu değer varyans analizindeki $\eta^2$ değerine eşittir. Çünkü her ikisi de bağımlı değişkende bağımsız değişkenler tarafından açıklanan varyans miktarını gösterir.

Eğer $R^2$ değerinin manidarlığı test edilirse, $F$ = 4,46 ve $p$ = ,040 değeri elde edilir. Bu $F$ değeri varyans analizinde elde edilen $F$ değerine eşittir. Ayrıca bu $F$ değeri, Bölüm 15'te $R^2$ değerinin manidarlığını test etmek için verilen formül kullanılarak hesaplanabilir:

$$F(p, N-p-1) = \frac{R^2(N-p-1)}{(1-R^2)p}$$

$$F(3,8) = \frac{,626(8)}{,374(3)} = 4,46$$

Dikkat edilirse, Gösterim 16.1'de Regresyon için kareler toplamı, Hata ve Toplam değerleri, Tablo 16.1'deki, Gruplar arası kareler toplamı, Hata ve Toplam değerlerine eşittir. Bu eşitlik, regresyon analizi ile varyans analizindeki kareler toplamı arasında tam bir uyum olduğunu açık bir şekilde gösterir.

Yukarıdaki analizler varyans analizi ile regresyon analizi arasındaki benzerliği açıkça göstermektedir. Bu öncelikle, "farklı işlemler değişik ortalamalar üretir" ifadesi ile ortalamanın, yordayıcı değişkenlerin bir fonksiyonu olup olmadığı problemi arasında önemli bir fark olmadığının bir gösterimidir. Basit bir şekilde madalyonun farklı yüzlerine bakılmaktadır.

Burada, tasarım matrisi oluşturmanın sadece en yaygın yolu ele alınmıştır. Bu matris, yararlı olabilecek bazı formlar ileri sürebilir. Ancak burada diğer formlara değinilmeyecektir. Bu konu hakkında bilgi edinmek için Cohen'a (1968) başvurulması önerilir.

## 16.3 Faktöriyel Desenler

Regresyon analizini, iki yönlü veya daha üst-düzey faktöriyel desenlere kolaylıkla genişletebiliriz. Bu işlemin yapılması regresyon analizi ile varyans analizine ilişkin bazı önemli özelliklerin görülmesini sağlayacaktır. (Bu yaklaşıma ilişkin etkili tartışmalar ve alınması gerekli olan kararlar Harris'te (2005) yer almaktadır.) İlk olarak eşit örneklem büyüklüklerinde iki yönlü-varyans analizini ele alacağız.

### Tam Model

İki-yönlü varyans analizine ilişkin en yaygın model aşağıdaki gibidir:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

Daha önce yaptığımız gibi, $a_i$ ve $\beta_j$ terimleri, tasarım matrisi kullanılarak genişletilebilir. Fakat daha sonra etkileşim terimi nasıl ele alınmalıdır? Bu sorunun cevabı, etkileşim te-

riminin, bu etkileşimin kendisini oluşturan bileşenlerin çarpımını temsil ettiği gerçeğine dayanır. 2 × 2 faktöriyel desenin en basit durumunun incelendiğini varsayalım. *Girdiler matrisinin satırlarına yerleştirildiginde, desenin karşı gelen hücresindeki tüm katılımcılara ait katsayıları temsil eder* ve tasarım matrisi aşağıdaki gibi yazılabilir:

$$\mathbf{X} = \begin{array}{c} \\ a_1b_1 \\ a_1b_2 \\ a_2b_1 \\ a_2b_2 \end{array} \begin{array}{c} \begin{array}{ccc} \boldsymbol{A_1} & \boldsymbol{B_1} & \boldsymbol{AB_{11}} \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \end{array}$$

İlk sütun *A*'nın temel etkisini temsil etmekte ve buradaki katılımcıları $A_1$ grubundakiler ve $A_2$ grubundakiler şeklinde ayırmaktadır. İkinci sütun ise *B*'nin temel etkisini temsil etmekte ve katılımcıları yine $B_1$ ve $B_2$ şeklinde ayırmaktadır. Üçüncü sütun ise *A* ve *B*'nin etkileşimidir. Buna göre, 1 = 1 × 1, –1 = 1 × –1, –1 = –1 × 1 ve 1 = –1 × –1 olur. Bir kez daha, her bir etki için bu etkiye ilişkin serbestlik derecesine sahip olduğumuz çok sayıda sütun vardır. Buradaki matrise 0 değeri girilmemiştir. Çünkü her bir değişkenin iki düzeyi olup bir katılımcı ya ilk ya da sonuncu düzeyde yer almalıdır.

Şimdi 2 × 3 faktöriyel deseni düşünelim. *A*'nın iki ve *B*'nin üç düzeyi için $sd_A = 1$, $sd_B = 2$ ve $sd_{AB} = 2$ olacaktır. Bu serbestlik dereceleri; matriste *A* için bir, *B* ve *AB* için ikişer sütun bulunması gerektiği anlamına gelir. Buna göre, aşağıdaki matris elde edilir:

$$\mathbf{X} = \begin{array}{c} \\ a_1b_1 \\ a_1b_2 \\ a_1b_3 \\ a_2b_1 \\ a_2b_2 \\ a_2b_3 \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{A_1} & \boldsymbol{B_1} & \boldsymbol{B_2} & \boldsymbol{AB_{11}} & \boldsymbol{AB_{13}} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & -1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{array}$$

$A_1$ sütunu katılımcıları $A_1$ olanlar ve $A_2$ olanlar şeklinde ikiye; ikinci sütun $B_1$'de yer alan katılımcıları $B_1$'de olmayan katılımcılardan ayırır. Bu işlemin aynısını üçüncü sütun $B_2$ için yapmaktadır. Yine, ilk *a*–1 ve ilk *b*–1 işlem düzeyindeki katılımcılar sorudaki işlem düzeyinde bulunup bulunmamasına bağlı olarak 1 ya da 0 şeklinde kodlanır. *a*'yıncı veya *b*'yinci işlem düzeylerindeki katılımcılar bu işlem etkisi ile ilgili her bir sütun için –1 ile kodlanır. $AB_{11}$ olarak adlandırılan sütun, $A_1$ ve $B_1$ sütunlarının çarpımı benzer şekilde, $AB_{12}$ sütunu $A_1$ ve $B_2$ sütunlarının çarpımıdır.

İki veya daha fazla temel etki ile bir veya daha fazla etkileşimin test edilmesi amaçlandığından, faktöriyel desenin analizi tek-yönlü varyans analizine göre daha zahmetlidir. İki-yönlü faktöriyel desenin nispeten basit bir örneği ele alınırsa, bu örneğin daha karmaşık faktöriyel desenlere genellenmesi zor olmayacaktır. Hesaplanmasının daha karmaşık olması dışında temel ilkeler aynıdır.

Bir örnek olmak üzere, her bir hücrede 4 katılımcının bulunduğu 2 × 4 faktöriyel desen ele alalım. Böyle bir desen, Tablo 16.2'deki geleneksel varyans analizi ile analiz edilmiştir. Bu tablo aynı zamanda, ortalamaları, kestirilen etkileri ve $\eta^2$ değerini de içermektedir.

**Tablo 16.2** 2 x 4 faktöriyel desen için örnek veri kümeleri ve özet tablo

**(a) Veri**

| | $B_1$ | $B_2$ | $B_3$ | $B_4$ | Ortalamalar |
|---|---|---|---|---|---|
| $A_1$ | 5 | 2 | 8 | 11 | |
| | 7 | 5 | 11 | 15 | |
| | 9 | 7 | 12 | 16 | |
| | 8 | 3 | 14 | 10 | |
| | 7,25 | 4,25 | 11,25 | 13,00 | 8,92750 |
| $A_2$ | 7 | 3 | 9 | 11 | |
| | 9 | 8 | 12 | 14 | |
| | 10 | 9 | 14 | 10 | |
| | 9 | 11 | 8 | 12 | |
| | 8,75 | 7,75 | 10,75 | 11,75 | 9,75000 |
| Ortalamalar | 8,000 | 6,000 | 11,000 | 12,375 | 9,34375 |

**(b) Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* | $\eta 2$ |
|---|---|---|---|---|---|
| *A* | 1 | 5,282 | 5,282 | <1 | ,014 |
| *B* | 3 | 199,344 | 66,448 | 11,452* | ,537 |
| *AB* | 3 | 27,344 | 9,115 | 1,571 | ,074 |
| Hata | 24 | 139,250 | 5,802 | | |
| Toplam | 31 | 371,220 | | | |

*$p < .05$

**(c) Kestirilen İşlem Etkileri**

$$\hat{\mu} = 9.34375$$
$$\hat{\alpha}_1 = \overline{A}_1 - \overline{X}_{..} = 8.9375 - 9.34375 = -0.40625$$
$$\hat{\beta}_1 = \overline{B}_1 - \overline{X}_{..} = 8.0000 - 9.34375 = -1.34375$$
$$\hat{\beta}_2 = \overline{B}_2 - \overline{X}_{..} = 6.0000 - 9.34375 = -3.34375$$
$$\hat{\beta}_3 = \overline{B}_3 - \overline{X}_{..} = 11.0000 - 9.34375 = 1.65625$$
$$\widehat{\alpha\beta} = \overline{AB}_{11} - \overline{A}_1 - \overline{B}_1 + \overline{X}_{..} = 7.2500 - 8.9375 - 8.0000 + 9.34375 = -.34375$$
$$\widehat{\alpha\beta}_{12} = \overline{AB}_{12} - \overline{A}_1 - \overline{B}_2 + \overline{X}_{..} = 4.2500 - 8.9375 - 6.0000 + 9.34375 = -1.34375$$
$$\widehat{\alpha\beta}_{13} = \overline{AB}_{13} - \overline{A}_1 - \overline{B}_3 + \overline{X}_{..} = 11.2500 - 8.9375 - 11.0000 + 9.34375 = 0.65625$$

© Cengage Learning 2013

Özet tablodan, *B*'nin ana etkisinin manidar olduğu, fakat *A*'nın ana etkisi ve *AB* etkileşiminin etkisinin manidar olmadığı görülmektedir.

Bu veri kümelerini çoklu regresyon bakış açısıyla analiz edebilmek için aşağıdaki tasarım matrisi ile başlayacağız. Bir kez daha, her satırdaki elemanlar karşılık gelen işlem kombinasyondaki tüm katılımcılara uygulanır.

$$
\mathbf{X} = \begin{array}{c} \\ a_1b_1 \\ a_1b_2 \\ a_1b_3 \\ a_1b_4 \\ a_2b_1 \\ a_2b_2 \\ a_2b_3 \\ a_2b_4 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \boldsymbol{A}_1 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \boldsymbol{AB}_{11} & \boldsymbol{A}_{12} & \boldsymbol{AB}_{13} \end{array} \\
\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}
\end{array}
$$

$A_1$'den $AB_{13}$'e kadar olan yedi yordayıcıya ait çoklu regresyon analizinin ilk adımı Gösterim 16.2'de sunulmuştur. Analiz sonuçları, SAS PROC CORR ve PROC REG programları kullanılarak elde edilmiştir. Bununla birlikte, tüm paket programlardan aynı sonuçlar elde edilebilir. Bunu R'de yapmak için bir program Web sitesinde bulunmaktadır.

Gösterim 16.2'de önemli özellikler vardır. Genel olarak **R** ile gösterilen, değişkenler arasındaki korelasyon matrisini düşünelim. $A' = [A_1]$, $B' = [B_1, B_2, B_3]$ ve $AB' = [AB_{11}, AB_{12}, AB_{13}]$ yordayıcı değişkenlerini tanımlayarak bu matrisi basitleştirdiğimizi varsayalım. Sonrasında korelasyon matrisi tekrar yazılırsa aşağıdaki matris elde edilecektir:

$$
\begin{array}{c} \\ \boldsymbol{A}' \\ \boldsymbol{B}' \\ \boldsymbol{AB}' \end{array}
\begin{array}{c}
\begin{array}{ccc} \boldsymbol{A}' & \boldsymbol{B}' & \boldsymbol{AB}' \end{array} \\
\begin{bmatrix} 1{,}00 & 0{,}00 & 0{,}00 \\ 0{,}00 & 1{,}00 & 0{,}00 \\ 0{,}00 & 0{,}00 & 1{,}00 \end{bmatrix}
\end{array}
$$

Etkilerin her birinin bir diğerinden bağımsız olduğuna dikkat ediniz (değişkenler arası korelasyonlar 0,00'dır). Bu tür bir örüntü, ancak her bir hücredeki katılımcı sayısı eşit (veya orantılı) olduğunda ortaya çıkar. Bu model, varyans analizi formülünü de basitleştirir. Örneklem büyüklükleri eşit olmadığında bu yapı kaybolur, bu da veri kümesinde eksik veri bulunması durumunda yapılacak işlemleri güçleştirir.

Sonraki dikkat edilecek husus, **b** vektörünün Parametre Kestirimi olarak adlandırıldığıdır. İlk girdi ($b_0$) sabit olarak adlandırılmıştır ve tüm gözlemlerin genel ortalamasıdır. İzleyen girdiler ($b_1 \ldots b_7$), ilgili işlem etkilerine ait kestirimlerdir. Buna göre, $b_1 = \alpha_1$, $b_2 = \beta_1$, $b_5 = a\beta_{11}$ vb. Regresyon katsayılarının test edilmesiyle ilgili testler işlem etkilerinin test edilmesini temsil eder. Satır ve sütunlar boyunca etkileşim terimlerinin toplamının 0'a eşit olması kısıtlamasından dolayı, sadece $(a - 1)(b - 1) = 3$ etkileşim etkisinin bulunması problem oluşturmaz. Buna göre; $a\beta_{12} = -1{,}34$ ise $a\beta_{22} = +1{,}34$ olmalıdır. Benzer şekilde, $a\beta_{14} = 0 - \Sigma\, a\beta_{1j} = -\Sigma a\beta_{1j} = 1{,}03$ olur.

$R^2 = {,}625$ değeri, aynı anda tüm yordayıcılar ile açıklanabilen değişkenlik yüzdesini belirtir. Eşit örneklem büyüklüklerinde, dolayısıyla bağımsız etkilerde, $\eta^2_A + \eta^2_B + \eta^2_{AB} = {,}014 + {,}537 + {,}074 = {,}625$ değerine eşittir. $R^2$'nin manidarlığı test edilirse, 7 ve 24 serbestlik derecelerinde 5,711'e eşit ve manidar (0,0006) bir $F$ değerine ulaşılır ($F = 5{,}711$, $p < {,}0006$). Bu sonuç bağımlı değişken ($Y$) ile işlem değişkenleri arasındaki ilişkinin şans sonucu gerçekleşmesi olasılığının beklenenden daha yüksek olduğunu gösterir.

Tablo 16.2'deki varyans analizi ve Gösterim 16.2'deki regresyon analizi sonuçları arasında iki paralellik daha görülebilir. İlk olarak, $KT_{\text{regresyon}} = KT_{\text{model}} = KT_Y(1 - R^2) = 231{,}969$. Bu, yordayıcı değişkenlerin doğrusal kombinasyonu ile açıklanabilen varyans miktarıdır. Bu değer, $KT_A + KT_B + KT_{AB}$ değerine eşittir. Bununla birlikte, Gösterim 16.2'de değişkenlik ayrı kaynaklara bölünememektedir. Son olarak, $KT_{\text{artık}} = KT_{\text{hata}} = KT_Y(1 - R^2) = 139{,}250$ olup varyans analizindeki hatanın kareler toplamıdır. Hatanın, işlem değişkenlerinin ayrı ayrı veya ortak etkilerine atfedilemeyen değişkenlik olduğu hatırlandığında, bu sonuç mantıklı gelmektedir.

```
Data Anova;
      infile 'Ex162.dat';
      input A1 B1 B2 B3 dv;
      AB11 = A1*B1;
      AB12 = A1*B2;
      AB13 = A1*B3;
Run;
Proc Corr Data = Anova;
      Var A1 B1 B2 B3 AB11 AB12 AB13;
Run;
Proc Reg Data = Anova;
      Model dv = A1 B1 B2 B3 AB11 AB12 AB13;
Run;
```

Pearson Correlation Coefficients, N = 32

Prob > |r| under H0: Rho=0

| | A1 | B1 | B2 | B3 | AB11 | AB12 | AB13 |
|---|---|---|---|---|---|---|---|
| A1 | 1,00000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 |
| | | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| B1 | 0,00000 | 1,00000 | 0,50000 | 0,50000 | 0,00000 | 0,00000 | 0,00000 |
| | 1,0000 | | 0,0036 | 0,0036 | 1,0000 | 1,0000 | 1,0000 |
| B2 | 0,00000 | 0,50000 | 1,00000 | 0,50000 | 0,00000 | 0,00000 | 0,00000 |
| | 1,0000 | 0,0036 | | 0,0036 | 1,0000 | 1,0000 | 1,0000 |
| B3 | 0,00000 | 0,50000 | 0,50000 | 1,00000 | 0,00000 | 0,00000 | 0,00000 |
| | 1,0000 | 0,0036 | 0,0036 | | 1,0000 | 1,0000 | 1,0000 |
| AB11 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 1,00000 | 0,50000 | 0,50000 |
| | 1,0000 | 1,0000 | 1,0000 | 1,0000 | | 0,0036 | 0,0036 |
| AB12 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,50000 | 1,00000 | 0,50000 |
| | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 0,0036 | | 0,0036 |
| AB13 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,50000 | 0,50000 | 1,00000 |
| | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 0,0036 | 0,0036 | |

The REG Procedure
Dependent Variable: dv

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 7 | 231,96875 | 33,13839 | 5,71 | 0,0006 |
| Error | 24 | 139,25000 | 5,80208 | | |
| Corrected Total | 31 | 371,21875 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2,40875 | R-Square | 0,6249 |
| Dependent Mean | 9,34375 | Adj R-Sq | 0,5155 |
| Coeff Var | 25,77928 | | |

SAS Institute Inc.'dan uyarlanmıştır.

**Gösterim 16.2** Tablo 16.2'deki veri kümesi için tüm yordayıcılar kullanılarak elde edilen regresyon çözümleri

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 9,34375 | 0,42581 | 21,94 | <,0001 |
| A1 | 1 | −0,40625 | 0,42581 | −0,95 | 0,3496 |
| B1 | 1 | −1,34375 | 0,73753 | −1,82 | 0,0809 |
| B2 | 1 | −3,34375 | 0,73753 | −4,53 | 0,0001 |
| B3 | 1 | 1,65625 | 0,73753 | 2,25 | 0,0342 |
| AB11 | 1 | −0,34375 | 0,73753 | −0,47 | 0,6454 |
| AB12 | 1 | −1,34375 | 0,73753 | −1,82 | 0,0809 |
| AB13 | 1 | 0,65625 | 0,73753 | 0,89 | 0,3824 |

**Gösterim 16.2** *(Devamı)*

## İndirgenmiş Modeller

Bu noktada sadece yordayıcı değişkenlerin tümünün aynı anda açıklayabildiği varyans miktarını biliyoruz. Bilmek istediğimiz, bu değişkenliğin $A$, $B$ ve $AB$ arasında nasıl parçalandığıdır. Bu bilgi, çeşitli indirgenmiş regresyon eşitlikleri yazılarak elde edilebilir.

Analizin sonraki evrelerinde, farklı etkiler ile ilgili çeşitli çoklu regresyon kareler toplamları hesaplanması gerekeceğinden, sembolleri değiştirip etki adlandırmalarını ($a$, $\beta$, ve $a\beta$ harflerini) alt indis olarak kullanacağız. Hesaplanan çoklu regresyon için model; $a$, $\beta$, ve $a\beta$'yi açıklamak için değişkenler içerir. Bu nedenle, bu çözümdeki kareler toplamını regresyon $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} \cdot KT_{\text{regresyon}}$ olarak göstereceğiz. Eğer son üç yordayıcı çıkarılırsa ($AB_{11}, AB_{12}, AB_{13}$) etkileşim ile ilgili bilgi taşıyan yordayıcılar silinmiş olacak fakat $a$ ve $\beta$ ile ilgili yordayıcılar kalacaktır. Buna bağlı olarak, $KT_{\text{regresyon}_{\alpha,\beta}}$ gösterimini kullanacağız. Yordayıcı olarak sadece; $A$, $AB_{11}$, $AB_{12}$ ve $AB_{13}$ terimlerini kullanırsak, modeli sadece $a$ ve $a\beta$ için hesaplayıp sonucu $KT_{\text{regresyon}_{\alpha,\alpha\beta}}$ ile göstereceğiz.

Örneğimiz için her bir regresyon çözümünü hesapladım, aşağıdaki sonuçları buldum:

$$KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} = 231{,}969$$
$$KT_{\text{regresyon}_{\alpha,\beta}} = 204{,}625$$
$$KT_{\text{regresyon}_{\beta,\alpha\beta}} = 226{,}687$$
$$KT_{\text{regresyon}_{\alpha,\alpha\beta}} = 32{,}635$$

Şimdi bu önemli bir kısımdır. Etkileşim terimi $Y$'deki değişkenliğin herhangi bir kısmını açıklıyorsa, etkileşim teriminin modelden çıkarılması açıklanabilen varyans miktarının düşmesine yol açacak ve bu azalma etkileşim terimi ile açıklanabilen varyans miktarına eşit olacaktır. Bu ve benzeri gerekçelerle aşağıdaki eşitliklere ulaşılır.

$$KT_{AB} = KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}}$$

$$KT_{A} = KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\beta,a\beta}}$$

$$KT_{B} = KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,a\beta}}$$

**Tablo 16.3** Tablo 16.2'deki veri kümesi için regresyon çözümü

$$KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} = 231{,}969 \qquad R^2 = {,}625$$
$$KT_{\text{kalan}_{\alpha,\beta,\alpha\beta}} = 139{,}250$$
$$KT_{\text{regresyon}_{\alpha,\beta}} = 204{,}6245 \qquad R^2 = {,}551$$
$$KT_{\text{regresyon}_{\beta,\alpha\beta}} = 226{,}687 \qquad R^2 = {,}611$$
$$KT_{\text{regresyon}_{\alpha,\alpha\beta}} = 32{,}625 \qquad R^2 = {,}088$$

$$KT_{AB} = KT_{\text{regresyon}_{\alpha,\beta,\alpha b}} - KT_{\text{regresyon}_{\alpha,\beta}} = 231{,}969 - 204{,}625 = 27{,}344$$
$$KT_{A} = KT_{\text{regresyon}_{\alpha,\beta,\alpha b}} - KT_{\text{regresyon}_{\beta,\alpha\beta}} = 231{,}969 - 226{,}687 = 5{,}282$$
$$KT_{B} = KT_{\text{regresyon}_{\alpha,\beta,\alpha b}} - KT_{\text{regresyon}_{\alpha,\alpha\beta}} = 231{,}969 - 32{,}625 = 199{,}344$$
$$KT_{\text{hata}} = KT_{\text{kalan}_{\alpha,\beta,\alpha b}} = 139{,}250$$

**Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* | 1 | 5,282 | 5,282 | <1 |
| *B* | 3 | 199,344 | 66,448 | 11,452* |
| *AB* | 3 | 27,344 | 9,115 | 1,571 |
| Hata | 24 | 139,250 | 5,802 | |
| | 31 | 371,220 | | |

*$p < ,05$

© Cengage Learning 2013

İlgili hesaplamalar Tablo 16.3'te görülmektedir. (İlgili yordayıcıların kullanılması durumunda, regresyon için kareler toplamının hesaplanmasına ait işlemlerin doğrulanmasını okuyucuya bıraktım.)

İlk olarak AB etkileşimine bakıldığında, Gösterim 16.2'den de görüldüğü gibi, etkileşim terimleri modelden silindiğinde, hesaplanan kareler toplamının azalması nedeniyle model aşağıdaki ifadeye indirgenir:

$$KT_{AB} = KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}} = 231{,}969 - 204{,}625 = 27{,}344$$

Bu azalma, etkileşim teriminden kaynaklanan değişkenliktir. Dolayısıyla,

$$KT_{AB} = 27{,}344$$

Benzer akıl yürütme ile diğer kareler toplamlarını da hesaplayabiliriz.[2]

Bu değerlerin geleneksel yöntemler ile elde edilen değerlerle tamamen aynı olduğuna dikkat edilmelidir. Ayrıca, $R^2$ değerindeki azalmanın hesaplanan $\eta^2$ ile örtüştüğü de gözden kaçırılmamalıdır.

Overall ve Spiegel'ın (1969) ifade ettikleri gibi; *A*, *B* ve *AB*'nin test edilmesi için burada kullanılan yaklaşım seçilebilecek tek yol değildir. Overall ve Spiegel (1969), burada kullanılan yaklaşıma alternatif olarak düşünülebilecek iki model daha sunmuşlardır. Örneklem

2 Çok sayıda araştırmacı (örneğin, Judd & McClelland) $KT_{\text{regresyon}}$ yerine hata terimindeki artışı kullanmayı tercih etmektedir. Ancak hangisi kullanılırsa kullanılsın sonuç aynı olmaktadır.

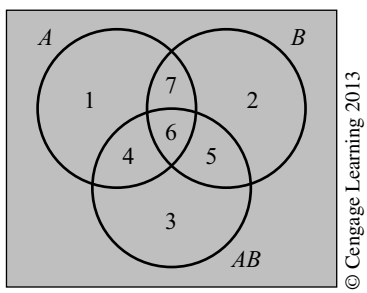
büyüklükleri eşit olduğunda farklı modellerden elde edilen sonuçlar eşdeğer olmaktadır. Çünkü bu durumda, etkilerden biri diğerinden bağımsız olmakta ve dolayısıyla birbirleri ile toplanabilmektedir. Öte yandan örneklem büyüklüklerinin farklı olması halinde, seçilen model üzerinde dikkatli bir biçimde düşünülmesi gerekir.

## 16.4 Eşit Olmayan Örneklem Büyüklükleri ile Varyans Analizi

Varyans analizinde en küçük-kareler yaklaşımı, özellikle eşit olmayan örneklem büyüklüklerinde faktöriyel desenler için yararlıdır. Bununla birlikte, çeşitli kareler toplamının elde edilmesi için kullanılan belirli bir sınırlandırılmış modelin seçiminde dikkatli olunmalıdır.

Varyans analizinin altında birbirinden farklı modeller yatabilir. Örneklem büyüklükleri eşit olduğunda, bu farklı modellerden elde edilen sonuçlar aynıdır. Örneklem büyüklükleri eşit olmadığında, farklı modellerden elde edilen sonuçlar değişiklik gösterir. Çünkü örneklem büyüklükleri farklı olduğunda; satır, sütun ve etkileşim etkileri dik (ortogonal) olmamakta ve bu nedenle açıklanan varyansta kesişmeler olabilmektedir. [Kesim 13.11'in, (s. 444-446) gözden geçirilmesini şiddetle tavsiye ederim.] Şekil 16.1'deki Venn diyagramını inceleyelim. Kare ile çevrili gri alan $KT_{\text{toplam}}$ değerini temsil eder ve daireler her bir etkiden kaynaklanan değişkenliğe karşılık gelir. Karenin içindeki fakat dairelerin dışında kalan alan $KT_{\text{hata}}$ değerini gösterir. Son olarak, dairelerin birleşiminden oluşan alan tüm değişkenlerin modele dâhil edilmesi halinde regresyon kareler toplamına karşılık olan $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}}$ değerini yansıtır. Örneklem büyüklükleri eşit olduğunda, dairelerin kesiştikleri herhangi bir alan bulunmaz ve her bir etki değişkenliğin bir diğerinden bağımsız parçası olur. Böyle bir durumda, bir etkinin modelden silinmesi sonucunda $KT_{\text{regresyon}}$ değerinde görülen azalmanın açık bir yorumu olur - azalma, çıkarılan etkiye ilişkin dairenin çevrelediği alan olmakta ve dolayısıyla söz konusu etkiye ilişkin kareler toplamına karşılık gelmektedir.

Fakat dairelerin kesiştikleri alan olması durumunda nasıl bir sonuçla karşılaşılacaktır? Eğer *A*, *B* ve *AB*'yi içeren bir model alınır ve sadece *A* ile *B*'yi içeren model ile karşılaştırılırsa, modeldeki azalma *AB* dairesinin alanı olmayacaktır. Çünkü *AB* dairesinin *A* veya *B* daireleri ile kesişim noktaları bulunmaktadır. Buna göre, $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}}$ yoluyla hesaplanan $KT_{AB}$, sadece *AB*'nin *A* ve *B*'den farkını temsil eder - bu alan "3" ile gösterilmiştir. Şimdiye kadar, önerilen tüm modeller uyum içindedir. $KT_{AB}$, *AB* dairesinin *A* ve B çıkarıldıktan sonra kalan kısmıdır. Fakat şimdi işler biraz karmaşıklaşmaktadır.



© Cengage Learning 2013

**Şekil 16.1** Değişkenliğin tüm bölümlerini temsil eden Venn diyagramı

Pastanın *A* ve *B* için ayrılan kısmına farklı ve anlamlı yaklaşımlar getirilmiştir. Overall ve Spiegel (1969), varyans analizi için üç model ileri sürmüşlerdir. 30 yıl önce başlamasına rağmen, bu modellerin kullanılması ve yorumlanması ile ilgili tartışmalar devam etmektedir. SPSS ve SAS programlarında yer verilen terminolojiden hareketle bu modeller için burada Tip I, Tip II ve Tip III adlandırmaları kullanılacaktır. (Overall ve Spiegel, sadece işleri daha karmaşık hale getirmek için bu yaklaşımları tersten numaralandırmışlardır.) Temel olarak, üç model arasındaki seçim örneklem büyüklüğü ile işlemin kendisi arasındaki ilişkinin nasıl görüldüğüne veya daha özel olarak satır ve sütun ortalamalarının elde edilmesinde çeşitli hücre ortalamalarının nasıl ağırlıklandırıldığına dayanmaktadır. Ancak, bu konuya açıklık getirmeden önce, rakip yöntemlerin incelenmesi gerekir.

**Metot III**

**Tip III kareler toplamı**

**Metot III** (veya **Tip III kareler toplamı**), önceki kesimde kullanılan yöntemdir. Bu durumda, her bir etkiye diğer etkilerin tümü için düzeltme uygulanır. Bu nedenle $KT_{AB}$, $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}}$ olarak; $KT_A$, $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\beta,\alpha\beta}}$ olarak ve $KT_B$, $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\alpha\beta}}$ olarak elde edilir. Şekil 16.1 açısından, her bir etki sadece bu etkiye ayrılan alan ile tanımlanır. Buna göre, $KT_A$ "1" numaralı alan; $KT_B$ "2" numaralı alan ve $KT_{AB}$ "3" numaralı alan ile temsil edilir.

**Metot II**

**Tip II *KT***

**Metot II** (veya **Tip II *KT***) pastayı farklı şekilde parçalar. $KT_{AB}$ yine "3" numaralı alan ile tanımlanır. Ancak şimdi etkileşim etkisini halletmiş olsak da; *A* veya *B*'nin etkisi ile açıklanan ve henüz halledilmeyen "1", "2", "4", "5", "6" ve "7" numaralı alanlar bulunmaktadır. Metot II esasen modelin genelini $KT_{\text{regresyon}_{\alpha,\beta}}$ olarak yeniden tanımlamakta ve $KT_{\text{regresyon}_{\alpha,\beta}} - KT_{\text{regresyon}_{\beta}}$. $KT_B = KT_{\text{regresyon}_{\alpha,\beta}} - KT_{\text{regresyon}_{\alpha}}$ olarak elde edilmektedir. Dolayısıyla *A,* "1" ve "4", *B* ise "2" ve "5" ile temsil edilen alandır. Metot II ve III Tablo 16.4'te özetlenmiştir.

**Tablo 16.4** Dik olmayan desenlerin çözümü için alternatif modeller

Metot III

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

| Kaynak | *sd* | *KT* | Venn Diyagramındaki Kısım |
|---|---|---|---|
| *A* | $a-1$ | $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\beta,\alpha\beta}}$ | 1 |
| *B* | $b-1$ | $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\alpha\beta}}$ | 2 |
| *AB* | $(a-1)(b-1)$ | $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}}$ | 3 |
| Hata | $N-ab$ | $KT_{\text{kalan}_{\alpha,\beta,\alpha\beta}}$ | |
| Toplam | $N-1$ | $KT_Y$ | |

Metot II

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

ve

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

| Kaynak | *sd* | *KT* | |
|---|---|---|---|
| *A* | $a-1$ | $KT_{\text{regreson}_{\alpha,\beta}} - KT_{\text{regreson}_{\beta}}$ | 1 + 4 |
| *B* | $b-1$ | $KT_{\text{regreson}_{\alpha,\beta}} - KT_{\text{regreson}_{\alpha}}$ | 2 + 5 |
| *AB* | $(a-1)(b-1)$ | $KT_{\text{regreson}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regreson}_{\alpha,\beta}}$ | 3 |
| Hata | $N-ab$ | $KT_{\text{kalan}_{\alpha,\beta,\alpha\beta}}$ | |
| Toplam | $N-1$ | $KT_Y$ | |

© Cengage Learning 2013

Şekil 16.1'deki Venn diyagramı açısından bakıldığında, bu yöntemlerin her ikisi de mantıklı görünmektedir. Bununla birlikte, Venn diyagramı ham bir yaklaşımdır ve bu ham yaklaşımla gidilebilecek noktaya kadar biraz zorlama ile ilerlenebilir.[3]

Carlson ve Timm'in (1974) belirttiği gibi, modelleri karşılaştırmanın daha uygun bir yolu test ettikleri hipotezlerin incelenmesidir. Carlson ve Timm'e (1974) göre, Metot III, hücre ortalamalarının eşit olarak ağırlıklandırılması durumunda grup etkisinin kestirimini içerir. Metot III özellikle, örneklem büyüklüğünün işlem koşullarından bağımsız olduğunun düşünüldüğü durumlarda uygun olur. Bunun ikna edici gösterimleri Overall, Spiegel ve Cohen'da (1975) sunulmuştur.[4] Ayrıca, Carlson ve Timm, sadece herhangi bir etkileşimin bulunmadığı durumlarda olmak üzere, Metot II'nin satır ve sütun ortalamalarını örneklem büyüklüğüne göre ağırlıklandırarak grup etkisine ilişkin kestirimleri ürettiğini göstermiştir. Etkileşim mevcut olduğunda, satır ve sütun etkisine ilişkin basit tahminler yapılamamakta ve gerçekte test edilen sıfır hipotezi tuhaf olmaktadır [Metot II'de sıfır hipotezinin ifadesi için Carlson ve Timm'e (1974) bakınız]. Metot II'ye benzeyen bir yönteme dayanan SPSS, birkaç yıl önce ışığı görmüş ve kendinden tanımlı bir özellik olarak Metot III'ü kullanmaya başlamıştır. Bu metot "Benzersiz *KT* (Unique *SS*)" olarak adlandırılmıştır. Çünkü her bir etki sadece tek başına açıklayabildiği değişkenlik kısmına atanmaktadır. SAS kendinden tanımlı bir özellik olarak daima Metot III'ü kullanmaktadır. Farklı yaklaşımlarda test edilen hipotezler hakkında oldukça etkili tartışmalar için Blair ve Higgins (1978) ile Blair'e (1978) başvurulabilir. Hector, von Felten ve Schmid (2010) alternatif analizler ile ilgili kusursuz bir tartışma sunmuşlardır. Cochran ve Cox, kayıp veri problemine ilişkin tam bir çözüm bulunmadığını tavsiye etmişlerdir (s. 82).

**Tip I *KT***

**Metot I**

**sıralı kareler toplamı**

**hiyerarşik kareler toplamı**

Kareler toplamını hesaplamanın ilk bakışta tuhaf görünebilecek üçüncü bir yolu daha vardır. Sadece gerektiğinden daha karmaşık hale getirmek için bu yöntem SPSS ve SAS programlarında **Tip I *KT*** ya da **Metot I** olarak adlandırılmaktadır. Bazen **sıralı kareler toplamı** olarak adlandırılsa da, burada SPSS ve SAS programları tarafından kullanılan terim olan **hiyerarşik kareler toplamı** ifadesiyle adlandıracağım. Bu yaklaşımın tuhaf yönü değişkenlere verilen isim sırasına bağlı olmasıdır. Buna göre, eğer SPSS veya SAS programına bağımlı değişkeni *A*, *B* ve *AB* değişkenleri ile kestirmesi talimatı verilirse, program tarafından ilk olarak, $KT_{\text{A}} = KT_{\text{regresyon}_{\alpha}}$, ardından $KT_{\text{B}} = KT_{\text{regresyon}_{\alpha,\beta}} - KT_{\text{regresyon}_{\alpha}}$ ve son olarak $KT_{\text{AB}} = KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta}}$ belirlenir. Bu durumda; ilk etki, açıklanması mümkün olan tüm kareler toplamını, ikinci etki ilk etki ile açıklananın *ötesindeki* tüm kareler toplamı ve son olarak etkileşim etkisi iki ana etkinin dışında açıklanabilen kareler toplamını yansıtır. Fakat programa, bağımlı değişkeni *B*, *A* ve *AB* değişkenleri ile tahmin etmesi komutu verilirse, $KT_{\text{B}} = KT_{\text{regresyon}_{\beta}}$ olacaktır ki bu eşitlik $KT_{B} = KT_{\text{regresyon}_{\alpha,\beta}} - KT_{\text{regresyon}_{\alpha}}$ eşitliğinden oldukça farklıdır. Bu yaklaşımın, sadece değişkenlerin belirli bir sıra içinde kontrol edilmek istenmesine yönelik güçlü bir gerekçe olduğu zamanda kullanılmasını önerebilirim.[5] Eğer *A*'nın çok önemli olduğu ve diğer değişkenler kontrol altında alınmadan önce bakılması gerektiğine dair gerekçeler varsa, belki bu yaklaşım kullanabilecek bir yöntem olabilir. Ancak, bir değişkenin ortak değişken olarak alındığı durumlar hariç olmak üzere, bu yöntemin kullanabileceği durumlar ile karşılaşma pek muhtemel değildir. Burada bu konuya yer verilmesinin tek nedeni, bilgisayar programlarını kullanırken yapmak zorunda kalacağınız bazı tercihleri açıklamaktır.

---

3 Bu tartışmadan Yöntem II'nin Yöntem III'e göre daima daha fazla değişkenlik açıklayabileceği izlenimi kolaylıkla edinilebilir. Bu mutlaka böyle değildir, çünkü iki etkinin çakışma derecesi etkiler arasındaki korelasyonu temsil etmekte ve açıklanabilen değişkenliğin dışında "kara delikler" olarak bastırıcı ilişkileri görülebilmektedir.

4 Eğer daha eski bu literatüre bakılırsa, bu çalışmalar yazıldığında Yöntem III'ün Yöntem I ve Yöntem I'in Yöntem III olarak adlandırıldığına dikkat edilmelidir.

5 Sürekli değişkenler ile standart çoklu regresyonu kullanan teorik çalışmalar için değişkenleri bu şekilde önceliklendirmenin iyi ve saygın bir geleceği vardır. Hipotetik örnekler üzerinde konuşan çok sayıda kişi görmeme rağmen, varyans analizi çerçevesinde benzer bir uygulamayı hiç görmedim.

R ailesi hakkında yazan birçok kişi test edilecek model hakkında güçlü görüşlere sahiptirler ve sözlerini sakınmazlar, bu kişiler güçlü bir şekilde Model I çözümünü savunurlar. Yani, "anova" ile elde edilen sonuçlar, R versiyonu tabanında komut vermektedir. Fox (2002) Tip II çözümünü destekler ve onun R'de ("car" alarak adlandırılmıştır) yazılmış fonksiyonları, uygun karşılaştırmaların belirtildiğini varsayarak Tip II ve Tip III çözümlerinin her ikisini de üretebilmektedir. (Fox'un ana fonksiyonu Anova olarak adlandırılır, büyük harf A ile gösterilir, temel sistem "a" harfini kullanır ve Tip I kareler toplamını çözer.) Ben Tip III yaklaşımını kullandım, çünkü davranış bilimlerindeki pek çok araştırmada örneklem büyüklükleri eşit değildir ve hücrelerden bazılarını diğerlerine göre daha fazla ağırlıklandırmak için bir gerekçe bulunmamaktadır. Bununla birlikte, manidar etkileşimler varsa muhtemelen ana etkilere bakmak ile ilgilenilmeyeceği akılda tutulmalıdır (basit etkiler daha açıklayıcı olacaktır). Bu durumda, Metot II ve Metot III aynı yere götürecektir. Öte yandan göze çarpan ancak manidar olmayan bir etkileşim varsa, iki modelden elde edilen sonuçlar farklı olacaktır. Tip II ve Tip III kareler toplamları üzerine olan çatışmalar uzun zaman öncesine dayanmakta ve görünürde hiçbir çözüm bulunmamaktadır. Ben Tip III kareler toplamını kullanmayı tercih ediyorum ancak adayın Tip II kareler yöntemini kullandığı bir tez savunmasında jüri üyesi olarak görev alırsam bir uyum olmaması olasılığı yüksektir. Howell ve McConaughy (1982), etkileşimin olduğu, Metot II ile test edilecek çok az sayıda tuhaf yokluk hipotezi olduğunu tartışmışlardır, fakat her ikisini de savunmak için kılıç kuşanmaya değmez. Bununla birlikte, literatürden çıkarılacak sonuç; birçok durumda ağırlıklandırılmamış ortalamalar test edilmek istenmekte ve buna bağlı olarak Metot III daha uygun olmaktadır. (Bu, SPSS'te kendinden tanımlı olarak kullanılan metottur. Metot III kareler toplamı, SAS programında ve SPSS'in son sürümlerinde Tip III olarak adlandırılan değerlerdir.) Ayrıca, bu metot, Bölüm 13'te ele alınan *ağırlıklandırılmamış ortalamalar çözümüne* yakındır. (Hatırlanacağı gibi, daha önce farklı bir işlemi ifade etmek için kullanılmadıysa, "ağırlıklandırılmamış ortalamalar" çözümü şeklindeki geleneksel adlandırma, **eşit ağırlıklandırılmış ortalamalar** çözümü anlamına gelir. Çünkü ağırlıklandırılmamış ortalamalar ifadesinin kullanımı örneklem büyüklüğünden bağımsız olarak tüm ortalamalara eşit muamele edildiğini yansıtır.) Metot III esasen eksik gözlemlerin tamamen tesadüfî olduğunu varsayar. Bundan dolayı, gözlem sayısının daha fazla olduğu bir hücreye uygulanacak ağırlıklandırmanın gözlem sayısının daha az olduğu bir hücreye uygulanacak ağırlıklandırmadan daha fazla olması için bir gerekçe yoktur. Eğer durum böyle değilse, farklı bir yöntem düşünülmelidir.

**eşit ağırlıklandırılmış ortalamalar**

Metot III'e örnek olmak üzere, önceki örnekte kullanılan veri kümesini ele alacağız fakat, eşit olmayan hücre büyüklükleri elde etmek için altı gözlem ekleyeceğiz. Veri kümesinin yanı sıra, ağırlıklandırılmış ve ağırlıklandırılmamış satır ve sütun ortalamaları ile çeşitli regresyon çözümlerinden elde edilen değerler Tablo 16.5'te verilmiştir. Ağırlıklandırılmamış ortalamalar, ortalamaların ortalamasıdır. (Bundan dolayı, satır$_1$'in ortalaması, bu satırdaki dört hücrenin ortalamasıdır.) Satır$_1$'in ağırlıklandırılmış ortalaması, örneğin, satır$_1$'deki puanların toplamının bu satırdaki puan sayısına bölümüne eşittir.

Tablo 16.5'ten $R^2_{\alpha,\beta,\alpha\beta} = ,541$ olduğunu görürüz. Bu sonuç, bağımlı değişkendeki varyansın yaklaşık %54'ünün yordayıcı değişkenlerin doğrusal kombinasyonu ile açıklanabildiğini gösterir. Ancak, bu değişkenliğin ne kadarının *A*'dan, ne kadarının *B*'den ve ne kadarının *AB*'den geldiği bilinmemektedir. Bundan dolayı indirgenmiş modellerin oluşturulması ve hesaplanmasına ihtiyacımız vardır.

## Etkileşim Etkilerinin Test Edlmesi

İlk olarak, etkileşim terimi ile ilgili yordayıcılar silinir ve $R^2_{\alpha,\beta}$ hesaplanır. Bu veri için $R^2_{\alpha,\beta} = ,483$ olup $R^2$ değerinde yaklaşık ,05'lik bir düşüşe karşı gelmektedir. Kestirilebilen

**Tablo 16.5** Dik olmayan faktöriyel desenler için örnek hesaplamalar

| | $B_1$ | $B_2$ | $B_3$ | $B_4$ | Ağırlıklandırılmamış Ortalamalar | Ağırlıklandırılmış Ortalamalar |
|---|---|---|---|---|---|---|
| | 5 | 2 | 8 | 11 | | |
| | 7 | 5 | 11 | 15 | | |
| $A_1$ | 9 | 7 | 12 | 16 | 8,975 | 8,944 |
| | 8 | 3 | 14 | 10 | | |
| | | 9 | | 9 | | |
| | 7 | 3 | 9 | 11 | | |
| | 9 | 8 | 12 | 14 | | |
| | 10 | 9 | 14 | 10 | | |
| $A_2$ | 9 | 11 | 8 | 12 | 9,590 | 9,950 |
| | | | 7 | 13 | | |
| | | | | 11 | | |
| | | | | 12 | | |
| Ağırlıklandırılmamış Ortalamalar | 8,000 | 6,475 | 10,625 | 12,029 | 9,282 | |
| Ağırlıklandırılmış Ortalamalar | 8,000 | 6,333 | 10,556 | 12,00 | | 9,474 |
| Tam Model | $R^2_{\alpha,\beta,\alpha\beta} = ,541$ | | | | | |
| | $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} = 216{,}017$ | | | | | |
| | $KT_{artık} = 183{,}457$ | | | | | |
| İndirgenmiş Modeller | $R^2_{\alpha,\beta} = ,490$ | | | | | |
| | $KT_{\text{regresyon}_{\alpha,\beta}} = 195{,}805$ | | | | | |
| | $R^2_{\beta,\alpha\beta} = ,532$ | | | | | |
| | $KT_{\text{regresyon}_{\beta,\alpha\beta}} = 212{,}552$ | | | | | |
| | $R^2_{\alpha,\alpha\beta} = ,77$ | | | | | |
| | $KT_{\text{regresyon}_{\alpha,\alpha\beta}} = 30{,}845$ | | | | | |

© Cengage Learning 2013

kareler toplamı ($KT_{\text{regresyon}}$) incelendiğinde, etkileşim teriminin çıkarılmasının $KT_{\text{regresyon}}$ değerinde (20,212 gibi) bir azalmaya yol açtığı görülmektedir.

$$\begin{aligned} KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} &= 216{,}017 \\ KT_{\text{regresyon}_{\alpha,\beta}} &= \underline{195{,}805} \\ KT_{AB} &= 20{,}212 \end{aligned}$$

Kareler toplamındaki bu düşüş, $AB$ etkileşiminden kaynaklanmaktadır ($KT_{AB}$).

Örneklem büyüklükleri eşit olmadığında, bu terimin neyi temsil ettiğinin anlaşılması oldukça önemlidir. $KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}}$'nin $KT_{\text{Y}}$ ($R^2_{\alpha,\beta,\alpha\beta}$) değerine eşit olduğu hatırlanmalıdır. Buna göre,

$$\begin{aligned} KT_{\text{AB}} &= KT_{\text{Y}}(R^2_{\alpha,\beta,\alpha\beta}) - KT_{\text{Y}}(R^2_{\alpha,\beta}) \\ &= KT_{\text{Y}}(R^2_{\alpha,\beta,\alpha\beta} - R^2_{\alpha,\beta}) \\ &= KT_{\text{Y}}(R^2_{0(\alpha,\beta,\alpha\beta)}) \end{aligned}$$

Parantez içindeki son terim, yordayıcı değişken ile *A* ve *B*'nin etkileri açısından kontrol altına alınmış etkileşim terimi arasındaki yarı kısmi korelasyonun karesidir. Bir başka deyişle, yordayıcı değişken ile *AB* etkileşiminin *A* ile *B*'ye dik olan kısmı arasındaki korelasyonun karesidir. Buna göre, $KT_{AB}$ değeri, *A* ve *B*'nin etkileri için düzeltme uygulanmak üzere $KT_{AB(\text{düzeltilmiş})}$ olarak düşünülebilir. (Örneklem büyüklükleri eşit olduğunda böyle bir durum söz konusu olmayacaktır. Çünkü *A*, *B* ve *AB* bağımsız olacak ve bu üç etkinin kontrol altına alınması gereken kesişim noktaları bulunmayacaktır.)[6]

## Ana Etkilerin Test Edilmesi

Metot III *KT* hesaplandığından, etkileşim etkisinin kestirimi ile doğrudan karşılaştırılabilir şekilde *A* ve *B*'nin ana etkilerini hesaplayacağız. Buradaki her bir ana etki, diğer ana etkiler ve etkileşim etkisi kontrol edildikten sonra, bu ana etkiden kaynaklanan kareler toplamını temsil eder.

$KT_A$ değerini elde etmek için, A'nın ana etkisine ait yordayıcı silinerek $KT_{\text{regresyon}_{\beta,\alpha\beta}}$ hesaplanır. Buradaki veri için $R^2_{\beta,\alpha\beta} = ,523$ olduğundan $R^2$ değerinde ,532 – ,523 = ,009'luk bir düşüş olacaktır. Kestirilebilir kareler toplamı açısından ($KT_{\text{regresyon}}$), *a*'nın modelden çıkarılması $KT_{\text{regresyon}}$ değerinde 3,465'lik bir düşüşe neden olur.

$$\begin{aligned} KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} &= 216{,}017 \\ KT_{\text{regresyon}_{\beta,\alpha\beta}} &= 212{,}552 \\ \hline KT_{\text{AB}} &= \quad 3{,}465 \end{aligned}$$

Bu azalma *A* ana etkisine atfedilecek kareler toplamıdır.

Aynı akıl yürütme ile $KT_{\text{B}}$ değerini, tüm model için kareler toplamı olan $KT_{\text{regresyon}}$ değeri ve $\beta$'nin çıkarıldığı modeli karşılaştırarak elde edebiliriz.

$$\begin{aligned} KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}} &= 216{,}017 \\ KT_{\text{regresyon}_{\alpha,\alpha\beta}} &= \quad 30{,}485 \\ \hline KT_{\text{B}} &= 185{,}172 \end{aligned}$$

Elde edilen sonuçlar kullanılan yöntemlerle birlikte Tablo 16.6'da verilmiştir. Dikkat edilirse, ana etkiler, etkileşim etkisi ve hata için hesaplanan kareler toplamı değerlerinin toplamı $KT_{\text{toplam}}$ değerine eşit değildir. Bu durum açıklanabilen varyansın çakışan kısımlarının (Şekil 16.1'deki "4", "5", "6", ve "7" numaralı kısımlar) herhangi bir yerde temsil edilmemesinden kaynaklanmaktadır. Ayrıca, eşit örneklem büyüklüklerinde olduğu gibi, $KT_{\text{hata}}$ değerinin tüm modeldeki $KT_{\text{artık}}$ olarak alındığı da göz önünde bulundurulmalıdır. Burada tekrar, $KT_{\text{hata}}$ toplam değişkenliğin, bağımsız değişkenlerin biri veya daha fazlası tarafından açıklanamayan kısmını temsil eder.

---

6 Bazı kişiler bağımsız olmayan grup etkisi kavramını anlamakta güçlük çekebilir. Uç bir örnek, bir satır etkisinin belirgin bir sütun etkisine yol açabileceğinin gösterilmesine yardımcı olabilir veya tersi. Aşağıdaki iki-yönlü tabloyu düşünelim. Ortalamalar arasındaki farka baktığımızda, *A*, *B* veya *AB*'den dolayı oluşan farka mı bakarız? Bunu anlatmanın bir yolu yoktur.

| | $B_1$ | $B_2$ | Ortalamalar |
|---|---|---|---|
| $A_1$ | $\overline{X} = 10$<br>$n = 20$ | $n = 0$ | 10 |
| $A_2$ | $n = 0$ | $\overline{X} = 30$<br>$n = 20$ | 30 |
| Ortalamalar | 10 | 30 | |

**Tablo 16.6** Metot III kullanılarak kareler toplamının hesaplanması-ağırlıklandırılmamış ortalamalar çözümü

**Method III (Ağırlıklandırılmamış Ortalamalar)**

| Kaynak | *sd* | *KT* |
|---|---|---|
| *A* | $a-1$ | $KT_Y(R^2_{\alpha,\beta,\alpha\beta} - R^2_{\beta,\alpha\beta})$ |
| *B* | $b-1$ | $KT_Y(R^2_{\alpha,\beta,\alpha\beta} - R^2_{\alpha,\alpha\beta})$ |
| *AB* | $(a-1)(b-1)$ | $KT_Y(R^2_{\alpha,\beta,\alpha\beta} - R^2_{\alpha,\beta})$ |
| Hata | $N-ab$ | $KT_Y(1 - R^2_{\alpha,\beta,\alpha\beta})$ |
| Toplam | $N-1$ | $KT_Y$ |

**Varyans Analizi İçin Özet Tablo**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| *A* | 1 | 3,465 | 3,465 | <1 |
| *B* | 3 | 185,172 | 61,724 | 10,094 |
| *AB* | 3 | 20,212 | 6,737 | <1 |
| Hata | 30 | 183,457 | 6,115 | |
| Toplam | 37 | (399,474) | | |

© Cengage Learning 2013

Daha önce belirttiğim gibi, Bölüm 13'te verilen ağırlıklandırılmamış ortalamalar çözümü, burada verilen çözüme (Metot III) yakındır. Bu bölümde söz konusu çözümün ele alınmasının temel nedeni, bilgisayar programında verilen çıktıların ve eşit olmayan örneklemlerin nasıl işlem gördüğünün anlaşılmasına yardımcı olmaktır.

SPSS programı kullanılarak elde edilen çözümler Gösterim 16.3'te görülmektedir. Bu gösterim, SPSS'te yapılan Tip III kareler toplamının Tablo 16.5'teki veriye uygun analizler yaptığını örneklendirmektedir.

Gruplar-Arası Etki Testi

Bağımlı Değişken: DV

| Kaynak | Tip III Kareler Toplamı | sd | Kareler Ortalaması | F | Mnd. |
|---|---|---|---|---|---|
| Düzeltilmiş Model | 216,017[a] | 7 | 30,860 | 5,046 | ,001 |
| Sabit | 3163,341 | 1 | 3163,841 | 517,370 | ,000 |
| Satırlar | 3,464 | 1 | 3,464 | ,566 | ,458 |
| Sütunlar | 185,172 | 3 | 61,724 | 10,093 | ,000 |
| Satırlar*Sütunlar | 20,212 | 3 | 6,737 | 1,102 | ,364 |
| Hata | 183,457 | 30 | 6,115 | | |
| Toplam | 3810,000 | 38 | | | |
| Düzeltilmiş Toplam | 399,474 | 37 | | | |

[a] $R^2$ = ,541 (Düzeltilmiş $R^2$ = ,434)

SPSS çıktısından uyarlanmıştır.

**Gösterim 16.3** Tablo 16.7'deki veri için SPPS analizi çıktıları

## 16.5 Tek Yönlü Kovaryans Analizi

**kovaryans analizi**

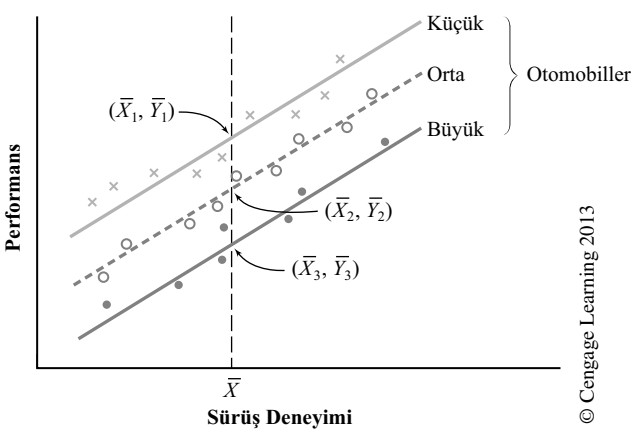
**Kovaryans analizi** deneysel çalışma veri kümelerinin analizi için çok yararlı bir yöntemdir. Varyans analizinde olduğu gibi, kovaryans analizi için yapılması gereken işlemler özellikle birden fazla birlikte değişkenliğin olduğu durumlarda hantal görünebilir. Bununla birlikte, çoklu regresyon analizi çerçevesinde, son derece basittir ve varyans analizindekine kıyasla daha az bir çalışma gerektirir. Burada kukla değişken olarak kodlanmış değişkenleri ve çoklu regresyonu açıklayacağım. Ancak, temel mantık anlaşıldığı takdirde, tasarım matrisini oluşturmaya gerek kalmadan SPSS ve SAS programını kullanarak elde edilen analiz çıktılarını anlayabileceksiniz.

Küçük otomobilleri kullanmanın daha kolay olduğu hipotezini test etmek için üç farklı boyuttaki otomobilin sürüş yeterliliğini karşılaştırmaya yönelik bir araştırmayı ele alalım. Elimizde üç farklı grup sürücü olsun. Ancak, her bir gruptaki sürücüler sürüş deneyimi açısından önemli derecede farklılık göstersin. Bu nedenle, sürücüler sürüş deneyimi açısından eşleştirilemesinler. Bu noktada, farklı gruplarda sürüş deneyimine ait ortalamaların eşit olduğu şeklinde basit bir varsayımda bulunalım. Bu varsayımı ileride ayrıntılı biçimde tartışacağız. Direksiyon hataları sayısını bağımlı değişken olarak kabul edelim ve araştırmada grafiği Şekil 16.2'de görünen biraz abartılı veri kümesini elde edelim. Bu şekilde, her bir gruba (otomobilin büyüklüğü) ait veri kümeleri sürüş deneyiminin (**ortak değişkenlik**) bir fonksiyonu olarak ayrı ayrı çizilmiş ve ayrı regresyon doğruları üst üste getirilmiştir.

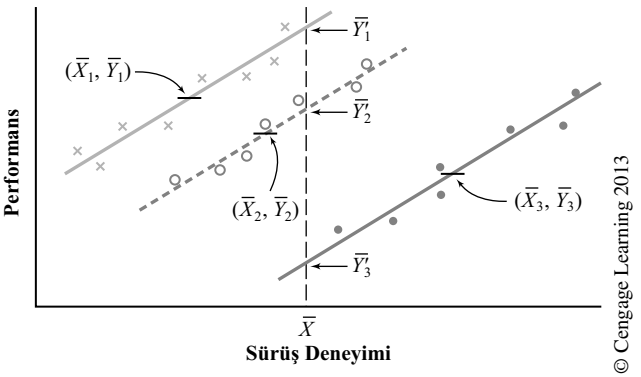
**ortak değişkenlik**

Şekil 16.2'deki en dikkat çekici noktalardan biri, her bir grup için hem sürüş performansı hem de sürüş deneyimindeki büyük değişkenliktir. Bu değişkenlik çok büyük olduğundan performans puanlarına ilişkin varyans analizinin manidar bir etki oluşturması neredeyse imkânsızdır. Performanstaki değişkenliğin büyük bir kısmı doğrudan sürücülerin sürüş deneyimlerindeki farklılıktan kaynaklanacaktır. Oysa araştırmada sürücüleri sürüş deneyimlerindeki farklılık yerine, otomobil boyutunun sürüş performansı üzerindeki etkisiyle ilgilenmekteyiz. Araştırma kapsamında cevaplamak istediğimiz hipotezin daha açık bir biçimde test edilmesi sürüş deneyiminden (ortak değişken) kaynaklanan varyansın dışarıda tutulabilmesine (kontrol altına alınmasına) bağlıdır. Kovaryans analizi tam olarak bu tür araştırmalarda hata terimini azaltmak için ileri sürülmüş bir istatistiksel tekniktir.

Kovaryans analizinin daha tartışmalı bir kullanımı, grupların ortak değişken (sürüş tecrübesi) ortalamalarının farklı olduğu durumlardır. Böyle bir durum (aynı deneysel çalışma için) Şekil 16.3'te gösterilmiştir. Bu şekilde, gruplardan ikisi $X$ ekseni boyunca hareket




**Şekil 16.2** Kovaryans analizinde hatanın azaltılmasını gösteren hipotetik veri kümesi




**Şekil 16.3** Kovaryans analizinde ortalamaların düzeltilmesini gösteren hipotetik veri

**düzeltilmiş *Y* ortalamaları**

ettirilmiştir. Regresyon doğrularının $X = \overline{X}$ dikey doğrusunu kestiği noktada, $Y'_1$, $Y'_2$ ve $Y'_3$ değerleri görülmektedir. Bunlar **düzeltilmiş *Y* ortalamaları** olup eğer *işlemler ortak değişken açısından farklılık göstermeseydi Y*'nin ortalamasının ne olacağına ilişkin en iyi kestirimi temsil edecekti. Kovaryans analizi daha sonra, kontrol altına alınan ortak değişkenden kaynaklanan hata terimini kullanarak düzeltir, bu ortalamalar arasında manidar fark bulunup bulunmadığını test eder. Burada, düzeltilmiş ortalamaların düzeltilmemiş ortalamalardan oldukça farklı olduğuna dikkat edilmelidir. Düzeltme işlemi, $\overline{Y}_1$'i arttırmış ve $\overline{Y}_3$'ü düşürmüştür.

Grupların ortak değişken ortalamaları açısından farklılık gösterip göstermediğinden bağımsız olarak kovaryans analizinin yapısı ve işlemleri aynı olmasına rağmen, Şekil 16.2 ve 16.3'te temsil edildiği gibi, problemlerin farklı biçimlerde görselleştirilmesi öğreticidir. İlk durumda basit bir şekilde hata terimi azaltılmaktadır. İkinci durumda, hem hata terimi azaltılmakta *hem de* bağımlı değişkene ait ortalamalar düzeltilmektedir. Bu bölümün sonraki kesimlerinde söz konusu ayrımdan daha derinlemesine söz edilecektir.

## Kovaryans Analizinin Varsayımları

**regresyonların homojenliği**

Kovaryans analizinde, varyans analizine ait normallik ve varyansların homojenliği varsayımlarına iki varsayım daha eklemeliyiz. İlk olarak, *Y* ve ortak değişken (*C*) arasındaki ilişki ne olursa olsun doğrusal olduğu varsayılmaktadır.[7] İkinci varsayım ise **regresyonların homojenliği**dir - yani gruplar boyunca regresyon katsayıları eşittir – $b_1^* = b_2^* = b_3^* = \ldots = b^*$. Bu varsayım, Şekil 16.2 veya Şekil 16.3'teki regresyon doğrularının paralel olduğu anlamına gelmekte ve ayrı regresyon doğruları için bir regresyon doğrusu (grup içinde toplanmış regresyon doğrusu) yerleştirilmesini desteklemektedir. Kısa süre içerisinde görüleceği gibi, bu test edilebilir bir varsayımdır. Ortak değişken hakkında herhangi bir varsayımda bulunulmadığı göz ardı edilmemelidir. Ortak değişken, sabit veya tesadüfi bir değişken olabilir. (Bu bölümün önceki kesimlerinde yapıldığı gibi, değişkenin farklı düzeylerini temsil etmek için kukla değişkenler oluşturulursa ortak değişken kategorik bir değişken de olabilir.)

## Kovaryans Analizinin Hesaplanması

Çoklu regresyon çerçevesinde bakıldığında kovaryans analizi, ortak değişkenin kontrol altına alınmak istenmesi dışında temel olarak varyans analizinden farklı değildir. Cohen'ın

7 Doğrusal olmayan ilişkileri ele almaya yönelik yöntemler bulunmakla birlikte bu yöntemlere burada yer verilmemiştir.

(1968) ifade ettiği gibi, "Nihayet, kovaryans, araştırmanın maddî konularda mantıksal olarak dikte ettirmesi yoluyla, *Y*'nin varyansının hesaplanmasında, bağımsız değişkenler arasında bir öncelik varsayar." (s. 439).

Eğer ortak değişkeni (*C*) ayrıştırdıktan sonra *Y*'deki değişkenlik belirlenmek isteniyorsa ve *Y*'deki değişkenlik sadece *C* ile ilişkili ise, işlem etkisi (*a*) ve hata ile ilişkili oluyorsa $KT_{\text{regresyon}_{c,\alpha}}$ açıklanabilen toplam varyansı temsil eder. Eğer, $KT_{\text{regresyon}_{c,\alpha}}$ ve $KT_{\text{regresyon}_{c}}$'yi karşılaştırırsak, iki kareler toplamı arasındaki fark, ortak değişkenden kaynaklanan varyansın ötesinde işlem etkisinden kaynaklanan varyans olacaktır.

Conti ve Musty (1984) tarafından yapılan ve Bölüm 11'de verilen çalışmanın bir değişiği örnek olarak alınacaktır. Hatırlanacağı gibi, bu çalışmada araştırmacılar, esrardaki ana aktif madde olan THC'nin farklı miktarlarının oluşturduğu etkinin belirlenmesi ile ilgilenmişlerdir. Bu kapsamda THC doğrudan farelerin beyinlerine enjekte edilmiştir. Hareket sistemi etkinliği araştırmanın bağımlı değişkenini oluşturmuştur. Daha geleneksel yollardan THC uygulandığında hareket sistemi etkinliği artmaktadır. Deneysel çalışmaların doğasından dolayı (tüm hayvanlar ilk olarak bazal koşullar altında gözlenmiş ve THC uygulandıktan sonra hayvanların tümü tekrar gözlenmiştir), hayvanlar verilen ilâca alıştıktan sonra, etkinlik azalmaktadır. THC beyindeki ödül merkezi üzerinde bir etkiye sahip olduğundan, THC'nin orta dozu kısmen beklenen bu düşüşü telâfi etmekte ve orta-doz gruplarda, düşük veya yüksek-doz gruplara göre daha yüksek etkinliğe neden olur.

Conti ve Musty (1984) bu alandaki verinin analizinde izlenen rutin yolu izlemiş ve enjeksiyon sonrası etkinliği enjeksiyon öncesi etkinliğin bir fonksiyonu olarak analiz etmiştir. Alternatif bir yol da, ön enjeksiyon puanları kontrol altına alınarak son enjeksiyon puanlarına kovaryans analizi uygulanmasıdır. Bu yaklaşımda, deneklerin enjeksiyon öncesi etkinliklerinin enjeksiyon sonrası etkinlikleri üzerinde etkili olduğu esas alınmakta ve enjeksiyon öncesi farklılıklara düzeltme işlemi uygulanmaktadır. Bu yöntem sayesinde, grupların enjeksiyon öncesi etkinlik puanlarının enjeksiyon sonrası etkinlik puanlarına karışmasının önüne geçilmektedir.

Daha sonra açıklık kazanacağı gibi, tüm hayvanların gruplara tesadüfi olarak atandığına dikkat edilmesi önemlidir. Bu nedenle, enjeksiyon öncesi aşamada grup ortalamalarının eşit olması beklenir. Deneysel işlemin herhangi bir *etkisi yoksa*, enjeksiyon öncesi aşamada gözlenen farklılıklar için düzeltme uygulandıktan sonra, enjeksiyon sonrası ortalamalar eşit olacaktır. Deneklerin gruplara tesadüfi olarak atanması enjeksiyon sonrası ölçümlerde grupların düzeltilmiş ortalamalarının eşit olması beklentisine imkân tanır (eğer $H_0$ hipotezi doğru ise). Yine bu işlem enjeksiyon sonrası ölçümlerde grupların ortalamaları arasında fark varsa, bu farkın katılımcıların gruplara atanış şekli yerine deneysel işlemin etkisine bağlanmasına imkân tanır.

Conti ve Musty'nin (1984) çalışmasına ait veri ve ilgili tasarım matrisi Tablo 16.7'de görülmektedir. Sonuçlarda elde edilen kareler toplamı değerini daha kolay yorumlanabilir hale getirmek için ham veri basit bir şekilde 100'e bölünmüştür.[8] Veri kümesinden sonra yer alan tasarım matrisinde, her bir gruptaki sadece ilk ve son gözlem verilmiştir. *X*'in sütun 6'dan sütun 9'a kadar olan kısımları, ortak değişkenlerin etkileşimini ve grup değişkenlerini temsil etmektedir. Bu sütunlar, gruplar boyunca regresyon katsayılarının homojenliği hipotezini test etmek için kullanılmaktadır.

$$H_0: b_1^* = b_2^* = b_3^* = b_4^* = b_5^*$$

8 Eğer her bir veri 100'e bölünmeseydi, elde edilen kareler toplamı ve kareler ortalaması değerleri, burada sunulan değerlerin 10000 katı ($100^2 = 10000$) olacaktı. Diğer yandan, *F* ve *t* değerleri bu durumdan etkilenmemektedir.

**Tablo 16.7** Conti ve Musty'nin (1984) çalışması için enjeksiyon öncesi ve sonrası ölçümlere ilişkin veri kümesi

| | Kontrol | | 0,1 $\mu$ g | | 0,5 $\mu$ g | | 1 $\mu$ g | | 2 $\mu$ g | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Ön | Son | Ön | Son | Ön | Son | Ön | Son | Ön | Son |
| | 4,34 | 1,30 | 1,55 | 0,93 | 7,18 | 5,10 | 6,94 | 2,29 | 4,00 | 1,44 |
| | 3,50 | 0,94 | 10,56 | 4,44 | 8,33 | 4,16 | 6,10 | 4,75 | 4,10 | 1,11 |
| | 4,33 | 2,25 | 8,39 | 4,03 | 4,05 | 1,54 | 4,90 | 3,48 | 3,62 | 2,17 |
| | 2,76 | 1,05 | 3,70 | 1,92 | 10,78 | 6,36 | 3,69 | 2,76 | 3,92 | 2,00 |
| | 4,62 | 0,92 | 2,40 | 0,67 | 6,09 | 3,96 | 4,76 | 1,67 | 2,90 | 0,84 |
| | 5,40 | 1,90 | 1,83 | 1,70 | 7,78 | 4,51 | 4,30 | 1,51 | 2,90 | 0,99 |
| | 3,95 | 0,32 | 2,40 | 0,77 | 5,08 | 3,76 | 2,32 | 1,07 | 1,82 | 0,44 |
| | 1,55 | 0,64 | 7,67 | 3,53 | 2,86 | 1,92 | 7,35 | 2,35 | 4,94 | 0,84 |
| | 1,42 | 0,69 | 5,79 | 3,65 | 6,30 | 3,84 | | | 5,69 | 2,84 |
| | 1,90 | 0,93 | 9,58 | 4,22 | | | | | 5,54 | 2,93 |
| Ortalama | 3,377 | 1,094 | 5,387 | 2,586 | 6,494 | 3,906 | 5,045 | 2,485 | 3,943 | 1,560 |

| Tasarım Matrisi | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Kov | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $CT_1$ | $CT_2$ | $CT_3$ | $CT_4$ | | |
| | 4,34 | 1 | 0 | 0 | 0 | 4,34 | 0 | 0 | 0 | | 1,30 |
| | … | … | … | … | … | … | … | … | … | | … |
| | 1,90 | 1 | 0 | 0 | 0 | 1,90 | 0 | 0 | 0 | | 0,93 |
| | 1,55 | 0 | 1 | 0 | 0 | 0 | 1,55 | 0 | 0 | | 0,93 |
| | … | … | … | … | … | … | … | … | … | | … |
| | 9,58 | 0 | 1 | 0 | 0 | 0 | 9,58 | 0 | 0 | | 4,22 |
| | 7,18 | 0 | 0 | 1 | 0 | 0 | 0 | 7,18 | 0 | | 5,10 |
| **X** = | … | … | … | … | … | … | … | … | … | **Y** = | … |
| (47 × 9) | 6,30 | 0 | 0 | 1 | 0 | 0 | 0 | 6,30 | 0 | (47 × 1) | 3,84 |
| | 3,94 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3,94 | | 2,29 |
| | … | … | … | … | … | … | … | … | … | | … |
| | 7,35 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 7,35 | | 2,35 |
| | 4,00 | –1 | –1 | –1 | –1 | –4,00 | –4,00 | –4,00 | –4,00 | | 1,44 |
| | … | … | … | … | … | … | … | … | … | | … |
| | 5,54 | –1 | –1 | –1 | –1 | –5,54 | –5,54 | –5,54 | –5,54 | | 2,93 |

© Cengage Learning 2013

Tam model (etkileşim yordayıcılarını da içeren) aşağıdaki gibi yazılabilir:

$$Y_{ij} = \tau_j + c + c\tau_j + e_{ij}$$

Burada; $\tau_j$ j'yinci işlemin etkisini, $c$ ortak değişkenliği, $c\tau_j$ regresyon katsayılarının homojenliği testinin sonucunu ve $e_{ij}$ $j$'yinci işlemdeki i'yinci katılımcıya ait hatayı temsil eder.

İki model arasındaki $KT_{\text{regresyon}}$ değerleri değişimi (hata terimi için daha karmaşık modeldeki artık kullanılarak) veya $R^2$ değerindeki azalma temel alınarak iki model karşılaştırılabilir. Bu durumda ikinci işlem biraz daha basit olur.

Bu modele ilişkin regresyon analizi sonucunda,

$$R^2_{\tau,c,c\tau} = ,8238$$

elde edilir.

Grup içi regresyonları içinde manidar fark bulunmaması -yani, regresyon doğrularının paralel ve böylece her bir işlem için ayrı ayrı hesaplanabilen regresyon doğrularının eğimleri homojen ise -regresyonların homojenliği olarak adlandırılır- etkileşim teriminin silinmesi açıklanabilen varyans yüzdesinde önemsiz bir azalmaya neden olacaktır. *CT* terimleri silindiğinde,

$$R^2_{\tau,c} = ,8042$$

elde edilir. Bu eksilmeye ilişkin *F* testi, iki model arasındaki farka ait *F* testidir.

$$F(1-r, N-f-1) = \frac{(N-f-1)(R^2_{\tau,c,c\tau} - R^2_{\tau,c})}{(f-r)(1-R^2_{\tau,c,c\tau})}$$

$$= \frac{(47-9-1)(,8238-,8042)}{(4)(,1762)} = 1,03$$

4 ve 37 serbestlik derecelerinde 1,03'e eşit bir *F* değerinde beş grup için regresyonların homojenliği (ortak regresyon katsayıları) varsayımının reddedilmesi için bir nedenimiz yoktur. Buna göre, ortak değişken ile grup etkisi etkileşimini içermeyecek şekilde düzeltilen tüm model temelinde analize devam edebiliriz.

$$Y_{ij} = \mu + \tau_j + c + e_{ij}$$

Bu model, kendisini indirgenmiş model ile kıyaslayacağımız modele temel teşkil eder.

Çoklu regresyon analizine ait sonuçlar (1) ortak değişken ve kukla grup değişkenleri, (2) sadece grup etkisi değişkenleri ve (3) sadece ortak değişken kullanılarak üç grup şeklinde Tablo 16.8'de görülmektedir.

Tablo 16.8 incelendiğinde, hem ortak değişkeni (ön ölçümler) hem de grup üyeliğine ilişkin kukla değişkenleri ($T_1$,..., $T_4$) yordayıcı olarak içeren modelde regresyon için kareler toplamının ($KT_{\text{regresyon}}$) 82,6435 olduğu görülmektedir. Bu değer, hem grup etkileri hem de ortak değişken aracılığıyla açıklanabilen toplam değişkenliğe karşı gelir. Yine tabloya göre, artık kareler toplamı ($KT_{\text{artık}}$) değeri 20,1254'tür. Bu değer kestirilemeyen değişkenliği yansıtır. Buradaki analize ilişkin kovaryans özet tablosu, hata için kareler toplamıdır.

Kukla grup üyeliği değişkenleri eşitlikten çıkarılıp sadece ortak değişken (ön ölçümler) yordayıcı olarak kullanılırsa, $KT_{\text{regresyon}}$ değeri 82,6435'ten 73,4196'ya düşmektedir. Grup üyeliklerini içeren model ile sadece ortak değişkenin yordayıcı olarak kullanıldığı model için hesaplanan $KT_{\text{regresyon}}$ değerleri arasındaki fark, ortak değişkenin ötesinde grup etkisinden kaynaklanan kareler toplamı değerine karşılık gelir. Örnekteki veri kümesi için bu değer aşağıdaki gibi bulunur:

$$\begin{aligned} KT_{\text{grup(düzeltilmiş)}} &= KT_{\text{regresyon}_{\tau,c}} - KT_{\text{regresyon}_c} \\ &= 82,6435 - 73,4196 \\ &= 9,2239 \end{aligned}$$

Bu son değer, kovaryans analizi için *düzeltilmiş* grup kareler toplamı olarak adlandırılır. Çünkü ortak değişkenin etkisi için düzeltme işlemi yapılmıştır. Bu durumda, beş grubun ön ölçümleri açısından farklılık gösterdiği dikkate alınarak, denel işlem öncesi ölçümler arasındaki farklılıklardan kaynaklanan etkilerin kontrol altına alınabilmesi için düzeltme işlemi uygulanmıştır.

Kovaryans analizi özet tablosunun oluşturulabilmesi için, ortak değişkenden kaynaklanan kareler toplamı şeklinde ek bir terime daha ihtiyacımız olacaktır. Bu terimi tanımlamanın çeşitli yolları bulunmakla beraber, en yaygın yöntem düzeltilmiş işlem etkisinin tanımlanmasına benzer bir yol izlemektir. Değişkenliğin işlem etkisiyle tanımlanamayan

**Tablo 16.8** Regresyon analizi

**(a) Tam Model**

$$\widehat{Y}_{ij} = 0{,}4347(\text{Ön}) - 0{,}5922(T_1) + 0{,}0262(T_2) + 0{,}8644(T_3) + 0{,}0738(T_4) + 0{,}2183$$

$$R^2_{\tau \cdot c} = {,}8042$$

**Regresyon İçin Varyans Analizi Özet Tablosu**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Regresyon | 5 | 82,6435 | 16,5287 | 33,6726 |
| Artık | 41 | 20,1254 | 0,4909 | |
| Toplam | 46 | 102,7689 | | |

**(b) İndirgenmiş Model-Grup Yordayıcıları Atlanarak**

$$\widehat{Y}_{ij} = {,}5311(\text{Ön}) - {,}26667$$

$$R^2_{c} = {,}7144$$

**Regresyon İçin Varyans Analizi Özet Tablosu**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Regresyon | 1 | 73,4196 | 73,4196 | 112,5711 |
| Artık | 45 | 29,3493 | 0,6522 | |
| Toplam | 46 | 102,7689 | | |

**(c) İndirgenmiş Model-Ortak Değişken (Ön Ölçümler) Çıkarılarak**

$$\widehat{Y}_{ij} = -1{,}2321(T1) + 0{,}2599\,(T2) + 1{,}5794(T3) + 0{,}1589(T4) + 2{,}3261$$

$$R^2_{\tau} = {,}4311$$

**Regresyon İçin Varyans Analizi Özet Tablosu**

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Regresyon | 4 | 44,3028 | 11,0757 | 7,9564 |
| Artık | 42 | 58,4661 | 1,3921 | |
| Toplam | 46 | 102,7689 | | |

© Cengage Learning 2013

kısmını ortak değişkene bağlayacağız. Bir diğer deyişle modeli, hem ortak değişken hem de işlem yordayıcılarıyla birlikte ele alacağız ve bunu sadece işlem yordayıcılarının bulunduğu model ile karşılaştıracağız. İki kareler toplamı arasındaki fark, işlem etkilerinin ötesinde ortak değişkenin açıklayabildiği kareler toplamıdır.[9] Buradaki veri kümesi için

$$\begin{aligned} KT_{\text{ortak değişken}} &= KT_{\text{regresyon}_{\tau,c}} - KT_{\text{regresyon}_{\tau}} \\ &= 82{,}6435 - 44{,}3028 \\ &= 9{,}2239 \end{aligned}$$

Şimdi, kovaryans analizine ait özet tabloyu oluşturmak için her bilgi elimizdedir. Kovaryans analizi için özet tablo Tablo 16.9'da verilmiştir. Bu tabloda, tam modeldeki hata terimi olan $KT_{artık}$ ile diğer kareler toplamı değerlerinin daha önce hesaplandığı gibi olduğuna dikkat edilmelidir. Ortak değişken için serbestlik derecesinin "1" olduğuna dikkat

9 Tüm programlar elde edilen çıktıları bu şekilde düzenlememektedir. Bu nedenle, farklı bir $KT_{\text{ortak değişken}}$ değeri ile karşılaşılması halinde şaşırılmamalıdır.

**Tablo 16.9** Kovaryans analizi için özet tablo

**Tek Yönlü Kovaryans Analizi İçin Genel Özet Tablo**

| Kaynak | *sd* | *KT* |
|---|---|---|
| Ortak Değişken | c | $KT_{\text{regresyon}_{(\tau,\text{c})}} - KT_{\text{regresyon}_{(\tau)}}$ |
| İşlem (düzeltilmiş) | $k - 1$ | $KT_{\text{regresyon}_{(\tau,\text{c})}} - KT_{\text{regresyon}_{(c)}}$ |
| Hata | $N - k - 1$ | $KT_{\text{regresyon}_{(\tau,\text{c})}}$ |
| Toplam | $N - 1$ | |

**Tablo 16.7'deki Veri İçin Özet Tablo**

| *Kaynak* | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Ortak Değişken | 1 | 38,3407 | 38,3407 | 78,108* |
| İşlem (düzeltilmiş) | 4 | 9,2239 | 2,3060 | 4,698* |
| Hata | 41 | 20,1254 | 0,4909 | |
| Toplam | 46 | 102,7689 | | |

Tam Model:

$$\hat{Y}_{ij} = 0{,}4347(Pre) - 0{,}5922(T_1) + 0{,}0262(T_2) + 0{,}8644(T_3) + 0{,}0738(T_4) + 0{,}2183$$

*$p < ,05$

© Cengage Learning 2013

edilmelidir. Çünkü bir ortak değişken vardır. Düzeltilmiş grup etkisi için serbestlik derecesi $(k - 1) = (5 - 1) = 4$ ve hata için serbestlik derecesi $N - k - c = 41$'dir ($k$ grup sayısını $c$ ise ortak değişken sayısını temsil eder).

Özet tablo incelendiğinde, $KT_{\text{işlem(düzeltilmiş)}} = 9{,}2239$ olduğu görülmektedir. Serbestlik derecesi 4 olduğundan, $KO_{\text{grup(düzeltilmiş)}} = 2{,}3060$ olur. Bu değerin $KO_{\text{hata}} = 0{,}4909$ değerine bölünmesiyle $F = 4{,}698$ elde edilir. Bu $F$ değeri (4,41) sd'de $p < ,05$ düzeyinde manidardır. Buna göre, enjeksiyon öncesi ölçümlerdeki etkinlikler için düzeltme uygulandıktan sonra, deneklerin enjeksiyon sonrası ölçüm puanlarında manidar fark bulunduğu ve uygulanan deneysel işlemin etkinlik üzerinde manidar bir etkiye sahip olduğu sonucuna varırız.

## Düzeltilmiş Ortalamalar

$F_{,05}\,(4,41) = 2.61 < F_{\text{hesaplanan}} = 4{,}698$ olduğundan;

$H_0$: $\mu_1$(düzeltilmiş) = $\mu_2$(düzeltilmiş) = $\mu_3$(düzeltilmiş) = $\mu_4$(düzeltilmiş) = $\mu_5$(düzeltilmiş) şeklindeki sıfır hipotezi reddedildi. Buna göre, ortak değişkenin etkisi kontrol altına alındıktan sonra, işlemlere ait ortalamalar arasında manidar fark bulunduğu sonucuna varılır. Bu farkı yorumlamak için ortak değişkenin etkisine göre düzeltilmiş grup ortalamalarının elde edilmesi gerekli değilse de, yararlıdır. Temel olarak; eğer grupların enjeksiyon öncesi ortalamaları arasında fark olmazsa, enjeksiyon sonrası ortalamalarının ne olacağının kestirilmesiyle ilgileniyoruz. Ortak değişkenin ve işlem değişkeninin yordayıcı olarak kullanıldığı regresyon çözümü aracılığıyla düzeltilmiş ortalamalar kolaylıkla elde edilebilir.

Revize edilmiş tam modelin analizinden, aşağıdaki değerler elde edilir (Tablo 16.8'e bakınız).

$$\widehat{Y}_{\text{ij}} = 0{,}4347\ (\text{Ön}) - 0{,}5922(T1) + 0{,}0262(T2) + 0{,}8644(T3) + 0{,}0738(T4) + 0{,}2183$$

Bu eşitlik, ortalamalar ve düzetilmiş ortalamaları temsil eden $\bar{Y}'_j$ açısından yazıldığında aşağıdaki eşitlik elde edilir:

$$\widehat{Y}'_j = 0{,}4347(\overline{Ön}) - 0{,}5922(T1) + 0{,}0262(T2) + 0{,}8644(T3) + 0{,}0738(T4) + 0{,}2183$$

Bu eşitlikte, $\overline{Ön} = 4{,}8060$ olup (enjeksiyon öncesi ölçümlerin ortalaması); $T_1$, $T_2$, $T_3$ ve $T_4$ ( 0, 1, -1) değişkenlerdir. (Ön ölçüm ortalamaları her bir ön ölçümden çıkarılır. Çünkü tüm deneklerin ortak değişken açısından ortalama puana sahip olduklarında *Y* için düzeltilmiş ortalamanın kestirimiyle ilgilenmekteyiz.) Örnekteki veri için işlemlerin düzeltilmiş ortalamaları:

$$\widehat{Y}'_1 = 0{,}4347(4{,}8060) - 0{,}5922(1) + 0{,}0262(0) + 0{,}8644(0) + 0{,}0738(0) + 0{,}2183$$
$$= 1{,}7153$$

$$\widehat{Y}'_2 = 0{,}4347(4{,}8060) - 0{,}5922(0) + 0{,}0262(1) + 0{,}8644(0) + 0{,}0738(0) + 0{,}2183$$
$$= 2{,}3336$$

$$\widehat{Y}'_3 = 0{,}4347(4{,}8060) - 0{,}5922(0) + 0{,}0262(0) + 0{,}8644(1) + 0{,}0738(0) + 0{,}2183$$
$$= 3{,}1719$$

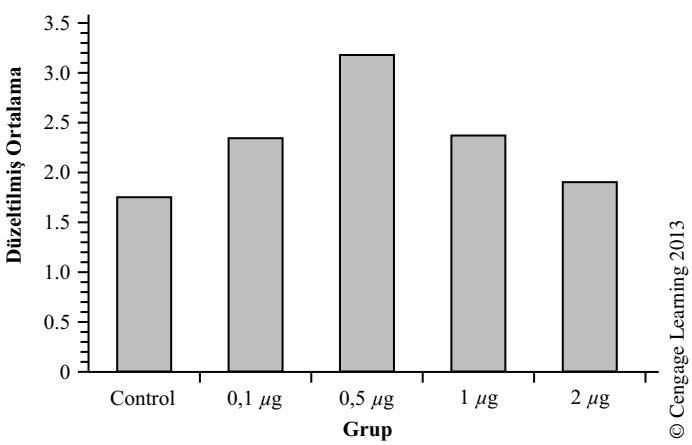
$$\widehat{Y}'_4 = 0{,}4347(4{,}8060) - 0{,}5922(0) + 0{,}0262(0) + 0{,}8644(0) + 0{,}0738(1) + 0{,}2183$$
$$= 2{,}3813$$

$$\widehat{Y}'_5 = 0{,}4347(4{,}8060) - 0{,}5922(-1) + 0{,}0262(-1) + 0{,}8644(-1) + 0{,}0738(-1) + 0{,}2183$$
$$= 1{,}9353.$$

Düzeltilmiş ortalamaların grafiği Şekil 16.4'te çizilmiştir.

Genel ortalama:

$$\bar{Y}'_{\cdot} = 0{,}4347(4{,}8060) - 0{,}5922(0) + 0{,}0262(0) + 0{,}8644(0) + 0{,}0738(0) + 0{,}2183$$
$$= 2{,}3075.$$




**Şekil 16.4** Gruba göre düzeltilmiş ortalamalar

Bu değer düzeltilmiş ortalamaların ortalamasıdır. (Örneklem büyüklükleri eşit olduğunda düzeltmiş genel ortalama, düzeltilmemiş genel ortalamaya eşittir.)[10]

Formüller açısından daha derin sulara dalmak ve gözlükleri takmak üzereyiz. Bu formüllerin okuyucu tarafından ezberlenmesine gerek yoktur. Önemli olan ne olup bittiğinin anlaşılması ve ihtiyaç duyulduğunda bu formülleri bulmak için nereye bakmak gerektiğinin hatırlanmasıdır. Bu formüllerin, istatistik programlarının çoğunda rapor edilen çıktılar arasında bulunması beklenilmemelidir.

Şimdi bu düzeltilmiş ortalamalar kullanılarak işlemler arasındaki karşılaştırmalar yapılabilir. Bununla birlikte bu durumda, genel kovaryans analizindeki hata terimini uyarlamak zorundayız. Eğer $KT_{\text{hata(ortak değişken)}}$ değerinin ortak değişken üzerindeki varyans analizinde hata kareleri toplamını temsil etmesine izin verilirse -Huitema (1980) kovaryans analizine ilişkin son derece okunaklı ve başarılı bir kaynaktır- belirtildiği gibi iki düzeltilmiş ortalama arasındaki farka ait bir test elde edilir. Huitema (1980), iki düzeltilmiş ortalama arasındaki farka ait test istatistiğini aşağıdaki ifadeyle verir:

$$F(1, N-a-1) = \frac{(\overline{Y}'_j - \overline{Y}'_k)^2}{KO'_{\text{hata}}\left[\left(\frac{1}{n_j}+\frac{1}{n_k}\right)+\frac{(C_j - C_k)^2}{KT_{\text{hata (ortak değişken)}}}\right]}$$

$KO'_{\text{hata}}$

Yukardaki formülde $\boldsymbol{KO'_{\text{hata}}}$ kovaryans analizindeki hata terimidir. Etkili hata terimlerine ve ortalamalar arası karşılaştırmalara ilişkin mükemmel bir tartışma için Winer (1971, s. 771) ve özellikle Huitema'ya (1980) bakınız. Huitema'nın yakın zamanda çıkması planlanan yeni bir basımı bulunmaktadır.

Bir örnek olarak; teoride büyük farklılık göstermesi beklenen $\overline{Y}'_1$ ve $\overline{Y}'_3$ değerlerini karşılaştırmak istediğimizi varsayalım. Önceki analizden, bilebiliriz veya hesaplayabiliriz.

$$KO'_{\text{hata}} = 0{,}4909$$
$$KT_{\text{hata(ortak değişken)}} = 202{,}938 \quad \text{(hesaplama gösterilmemiştir.)}$$
$$\overline{C}_1 = 3{,}3770 \qquad \overline{C}_3 = 6{,}4944$$
$$\overline{Y}'_1 = 1{,}7153 \qquad \overline{Y}'_3 = 3{,}1719$$

$$F(1,41) = \frac{(1{,}7153 - 3{,}1719)^2}{0{,}4909\left[\left(\frac{1}{10}+\frac{1}{9}\right)+\frac{(3{,}3770 - 6{,}4944)^2}{202{,}938}\right]} = \frac{2{,}1217}{0{,}1271} = 16{,}69$$

Kritik $F$ değeri; $F_{.05}(1,41) = 4{,}08$'dir. Hesaplanan $F$ değeri kritik $F$ değerinden büyük olduğundan, "iki işlemin ortalamalarının evrende eşit olduğu" şeklinde ifade edilen sıfır hipotezi reddedilir. Enjeksiyon öncesi ölçümlerdeki farklılıklar kontrol altına alındıktan sonra bile, enjeksiyon sonrası ölçümler arasında manidar fark olduğu sonucuna ulaşırız.

Gösterim 16.4 varyans analizi için SPSS çıktılarını içermektedir. (Öntest ve Sontest ortalamaları "Compare means" işlemi kullanılarak hesaplanmıştır.) Seçenekler menüsünden, "spread versus level" grafiğinin istendiğine dikkat ediniz. Bu grafikler grup varyanslarına karşı grup ortalamalarını gösterir. Varyanslar homojen ve ortalamalar bağımsızsa, bu grafik tesadüfi bir görünüme sahip olmalıdır. Grafik, ortalamaların büyüklüğü ile varyansların büyüklüğü arasında bir korelasyon olduğunu ortaya koymaktadır. Ancak, düzeltilmiş ortalamalar ile bu ortalamaların standart hataları arasındaki grafik çizildiğinde, ilişkinin önemli ölçüde azaldığı gözden kaçırılmamalıdır.

---

10 Düzeltmiş ortalamaların hesaplanmasına alternatif bir yaklaşım aşağıdaki şekilde tanımlanır:

$$\overline{Y}'_j = \bar{Y}_j - b_w(\overline{C}_j - \overline{C})$$

Burada, $\overline{C}_j$ değeri Grup j için ortak değişken ortalaması, $\overline{C}$ ortak değişken genel ortalaması ve $b$w tam modelde ortak değişkene ilişkin regresyon katsayıdır (burada, $b_w = 0{,}4347$). Düzeltilmiş ortalamaların hesaplanmasına ilişkin daha geleneksel bu yaklaşım, düzetilmiş ortalamaların grubunda ortak değişkenden sapmasının bir fonksiyonu olduğunu açık hale getirmektedir. İki yaklaşımdan hangisi kullanılırsa kullanılsın, düzeltilmiş ortalamalara ilişkin elde edilen değerler aynıdır.

Rapor

| Grup | | ÖNTEST | SONTEST |
|---|---|---|---|
| Kontrol | Ortalama | 3,3770 | 1,0940 |
| | N | 10 | 10 |
| | Standart Sapma | 1,3963 | ,5850 |
| 0,1 $\mu$g | Ortalama | 5,3870 | 2,5860 |
| | N | 10 | 10 |
| | Standart Sapma | 3,4448 | 1,5332 |
| 0,5 $\mu$g | Ortalama | 6,4944 | 3,9056 |
| | N | 9 | 9 |
| | Standart Sapma | 2,3781 | 1,4768 |
| 1 $\mu$g | Ortalama | 5,0450 | 2,4850 |
| | N | 8 | 8 |
| | Standart Sapma | 1,6876 | 1,1874 |
| 2 $\mu$g | Ortalama | 3,9430 | 1,5600 |
| | N | 10 | 10 |
| | Standart Sapma | 1,2207 | ,8765 |
| Toplam | Ortalama | 4,8060 | 2,2857 |
| | N | 47 | 47 |
| | Standart Sapma | 2,3788 | 1,4947 |

Denekler (Katılımcılar)-Arası Etki Testleri

Bağımlı Değişken: SONTEST

| Kaynak | Tip III Kareler Toplamı | sd | Kareler Ortalaması | F | Sig. | Eta Kare |
|---|---|---|---|---|---|---|
| Düzeltilmiş Model | 82,644[a] | 5 | 16,529 | 33,673 | ,000 | ,804 |
| Sabit | ,347 | 1 | ,347 | ,707 | ,405 | ,017 |
| ÖNTEST | 38,341 | 1 | 38,341 | 78,108 | ,000 | ,656 |
| GRUP | 9,224 | 4 | 2,306 | 4,698 | ,003 | ,314 |
| Hata | 20,125 | 41 | ,491 | | | |
| Toplam | 348,327 | 47 | | | | |
| Düzeltilmiş Toplam | 102,769 | 46 | | | | |

[a] $R^2$ = ,804 (Düzeltilmiş $R^2$ = ,780)

**Kestirilen Marjinal Ortalamalar**

İşlem Grubu

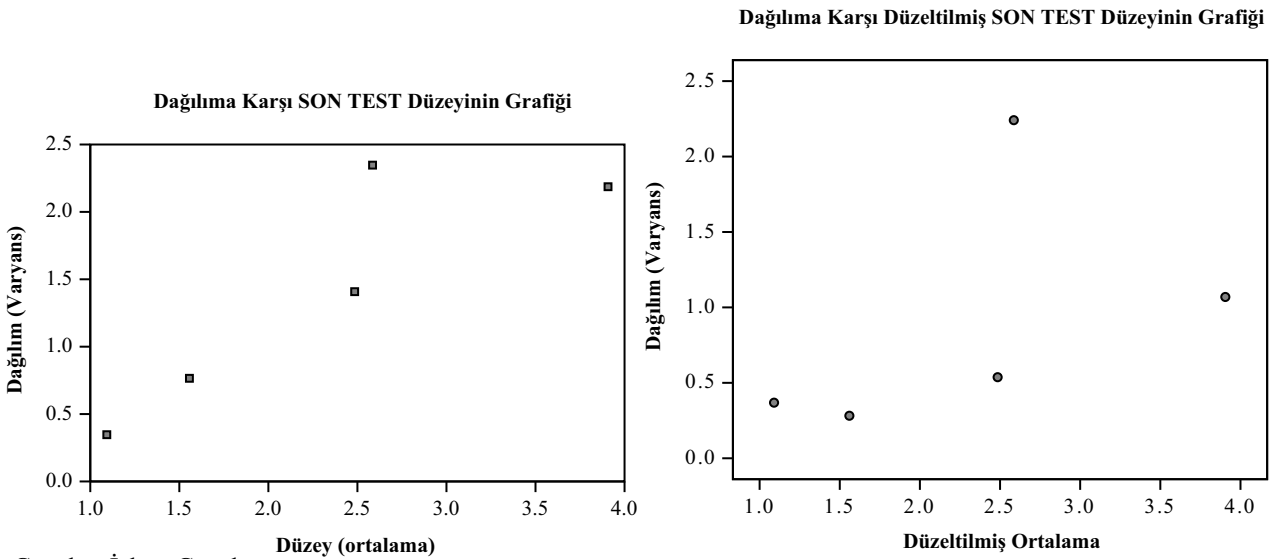
Bağımlı Değişken: Sontest

| İşlem Grubu | Ortalama | Standart Hata | %95 Olasılıklı Güven Aralığı | |
|---|---|---|---|---|
| | | | Alt Sınır | Üst Sınır |
| Kontrol | 1,715[a] | ,232 | 1,246 | 2,185 |
| 0,1 $\mu$g | 2,333[a] | ,223 | 1,882 | 2,785 |
| 0,5 $\mu$g | 3,172[a] | ,248 | 2,671 | 3,672 |
| 1 $\mu$g | 2,381[a] | ,248 | 1,880 | 2,882 |
| 2 $\mu$g | 1,935[a] | ,226 | 1,480 | 2,391 |

SPSS Inc.'dan uyarlanmıştır..

**Gösterim 16.4** Conti ve Musty'nin çalışması için SPSS analiz çıktıları

*(devam edecek)*




Gruplar: İşlem Grupları

**Gösterim 16.4** *(Devamı)*

## 16.6 Kovaryans Analizinde Etki Büyüklüğünün Hesaplanması

Beklenildiği gibi, etki büyüklüğünün hesaplanması kovaryans analizinde varyans analizindekine kıyasla biraz daha karmaşık bir işlemdir. Çünkü kovaryans analizinde, karşılaştırılan ortalamalar ve kullanılan hata terimleri için seçim yapılması gerekmektedir. Hatırlanacağı gibi, faktöriyel desenler ile tekrarlı ölçüm desenlerinde etki büyüklüğü için hata teriminin seçimi ile ilgili benzer problemler söz konusuydu.

Daha önce olduğu gibi, etki büyüklüğüne *r*-ailesi ve *d*-ailesi açısından bakılacaktır. Normal şartlarda, *F* testine bakılırken *r*-ailesinin; belirli karşılaştırmalara bakıldığında ise *d*-ailesinin kullanılması önerilmektedir. Burada ilk olarak *r*-ailesinden bir örnek ile başlanacak ve daha sonra *d*-ailesinden bir örnek ile devam edilecektir. Kullanılacak örnek Conti ve Musty'nin çalışmasına dayanmakta olup bu çalışmaya ait veri kümesi geniş kapsamlı sıfır hipotezinde manidar bir *F* değeri üretmektedir. Muhtemelen bu örnek hakkında konuşmanın en uygun yolu, 0'dan 2 $\mu$g'ye artan Grup metrik değişkeninden faydalanılmasını sağlayacaktır.[11] Bununla birlikte, burada "ikinci en uygun" yaklaşım alınacaktır. Çünkü birçok çalışmada buradaki ısmarlama değişken gibi düzgün dağılıma sahip olan bir bağımsız değişken yoktur.

### *r*-Ailesi Ölçümü

*r*-ailesi ölçümü olarak $\eta^2$ değeri kullanılacaktır. Bu değerin pozitif yönlü olduğu kabul edilmektedir. $\eta^2$'nin gruba karşılık gelen *KT* değerinin toplam *KT* değerine bölünmesiyle elde edildiği hatırlanmalıdır. Ancak gruba karşılık gelen *KT* değerlerinden hangisinin kullanılacağının belirlenmesi gerekir. Daha açık bir anlatımla, bağımlı değişken için varyans analizindeki *KT* değerinin mi yoksa kovaryans analizindeki *KT* değerinin mi kullanılaca-

11 SPSS düzeltilmiş ortalamalar üzerinden çoklu karşılaştırmaları test etmektedir. Sadece CONTRAST butonu tıklanmalı ve çoklu karşılaştırmalar talep edilmelidir. Buradaki örnek için anlamlı bir kuadratik bileşen vardır.

ğına karar verilmelidir. Kline (2004), bu iki alternatifin de kullanılabileceğini belirtmekle birlikte, bir öneri olarak dile getirmeden kovaryans analizindeki[12] $KT_{\text{toplam}}$ değerini kullanmaktadır. Eğer ortak değişken evrende doğal olarak değişiyorsa (farklı hayvanların ön test puanları açısından değişiklik göstermesi beklendiği için buradaki örnekte olduğu gibi), kovaryans analizinden elde edilen $KT_{\text{işlem}}$ değerinin bu analizde hesaplanan $KT_{\text{toplam}}$ (düzeltilmemiş) değerine bölünmesi son derece mantıklı bir uygulama olmaktadır. Bu işlem sonucunda, bağımsız değişken tarafından açıklanabilen "normal değişkenlik" yüzdesine karşılık gelen $\eta^2$ değeri elde edilmektedir.[13] Buna göre;

$$\eta^2 = \frac{KT_{\text{işlem(düzeltilmiş)}}}{KT_{\text{toplam}}} = \frac{9{,}2239}{102{,}7689} = {,}09 \text{ olur.}$$

Aynı cevabı üreten alternatif bir yaklaşım; bağımlı değişkenin sadece ortak değişken tarafından yordandığı model ile bağımlı değişkenin hem ortak değişken (ön test) hem de bağımlı değişken tarafından yordandığı model arasındaki $R^2$ farkı olarak $\eta^2$'nin alınmasıdır. İlk modelin ardından açıklanan varyans miktarı açısı için ikinci modelde gözlenen artış ortak değişkenin kontrol altına alınmasından sonra grup etkisini temsil etmektedir. Buradaki örnek için sadece ortak değişken kullanıldığında $R^2$ değeri ,417 olarak hesaplanmaktadır. (Bu değer, ortak değişkenin bağımsız değişken olarak kullanıldığı bir varyans analiziyle ya da ortak değişkenin bağımsız değişken olarak alındığı bir regresyon analiziyle elde edilebilir.) Grup etkisi analize dahil edildiğinde, $R^2$ değeri ,804 olmaktadır. Bu değerler aşağıdaki tabloda gösterilmiştir:

| Adım | Yordayıcılar | $R^2$ | $R^2$'deki değişim | $F$'deki değişim |
|---|---|---|---|---|
| 1 | Ön test | ,714 | | |
| 2 | Ön Test, İşlem | ,804 | ,090 | 4,689 |

$\eta^2$ değeri bu iki $R^2$ değeri arasındaki farka eşit olup ortak değişken kontrol edildikten sonra bağımsız değişkenin (işlem etkisinin) açıklanan varyans miktarına katkısını gösterir.

## *d*-Ailesi Ölçümü

*d*-ailesi ölçümleri genellikle daha yorumlanabilir niteliktedir ve herhangi iki ortalamanın karşılaştırılmasında daha sık kullanılır. Burada kullanılan örnekteki bağımsız değişken sürekli bir yapıda olduğundan, söz konusu örnek iki ortalamanın karşılaştırılması için çok uygun değildir. Ancak burada örnek olarak kontrol grubu ile 0,5 $\mu$g grubu arasındaki karşılaştırmayı kullanacağım. Çünkü Conti ve Musty'nin teorisine göre en yüksek farkın bu iki grup arasında görülmesi beklenmektedir. Kovaryans analizi ile çalıştığımızdan, karşılaştırmada kullanılması uygun olan ortalamalar düzeltilmiş ortalama ($\overline{Y}_i$) değerleridir. 0,5 $\mu$g ve kontrol grubu için düzeltilmiş ortalama değerleri sırasıyla 3,1719 ve 1,7153'tür. (Hatırlayacağınız gibi, önceki kısımda bu iki ortalama arasındaki farkın manidarlığına ilişkin bir test uygulamış ve farkın manidar olduğunu saptamıştık.)

Genellikle, etki büyüklüğü ölçümü olarak,

$$d = \frac{\hat{\omega}}{\hat{\sigma}}$$

eşitliğini kullanırız. İki grubun ortalamasını karşılaştırken, katsayılar [–1, 0, 1, 0, 0] olduğundan $\hat{\psi}$ basit bir şekilde iki ortalama arasındaki fark olmaktadır. Burada ele alınan örnek için $\hat{\psi}$ = 3,1719 – 1,7153 = 1,4566'dır. Fakat daha önemli olan soru, standart sapmanın kestirimi için

12 Eğer etki büyüklüğü istenirse, SPSS de aynı değeri kullanmaktadır. Bu değer basit bir şekilde, $KT_{\text{işlem}} + KT_{\text{hata}}$ toplamıdır.

13 Ön test kontrol altında tutulduktan sonra, doz ve etkinlik düzeyi arasındaki kuadratik ilişki için $\eta^2$ değeri ile ilgilenilirse, sadece $KT_{\text{kuadratik}}$ değerinin $KT_{\text{toplam}}$ değerine bölünmesi yeterlidir.

hangi değerin kullanılacağıdır. Bu konudaki çeşitli seçeneklerden biri, varyans analizindeki $KO_{\text{hata}}$ değerinin karekökü olabilir. Çünkü bu karekök değeri her bir grup içindeki ortalama değişkenliğe ilişkin bir kestirimdir ve bundan dolayı orijinal ölçümlerin biriminde ortalama farkı standartlaştırmaktadır. ($\eta^2$ değeri hesaplanırken varyans analizindeki $KT_{\text{toplam}}$ değerinin kullanıldığı hatırlanmalıdır.) Kovaryans analizinde $KO_{\text{hata}}$ değerinin karekökü, bu konuda bir alternatiftir. Düzeltilmiş puanlar biriminde ortalama farkını standartlaştıran bu değerin anlaşılması biraz daha zordur. Cortina ve Nouri (2000), buradaki örnekte yaptığımız gibi, ortak değişkenin evrende normal olarak değiştiği durumlarda bu değişkenliğin hatanın kestirimine dâhil edilmek isteneceği şeklinde mantıklı bir öneri ileri sürmüştür. Bu öneri, son test puanlarına ilişkin varyans analizindeki $KO_{\text{hata}}$ değerinin karekökünün kullanılacağı anlamına gelmektedir. Buradaki örnek için $KO_{\text{hata}} = 1{,}392$ olup bu değer grup varyanslarının ağırlıklandırılmış ortalamasıdır. Buradan, aşağıdakiler elde edilir:

$$d = \frac{\hat{\psi}}{\hat{\sigma}} = \frac{\overline{X}_3 - \overline{X}_1}{\sqrt{KO_{\text{hata}}}} = \frac{3{,}1719 - 1{,}7153}{\sqrt{1{,}392}} = \frac{1{,}4566}{1{,}1798} = 1{,}23$$

THC'nin en uygun dozda ($0{,}5\ \mu_g$) enjeksiyonu, enjeksiyon sonrası etkinlikte "1" birimlik ve kontrol grubuna göre çeyrek standart sapmalık bir artış ile sonuçlanmaktadır.

## 16.7 Kovaryans Analizi Sonuçlarının Yorumlanması

Kovaryans analizi sonuçlarının yorumlanmasında, verinin doğası ve daha önemlisi deneysel işlemin tasarımına bağlı olarak bazı sorunlar ortaya çıkar. Huitema (1980), bu tartışmaların çoğu hakkında okunaklı ve eksiksiz bir kaynaktır. Anderson (1963), Evans ve Anastasio (1968), Huitema (2005), Lord (1967, 1969), Maxwell ve Cramer (1975), Reichardt (1979), Smith (1957) ve Weisberg (1979) bu problemlerin ele alındığı diğer önemli kaynaklardır.

Kovaryans analizinin ideal uygulaması, katılımcıların gruplara (veya faktöriyel desenlerdeki hücrelere) tesadüfi olarak atandığı deneylerdir. Bu durumda, deneysel işlem uygulanmadan önce ortak değişkenin ölçüldüğü varsayılarak her bir grup veya hücre için ortak değişken ortalamasının beklenen değeri aynıdır ve gözlenebilecek herhangi bir farklılık sadece şansa bağlanabilir. Kovaryans analizi öncelikle, hata terimini azaltır. Ayrıca ortak değişken için gruplar arasında şansa bağlı olarak gözlenen farklılıkların bağımlı değişken ortalamalarında sebep olduğu yanlılıkları ortadan kaldırır. Bu Conti ve Musty'nin (1984) çalışmasında tartışılan durumdur.

Ortak değişkenin deneysel işlemden *sonra* ölçüldüğü ve uygulanan deneysel işlemin ortak değişkeni etkilediği bir çalışmada, tesadüfî desen, kovaryans analizi sonuçlarının yorumlanması son derece güçtür. Böyle bir durumda, katılımcılar gruplara tesadüfi olarak atansalar bile, grup ortak değişken ortalamalarının beklenen değerleri eşit değildir. Analiz sonuçlarının yorumlanması zordur. Çünkü gerçekte ortak değişken farklılıkları uygulanan işlemden dolayı oluşan etkilerin ayrılmaz bir parçası olduğu halde, ortak değişken açısından farklılık göstermemeleri durumunda grupların nasıl olacağı sorusuna cevap aranmaktadır. Ortak değişkenin hatalı ölçülmesi durumunda (yani tam olarak güvenilir değilse), bu problem çok daha ciddi bir sorun olur. Bu durumda yorumlama ile ilgili diğer sorunların üstesinden gelmek için **gerçek puan kovaryans analizi** olarak adlandırılan alternatif bir analiz daha uygun olabilir. Huitema'da (1980, Bölüm 14) böyle bir analize yer verilmiştir.

**gerçek puan kovaryans analizi**

Katılımcılar gruplara tesadüfi olarak atanmadığında, kovaryans analizi sonuçlarının yorumlanması daha zor olur. Bu problemin en yaygın örneği, **eşit olmayan grup deseni** olarak adlandırılır. Bu desende iki veya daha fazla müdahale edilmemiş grup seçilir (örneğin, öğrencilerin okulları veya sınıfları), her iki gruptaki katılımcılardan ön teste ilişkin ölçümler elde edilir, gruplardan birine denel işlem uygulanır, daha sonra iki gruptan son teste ait elde edilen ölçümler karşılaştırılır. Denekler gruplara tesadüfi olarak atanmadığından, ön testte bulunan farklılıkların şanstan kaynaklandığını varsayabilmek için bir da-

**eşit olmayan grup deseni**

yanak yoktur. Benzer şekilde, denel işlemin bir etkisinin olmaması halinde iki grubun son test puanlarının eşit olacağı beklentisine ilişkin bir dayanak da bulunmamaktadır. Huitema (1980, s. 149); gruplar denel işlem öncesinde farklılık gösterdiğinde, deneysel işlemin bir etkisi olmasa bile ölçümlerde görülen ortalamaya kayma eğiliminin son test puanlarında farklılığa yol açabildiğini etkili bir biçimde gösteren bir örneğe yer vermiştir. Ortalamaya kayma eğilimi, ortalamadan aşırı derecede düşük veya aşırı derecede yüksek olan ölçümler ikinci kez ölçüldüğünde bu ölçümlerin ortalamaya daha yakın değerlere sahip olacağını ifade eden istatiksel bir kavramdır. Bu şartlar altında yararlı olabilecek alternatif bir analiz için Huitema'ya (1980) başvurulabilir. Maris (1998) bu konuya biraz daha farklı bir bakış açısı getirmektedir.

Katılımcıların gruplara tesadüfi olarak atanmadığı desenlerde analiz sonuçlarının yorumlanmasına ilişkin problemlerin üstesinden gelmek kolay değildir. Bu durum tesadüfi atamanın tesadüfi seçimden neden daha önemli olduğunun gerekçelerinden biridir. Hem verinin yorumlanması hem de değişkenler arasındaki nedensel ilişkiler hakkında çıkarım yapılabilmesi açsısından tesadüfi atamanın üstün yönlerini abartmak zordur. Lord'un (1967) muhtemelen hafif abartı ile söylediği gibi, "yazarın düşüncesinde, bu çalışmalardaki veri kümelerinde genellikle olduğu gibi, gruplar arasındaki önceden var olan ve kontrol edilemeyen farklılıklarla ilgili mantıksal veya istatistiksel bir açıklama göz önünde bulundurulamaz." (s. 305). (Lord tesadüfi atamadan dolayı oluşan farklılıkları kastetmemektedir.) Anderson (1963), aşağıdaki ifadesiyle benzer noktaya değinmiştir: Bir kimse, "veri mevcut durumdaki gibi olmasaydı tam olarak nasıl olabilirdi sorusunun" anlamını merak edebilir (s. 170). Bunların tamamı grupların ortak değişken açısından farklılık gösterdiği veri kümelerinin analizinde kovaryans analizinin yeri olmadığı anlamına gelmez. Bununla birlikte, kovaryans analizini kullanan herhangi bir araştırmacının elindeki veri ve çıkardığı sonuçların pratikteki geçerliği hakkında dikkatli bir biçimde düşünmesi gerekir.

## 16.8 Kovaryans Analizi Sonuçlarının Raporlaştırılması

Varyans analizi ile kovaryans analizi sonuçlarının ifade edilişi arasındaki tek fark, kovaryans analizinde ortak değişkene ve düzeltilmiş ortalamalara başvurulmasıdır. Conti ve Musty tarafından gerçekleştirilen deney için aşağıdakiler yazılabilir:

> Conti ve Musty (1984), THC'nin farelerde etkinlik etkisini incelemişlerdir. Araştırmada orta doz THC'nin etkinlikte en fazla artışı (veya maddeye uyum sağlamadan dolayı en az düşüş) kestirmişlerdir. Enjeksiyon öncesi ölçümlerden sonra, beş farklı gruptaki farelere tesadüfî olarak 0 $\mu_g$, 0,1 $\mu_g$, 0,5 $\mu_g$, 1 $\mu_g$, 2 $\mu_g$ THC enjekte edilmiştir. Etkinlik düzeyi, enjeksiyon sonrası 10 dakika aralıklarla ölçülmüştür. Ön test etkinlikleri arasında önemli farklılıklar bulunduğundan, ön test puanları araştırmada ortak değişken olarak kullanılmıştır.
>
> Kovaryans analizi sonuçları manidar bulunmuştur [$F_{(4,41)} = 4{,}694$, $p < {,}003$]. Analiz sonuçlarına göre, orta doz THC'nin etkinlik artışında en yüksek etkiye sahip olduğu belirlenmiştir. Ortak değişkene göre düzeltilmemiş $KT_{toplam}$ değeri kullanılarak Eta Kare değeri ,09 bulunmuştur. Kontrol grubunun ortalaması ile 0,5 $\mu_g$ alan grubun ortalamalarının karşılaştırılmasıyla, [$F_{(1,41)} = 16{,}69$, $p < {,}05$] değeri manidar bulunmuş ve standartlaştırılmış etki büyüklüğü ($\hat{d}$) 1,23 olarak hesaplanmıştır.

## 16.9 Faktöriyel Kovaryans Analizi

Kovaryans analizi tek değişkenli desenlere uygulanabildiği gibi çok değişkenli desenlere de uygulanabilir. Metodolojik hususlardan dolayı, ortak değişken bir kez daha analizde öncelik üstlenen bir değişken olarak işlem görür. Bu kısımda sadece eşit hücre büyüklüklerinin olduğu durumlar ele alınacaktır. Bununla birlikte, eşit olmayan örneklem büyüklüklerine genellenmesi anlık bir iştir.

Analizin mantığı basittir ve daha önceki örneklere benzer. $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}}$ ortak değişken, *A*, *B* ana etkileri ve *A* ile *B* etkileşiminin doğrusal kombinasyonundan kaynaklanan değişkenliktir. Benzer şekilde, $KT_{\text{regresyon}_{c,\alpha,\beta}}$ ortak değişken ile *A* ve *B* ana etkilerinden kaynaklanan varyans miktarıdır. Bu iki kareler toplamı arasındaki fark;

$$KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\alpha,\beta}}$$

ortak değişken ile ana etkiler kontrol altına alındıktan sonra *AB* etkileşiminden kaynaklanan değişkenliğe karşılık gelir. Eşit örneklem büyüklüklerinde iki ana etki ile etkileşim dik olduğundan, bu desenlerde tamamı gerçekten kontrol altında tutulabilen değişken ortak değişkendir.

Aynı mantıkla,

$$KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\alpha,\alpha\beta}}$$

ortak değişken kontrol altına alındıktan sonra *B* değişkeninden kaynaklanan değişkenliği temsil ederken;

$$KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\beta,\alpha\beta}}$$

ortak değişkenin kontrol altına alınmasının ardından *A* temel etkisinden kaynaklanan değişkenliği temsil eder. Hata terimi ise *A*, *B* ana etkilerinin, *AB* etkileşiminin ve ortak değişkenin kontrol altına alınmasından arta kalan değişkenlik olup

$$KT_{\text{artık}_{c,\alpha,\beta,\alpha\beta}}$$

ile gösterilir. Analizin genel yapısı Tablo 16.10'da görülmektedir. Burada, hata teriminin her bir ortak değişken için bir serbestlik derecesi kaybettiğine dikkat edilmelidir. Bağımsız değişken ile ortak değişkenin açıkladıkları varyansın çakışan kısımları bulunduğundan, bu ikisinin kareler toplamı $KT_{\text{toplam}}$ değerine eşit değildir.

Bir örnek olarak Spilich ve ark.'nın (1992) Bölüm 13'te yer verdiğimiz, sigara içmenin bir fonksiyonu olarak performans çalışmasından yararlanacağız. Çalışmada, katılımcılar üç gruba ayrılmıştır: Birinci gruptakiler (Aktif Kullanıcılar), görev sırasında veya görevden hemen önce sigara içenler; ikinci gruptakiler Gecikmiş Kullanıcılar (GK) (Delayed Smoking (DS)), üç saatten beri sigara içmeyenler; üçüncü gruptakiler de Sigara İçmeyenler (NonSmoking). Görev sırasında yapılan hata sayısı araştırmanın bağımlı değişkenini temsil etmektedir. Bu örneği kovaryans analizine uygun hale getirmek için katılımcıların dikkat dağınıklığı şeklindeki hipotetik bir değişken çalışmaya dahil edilmiştir. (Yüksek dikkat dağınıklığı puanları, katılımcıların dikkatlerinin daha kolay dağılabildiğini gösterir.)

Veri kümesi Tablo 16.11'de sunulmuştur. Bu tablodaki veri bir ortak değişken (dikkat dağınıklığı) içeren 3×3 şeklinde bir faktöriyel deseni temsil eder.

Tablo 16.12 tasarım matrisinin kısaltılmış halini içermektedir. Bu tablo, her bir hücre için sadece hücredeki ilk ve son katılımcıya ilişkin girdileri göstermektedir. Matris; ortak değişken için bir sütun (*C* ile gösterilmiştir), alışılmış tasarım matrisi bileşenleri olan ana etkiler için Görev (T), Grup etkisi için (G) ve Görev x Grup etkileşimi için TG sütunlarını içermektedir. Ayrıca, ortak değişken ile Grup × Görev etkileşimini içeren sütunlar

**Tablo 16.10** Faktöriyel desenler için kovaryans analizinin yapısı

| Kaynak | *sd* | *KT* |
|---|---|---|
| $A_{\text{düzeltilmiş}}$ | $a - 1$ | $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\beta,\alpha\beta}}$ |
| $B_{\text{düzeltilmiş}}$ | $b - 1$ | $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\alpha,\alpha\beta}}$ |
| $AB_{\text{düzeltilmiş}}$ | $(a - 1)(b - 1)$ | $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{c,\alpha,\beta}}$ |
| Hata | $N - ab - c$ | $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}}$ |
| Kovaryans | $c$ | $KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta}} - KT_{\text{regresyon}_{\alpha,\beta,\alpha\beta}}$ |
| Toplam | $N - 1$ | |

© Cengage Learing 2013

**Tablo 16.11** Sigara kullanımı ve performansa ilişkin hipotetik veri (Spilich ve ark., 1992)

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Örüntü Tanıma** | | | | | | | | | | | | | | | | |
| Kullanmayanlar | Hatalar | 9 | 8 | 12 | 10 | 7 | 10 | 9 | 11 | 8 | 10 | 8 | 10 | 8 | 11 | 10 |
| | Dik. Dağ. | 107 | 133 | 123 | 94 | 83 | 86 | 112 | 117 | 130 | 111 | 102 | 120 | 118 | 134 | 97 |
| Gecikmiş Kullanıcılar | Hatalar | 12 | 7 | 14 | 4 | 8 | 11 | 16 | 17 | 5 | 6 | 9 | 6 | 6 | 7 | 16 |
| | Dik. Dağ. | 101 | 75 | 138 | 94 | 138 | 127 | 126 | 124 | 100 | 103 | 120 | 91 | 138 | 88 | 118 |
| Aktif Kullanıcılar | Hatalar | 8 | 8 | 9 | 1 | 9 | 7 | 16 | 19 | 1 | 1 | 22 | 12 | 18 | 8 | 10 |
| | Dik. Dağ. | 64 | 135 | 130 | 106 | 123 | 117 | 124 | 141 | 95 | 98 | 95 | 103 | 134 | 119 | 123 |
| **Bilişsel Görev** | | | | | | | | | | | | | | | | |
| Kullanmayanlar | Hatalar | 27 | 34 | 19 | 20 | 56 | 35 | 23 | 37 | 4 | 30 | 4 | 42 | 34 | 19 | 49 |
| | Dik. Dağ. | 126 | 154 | 113 | 87 | 125 | 130 | 103 | 139 | 85 | 131 | 98 | 107 | 107 | 96 | 143 |
| Gecikmiş Kullanıcılar | Hatalar | 48 | 29 | 34 | 6 | 18 | 63 | 9 | 54 | 28 | 71 | 60 | 54 | 51 | 25 | 49 |
| | Dik. Dağ. | 113 | 100 | 114 | 74 | 76 | 162 | 80 | 118 | 99 | 146 | 132 | 135 | 111 | 106 | 96 |
| Aktif Kullanıcılar | Hatalar | 34 | 65 | 55 | 33 | 42 | 54 | 21 | 44 | 61 | 38 | 75 | 61 | 51 | 32 | 47 |
| | Dik. Dağ. | 108 | 191 | 112 | 98 | 128 | 145 | 76 | 107 | 128 | 128 | 142 | 144 | 131 | 110 | 132 |
| **Sürüş Simülasyonu** | | | | | | | | | | | | | | | | |
| Kullanmayanlar | Hatalar | 3 | 2 | 0 | 0 | 6 | 2 | 0 | 6 | 4 | 1 | 0 | 0 | 6 | 2 | 3 |
| | Dik. Dağ. | 130 | 83 | 91 | 92 | 109 | 106 | 99 | 109 | 136 | 102 | 119 | 84 | 68 | 67 | 114 |
| Gecikmiş Kullanıcılar | Hatalar | 7 | 0 | 6 | 0 | 12 | 17 | 1 | 11 | 4 | 4 | 3 | 5 | 16 | 5 | 11 |
| | Dik. Dağ. | 93 | 102 | 108 | 100 | 123 | 131 | 99 | 116 | 81 | 103 | 78 | 103 | 139 | 101 | 102 |
| Aktif Kullanıcılar | Hatalar | 15 | 2 | 2 | 14 | 5 | 0 | 16 | 14 | 9 | 17 | 15 | 9 | 3 | 15 | 13 |
| | Dik. Dağ. | 110 | 96 | 112 | 114 | 137 | 125 | 168 | 102 | 109 | 111 | 137 | 106 | 117 | 101 | 116 |

© Cengage Learning 2013

da (CTG) ekledim. Bu sütunlar, tüm *i* ve *j* değerleri için $H_0$: $b_i = b_j$ hipotezini test etmek amacıyla kullanılacaktır. Çünkü regresyonların homojenliği varsayımı tüm kovaryans analizleri için geçerlidir.

Sadece ortak değişken içeren etkileşimlerin neyi temsil ettiğinin dikkate alınması önemlidir. $CT_1$ ve $CT_2$ terimlerini ekleseydim ve bu terimleri kullanarak Dikkat Dağınıklığının bir fonksiyonu olarak Hataların regresyon doğrularının eşit olduğuna ilişkin sıfır hipotezini test edebilirdim. Benzer şekilde, $CG_1$ ve $CG_2$ terimlerini ekleseydim, her üç regresyon grubundaki regresyonların homojenliğini test edebilirdim. Dokuz hücrenin her birinde regresyonların homojenliği hipotezinin test edilmesi ile daha fazla ilgilendiğimden, sadece $CTG_{ij}$ terimleri eklenmiştir.

İlk regresyon *X*'teki tüm yordayıcılara dayalıdır. Bu analizden aşağıdaki eşitlikler bulunur.

$$KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta,c\alpha\beta}} = 36724{,}771$$

$$KO_{\text{kalan}_{c,\alpha,\beta,\alpha\beta,c\alpha\beta}} = 71{,}134$$

Görev × Grup etkileşiminin ortak değişken (Dikkat Dağınıklığı) ile etkileşimine ait terim çıkarıldığında, kareler toplamına ait aşağıdaki değer elde edilir.

$$KT_{\text{regresyon}_{c,\alpha,\beta,\alpha\beta,}} = 36389{,}60175$$

Bu iki kareler toplamı arasındaki fark 335,16925'e eşittir. İlk modele ait *sd* 13 iken ikinci modelin serbestlik derecesi 9 olup iki kareler toplamı arasındaki farka *ait sd* 13 – 9 = 4'tür. Buna göre kareler toplamı farkına ait değer kareler ortalamasına dönüştürülebilir.

$$KO_{\text{fark}} = 335{,}16925/4 = 83{,}79231$$

**Tablo 16.12** İçme verisi için kovaryans analizi tasarım matrisi

| | $C$ | $T_1$ | $T_2$ | $G_1$ | $G_2$ | $TG_{11}$ | $TG_{12}$ | $TG_{21}$ | $TG_{22}$ | $cTG_{11}$ | $cTG_{12}$ | $cTG_{21}$ | $cTG_{22}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 107 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 107 | 0 | 0 | 0 | | 9 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 97 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 97 | 0 | 0 | 0 | | 10 |
| | 101 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 101 | 0 | 0 | | 12 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 118 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 118 | 0 | 0 | | 16 |
| | 64 | 1 | 0 | -1 | -1 | -1 | -1 | 0 | 0 | -64 | -64 | 0 | 0 | | 8 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 123 | 1 | 0 | -1 | -1 | -1 | -1 | 0 | 0 | -123 | -123 | 0 | 0 | | 10 |
| | 126 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 126 | 0vv | | 27 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 143 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 143 | 0 | | 49 |
| | 113 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 113 | | 48 |
| $X=$ | … | … | … | … | … | … | … | … | … | … | … | … | … | $Y=$ | … |
| | 96 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 96 | | 49 |
| | 108 | 0 | 1 | -1 | -1 | 0 | 0 | -1 | -1 | 0 | 0 | -108 | -108 | | 34 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 132 | 0 | 1 | -1 | -1 | 0 | 0 | -1 | -1 | 0 | 0 | -132 | -132 | | 47 |
| | 130 | -1 | -1 | 0 | 0 | -1 | 0 | 1 | 0 | -130 | 0 | -130 | 0 | | 3 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 114 | -1 | -1 | 1 | 0 | -1 | 0 | 1 | 0 | -114 | 0 | -114 | 0 | | 3 |
| | 93 | -1 | -1 | 0 | 1 | 0 | 0 | 0 | -1 | 0 | -93 | 0 | -93 | | 7 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 102 | -1 | -1 | | 0 | 0 | 0 | 0 | -1 | 0 | -102 | 0 | -102 | | 11 |
| | 110 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 110 | 110 | 110 | 110 | | 15 |
| | … | … | … | … | … | … | … | … | … | … | … | … | … | | … |
| | 116 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 116 | 116 | 116 | 116 | | 13 |

Bu iki model arasındaki fark, daha kapsamlı tam modeldeki $KO_{\text{kalan}}$ değerinden yararlanarak aşağıdaki gibi test edilebilir:

$$F = \frac{KO_{fark}}{KO_{kalan}} = \frac{83,792}{71,134} = 1,18$$

Bulunan $F$ değerine ilişkin $sd$, $[(f - r), (N - f - 1)] = 4$ ve 21'dir. İlgili serbestlik derecesindeki kritik $F$ değeri ise; $F_{,05}(4,121) = 2,45$'tir. Hesaplanan $F$ değeri kritik $F$ değerinden küçük olduğundan regresyonların homojenliğine ilişkin yokluk hipotezini reddedemeyiz. Dikkat dağınıklığı ile ilgili regresyon doğrularının dokuz hücrede aynı eğime sahip olduğuna dair şüphe duymak için bir gerekçe olmadığı sonucuna varırız. Bu, izleyen indirgenmiş modelleri karşılaştırmak için daha basit bir tam modeli benimsememize imkân tanır. Basitleştirilmiş tam model;

$$\hat{Y} = b_0 + b_1 C + b_2 T_1 + b_3 T_2 + b_4 G_1 + b_5 G_2 + b_6 TG_{11} + b_7 TG_{12} + b_8 TG_{21} + b_9 TG_{22}$$

veya daha klâsik varyans analizi terimleriyle

$$Y_{\text{ijk}} = \mu + C_k + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

Kovaryans analizi için ihtiyaç duyulan çeşitli çoklu regresyon çözümleri Tablo 16.13'te gösterilmiştir. Tam modeller ile indirgenmiş modeller arasındaki fark hesaplanarak ve test edilerek, kovaryans analizinin tamamı hesaplanabilir.

Gösterim 16.5 bu veri kümelerine ilişkin SPSS analiz sonuçlarını göstermektedir. Analizin uygulanması durumunda elde edilecek sonuçları, bu gösterimdeki sonuçlarla karşılaştırmanız gerekir.

**Tablo 16.13** Tablo 16.11'deki veri için çeşitli modellere ait regresyon analizi sonuçları

| Model | $KT_{regresyon}$ | $KO_{kalan}$ | $R^2$ |
|---|---|---|---|
| *C, T, G, TG* | 36.389,602 | 71,539 | ,803 |
| *C, T, G* | 35.154,816 | | ,775 |
| *C, G, TG* | 12.519,117 | | ,276 |
| *C, T, TG* | 35.416,789 | | ,781 |
| *T, G, TG* | 31.744,726 | | ,700 |

© Cengage Learning 2013

**Gruplar-Arası Etki Testleri**

Bağımlı Değişken: HATALAR

| Kaynak | Tip III KarelerToplamı | sd | Kareler Ortalaması | F | Mnd. | Eta Kare | Merkezi olm. parametre | Gözlenen Güç[a] |
|---|---|---|---|---|---|---|---|---|
| Düzeltilmiş Model | 36389,602[b] | 9 | 4043,289 | 56,519 | ,000 | ,803 | 508,671 | 1,000 |
| Kesim nok. | 892,395 | 1 | 892,395 | 12,474 | ,001 | ,091 | 12,474 | ,939 |
| DİK. DAĞ. | 4644,876 | 1 | 4644,876 | 64,928 | ,000 | ,342 | 64,928 | 1,000 |
| GÖREV | 23870,485 | 2 | 11935,243 | 166,836 | ,000 | ,727 | 333,673 | 1,000 |
| Uyuş. Mad Grup | 563,257 | 2 | 281,629 | 3,937 | ,022 | ,059 | 7,873 | ,699 |
| GÖREV × SMKGRUP | 1626,511 | 4 | 406,628 | 5,684 | ,000 | ,154 | 22,736 | ,977 |
| Hata | 8942,324 | 125 | 71,539 | | | | | |
| Toplam | 90341,000 | 135 | | | | | | |
| Düzeltilmiş Toplam | 45331,926 | 134 | | | | | | |

[a] Alfa kullanılarak hesaplanmıştır = ,05

[b] $R^2$ = ,803 (Düzeltilmiş $R^2$ = ,789)

**1. Görev * Madde Kullanımı Grubu**

Bağımlı Değişken: HATALAR

| Görev | Sigara İçme Grupları | Ortalama | Std. Hata | %95 Güven Aralığı | |
|---|---|---|---|---|---|
| | | | | Alt Sınır | Üst Sınır |
| Örüntü Tanıma | Kullanmayanlar | 9,805[a] | 2,184 | 5,482 | 14,128 |
| | Gecikmiş Kullanıcılar | 9,732[a] | 2,184 | 5,410 | 14,054 |
| | Aktif Kullanıcılar | 9,558[a] | 2,184 | 5,235 | 13,882 |
| Bilişsel | Kullanmayanlar | 27,770[a] | 2,188 | 23,440 | 32,101 |
| | Gecikmiş Kullanıcılar | 40,436[a] | 2,185 | 36,112 | 44,760 |
| | Aktif Kullanıcılar | 43,785[a] | 2,233 | 39,366 | 48,204 |
| Sürüş | Kullanmayanlar | 8,505[a] | 2,191 | 4,169 | 12,842 |
| | Gecikmiş Kullanıcılar | 8,921[a] | 2,200 | 4,568 | 13,275 |
| | Aktif Kullanıcılar | 5,820[a] | 2,226 | 1,414 | 10,226 |

[a] Modelde görünen ortak değişkenler değerlendirilmiştir: Dikkat Dağınıklığı = 112,52

*(Devam ediyor.)*

SPSS In.c'dan adapte edilmiştir.

**Gösterim 16.5** Spilich verisi için SPSS programından elde edilen kovaryans analizi çıktıları

**2. Görev**

Bağımlı Değişken: HATALAR

| Görev | Ortalama | Standart Hata | % 95 Güven Aralığı | |
|---|---|---|---|---|
| | | | Alt Sınır | Üst Sınır |
| Örüntü Tanıma | 9,699[a] | 1,261 | 7,203 | 12,194 |
| Bilişsel | 37,330[a] | 1,274 | 34,810 | 39,851 |
| Sürüş | 7,749[a] | 1,273 | 5,230 | 10,268 |

[a] Modelde görünen ortak değişken değerlendirilmiştir: DİKKAT DAĞINIKLIĞI = 112,52

**3. İçici Grubu**

Bağımlı Değişken: HATALAR

| İçme Grubu | Ortalama | Standart Hata | % 95 Güven Aralığı | |
|---|---|---|---|---|
| | | | Alt Sınır | Üst Sınır |
| İçmeyenler | 15,360[a] | 1,264 | 12,859 | 17,862 |
| Gecikmiş içiciler | 19,696[a] | 1,266 | 17,191 | 22,202 |
| Aktif içiciler | 19,721[a] | 1,261 | 17,225 | 22,217 |

[a] Modelde görünen ortak değişken değerlendirilmiştir: DİKKAT DAĞINIKLIĞI = 112,52

**Gösterim 16.5** *(Devamı)*

Karşılaştırma amacıyla Gösterim 13.1'deki varyans analizi sonuçları aşağıda verilmiştir. Bu sonuçlar, ortak değişken olmaması durumunda aynı veri kümesi için elde edilen varyans analizi bulgularını gösterir.

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Görev | 2 | 28.661,526 | 14.330,763 | 132,895* |
| Grup | 2 | 1813,748 | 906,874 | 8,410* |
| Görev × Grup | 4 | 1269,452 | 317,363 | 2,943* |
| Hata | 126 | 13.587,200 | 107,835 | |
| Toplam | 134 | 45.331,926 | | |

*p* < ,05

Bu analizde görev değişkenine ait manidar bir etkiye sahip olduğumuza dikkat ediniz. Bu ilginç değildir, çünkü görevler birbirinden oldukça farklıdır ve bazı görevlerin diğerlerine göre daha fazla hataya yol açması beklenir. Grup değişkenin de etkisi manidardır, fakat etkileşim terimi ile ilgili bulgular daha fazla dikkatimizi çekmektedir. İçiciler grubu belirli durumlarda fark yaratırken (daha yoğun bir şekilde bilişsel süreç gerektiren işlemlerde), diğer bazı durumlarda fark oluşturup oluşturmadığının belirlenebilmesi amacıyla Görev × Grup etkileşimine de bakılmıştır. Ayrıca, kovaryans analizinde 71,539 olan $KO_{\text{hata}}$ değerinin burada 107,835 olduğu da gözden kaçırılmamalıdır.

Kovaryans analizine baktığımızda görülen ilk şey, $KO_{\text{hata}}$ (71,539) değerinin varyans analizindekine kıyasla üçte bir oranında daha küçük olduğudur. Bu durum, ortak değişkenin (Dikkat Dağınıklığı) varyans analizi ile Hatada açıklanamayan değişkenliğin önemli bir kısmını açıklayabilmesinden (bu kısım varyans analizi tarafından açıklanmamaktadır) kaynaklanmaktadır.

Gösterim 16.5'te etkiler için *kısmi* Eta Kare değeri görülmektedir. Bu etki-büyüklüğü değeri ölçümü, iki $R^2$ değeri arasındaki fark ($1 - R^2_{\text{indirgenmiş}}$) değerine bölünerek hesaplanabilir. Örneğin; Görev için kukla değişkeni içermeyen modelde $R^2$ = ,276 olup açıklanamayan varyans oranı %72,4'tür [1 – ,276 = ,724]. Görev değişkeni modele eklendiğinde (tam model olmaktadır), $R^2$ = ,803 olarak elde edilir. İki model için hesaplanan $R^2$ değerleri arasındaki fark, ,803 – ,276 = ,527'dir. Bu değer; *önceki modelde açıklanamayan varyansın* %72,8'lik bir kısmına denk düşmektedir [,527 / ,724 = %72,8]. SPSS bu değerden kısmi $\eta^2$ olarak söz etmese de bu değer Gösterim 16.5'te Görev değişkeni için verilen değerdir. Benzer hesaplamalar, diğer değerleri de yeniden üretecektir.

## Düzeltilmiş Ortalamalar

Düzeltilmiş ortalamaların elde edilmesi yöntemi, Conti ve Musty'nin örneğinde kullanılan yöntemin bir uzantısıdır. Eğer işlem kombinasyonları ortak değişken açısından farklılık göstermezse hücre ortalamalarının ne olacağı sorusuna cevap bulmak isteriz.

Tam modelden, elimizde aşağıdaki gibi bir eşitlik bulunmaktadır:

$$\hat{Y} = b_0 + b_1 C + b_2 T_1 + b_3 T_2 + b_4 G_1 + b_5 G_2 + b_6 TG_{11} + b_7 TG_{12} + b_8 TG_{21} + b_9 TG_{22}$$

Buna ait eşitlik aşağıdaki gibi yazılabilir:

$$\hat{Y} = -14{,}654 - 8{,}561\, T_1 + 19{,}071 T_2 - 3{,}794 G_1 + 1{,}437 G_2 + 3{,}901 TG_{11} - 1{,}404 TG_{12} - 5{,}766 TG_{21} + 1{,}668\, TG_{22} + 0{,}293 \quad \text{Dikkat Dağınıklığı}$$

Grupların ortak değişken açısından farklılık göstermemesi durumunda $Y$ ortalamasının ne olacağını bilmek isteğidimizden, tüm işlemler için $C = \bar{C}$ = 112,518 olarak alacağız.

Hücre$_{11}$'deki tüm gözlemler için $C$'nin $\bar{C}$ ile değiştirilmesiyle tasarım matrisinin uygun satırı,

1 0 1 0 1 0 0 0 112,518

Regresyon katsayıları uygulanıp sabit dikkate alındığında,

$$\begin{aligned}\hat{Y} &= -14{,}654 - 8{,}561(1) + 19{,}071(0) - 3{,}794(1) + 1{,}437(0) + 3{,}901(1) - 1{,}404(0) \\ &\quad - 5{,}766(0) + 1{,}668(0) + 0{,}293(112{,}518) \\ &= 5{,}860\end{aligned}$$

Bu işlem tüm hücrelere uygulandığında, aşağıdaki tabloda görülen düzeltilmiş hücre ortalamaları elde edilir:

| | Örüntü Tanıma | Bilişsel | Sürüş | Satır Ortalamaları |
|---|---|---|---|---|
| Kullanmayanlar | 9,805 | 27,770 | 5,820 | 14,465 |
| Gecikmiş kullanıcılar | 9,732 | 40,436 | 8,921 | 19,696 |
| Aktif Kullanıcılar | 9,558 | 43,785 | 8,505 | 20,616 |
| Sütun Ortalamaları | 9,699 | 37,330 | 7,749 | 18,259 |

© Cengage Learning 2013

Bu değerler, gösterim 16,5'te verilen hücre ortalamalarıdır. Satır ve sütün ortalamaları ilgili satır ve sütunda bulunan hücrelerdeki değerlerin ortalaması alınarak hesaplanabilir.

## Düzeltilmiş Ortalamaların Test Edilmesi

Düzeltilmiş ortalamalar Şekil 16.5'te gösterilmiştir. Bu şekil, etkileşimi ve ana etkilere atfedilebilecek anlamı gösterir. İçici olmaktan dolayı oluşan farklılıklar, üst düzey bilişsel süreç gerektiren durumlar ile sınırlı göründüğünden bu veri kümesiyle daha fazla analiz yapılması muhtemelen gereksizdir. Bununla birlikte bütünlüğün sağlanabilmesi amacıyla,

Kullanmayanlar grubu ile Aktif ve Gecikmiş Kullanıcılar gruplarına ait ortalamaların birleşimi arasında bir karşılaştırma yapılmasını isteyebileceğinizi varsayacağız. Bu durumda, $\overline{X}_1'$ ile $\overline{X}_2'$ ve $\overline{X}_3'$'ün kombinasyonu karşılaştırılacaktır. Ortak değişkendeki farklılıkların hesaplanabilmesi için bu karşılaştırma hata teriminde bazı değişikliklerin yapılmasını gerektirir. Winer (1971) tarafından yapılan düzeltme aşağıdaki gibidir:

$$KO_{hata}^{m} = KO_{hata}^{l}\left[1 + \frac{\frac{KT_{g(c)}}{g - 1}}{KT_{hata(c)}}\right]$$

Bu eşitlikteki $KT_{g(c)}$ ve $KT_{hata\,(c)}$ varyans analizinde sırasıyla *ortak değişkene* ait grup değişkeni ve hata teriminden kaynaklanan kareler toplamlarını temsil eder. $KO_{hata}'$ da genel kovaryans analizindeki hata terimini yansıtır. Ezberlenmesi kolay olmayan bu formülün hatırlanması için bir neden görmüyorum.

Örneğimiz için,

$KT_{hata}' = 71{,}538$

$KT_{g(c)} = 2701{,}215$

$KT_{hata\,(c)} = 54285{,}867$

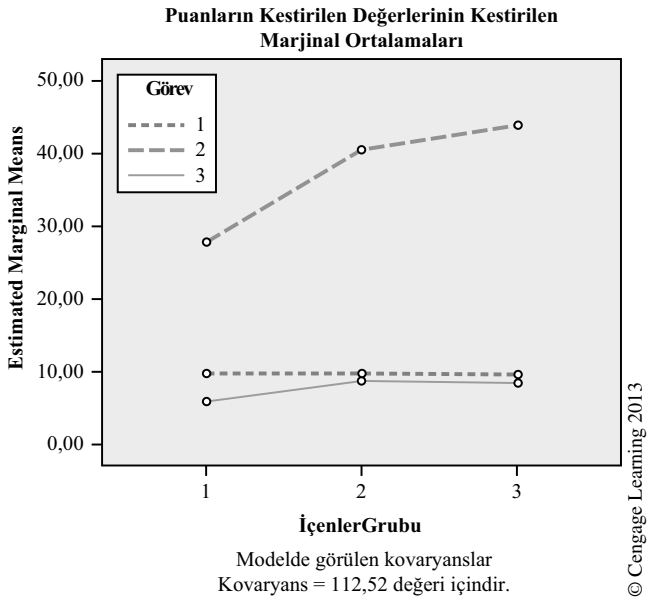
böylece;

$$KO_{hata}'' = 71{,}538\left[1 + \frac{\frac{2701{,}215}{2 - 1}}{54285{,}867}\right] = 75{,}098$$

bulunur. Düzeltilmiş ortalamaları karşılaştırmak için aşağıdaki eşitlikten yararlanırız:

$$\psi = 2(14{,}465) - 1(19{,}696) - 1(20{,}616) = -11{,}382$$

$$F(1, 125) = \frac{n\psi^2}{\sum \alpha_i^2 KO_{hata}''} = \frac{48(-11382)^2}{6(75{,}098)} = 12{,}938$$




**Şekil 16.5** Görev ve Grup değişkenlerinin bir fonksiyonu olarak düzeltilmiş hücre ortalamaları

$F_{,05}(1,25) = 3,92$ olduğundan yokluk hipotezini reddedebilir ve diğer iki grubun ortalaması ile karşılaştırdığımızda Aktif İçiciler grubunun daha kötü bir performans ortaya koyduğu sonucuna varabiliriz.

Bir başka araştırmacı, sadece Bilişsel görev için Grup etkisinin belirlenmesiyle ilgilenebilir. Bu basit etki incelenirken yine, bir şekilde hata teriminin değiştirilmesi gerekir. Grup etkisine, verinin sadece bir kısmı için bakılması ve Bilişsel görev katılımcılarının ortak değişken ortalamasının tüm katılımcıların ortak değişken ortalamasından farklılık göstermesi, hata teriminde değişiklik yapılmasını gerekli kılan faktördür. Muhtemelen buradaki en emin yol, sadece Bilişsel görevi gerçekleştiren katılımcılar için ayrı bir kovaryans analizinin uygulanmasıdır. Bu yöntemin hataya ilişkin serbestlik derecesi gibi bir dezavantajı olmasına rağmen, hata teriminin düzeltilmesinde gerekli olan rahatsız edici varsayımları ortadan kaldırması gibi bir avantajı vardır.

Testler hakkındaki tartışmayı tamamlamak için iki düzeltilmiş hücre ortalamasını (bu hücrelerin aynı satır veya sütunda olması önemli değildir) karşılaştırmak isteyen bir araştırmacı düşünelim. Bu karşılaştırma için düzeltilmiş hata terimi Winer (1971) tarafından aşağıdaki gibi verilmiştir:

$$KO''_{\text{hata}} = \frac{2KO'_{\text{hata}}}{n}\left[1 + \frac{\dfrac{KT_{\text{hücreler(c)}}}{tg - 1}}{KT_{\text{hata(c)}}}\right]$$

Buradaki $KT_{\text{hücreler(c)}}$ ortak değişkene ait varyans analizindeki kareler toplamıdır.

Karşılaştırma yaparken hata terimine düzeltme uygulanması konusunda neden sürekli olarak endişe duyduğumuzu merak edebilirsiniz. Bölüm 9'daki regresyon doğrusunun etrafındaki güven aralığının nasıl göründüğü (eğilimin gerçekte eliptik bir yapısı vardı) hatırlandığında cevabın genel yapısı açık hale gelir. $X = \bar{X}$ olduğunda $\hat{Y}$ hakkında nispeten eminizdir. Bununla birlikte $X$, $\bar{X}$'dan uzaklaştıkça yapılan kestirime ait güven azalmakta ve güven aralığı genişlemektedir. Şimdi tekrar Şekil 16.3'teki grafiğe dönersek, problemin doğrudan düzeltilmiş ortalamalardan geldiğini göreceksiniz. Bu grafikte, $\bar{Y}'_1$ değeri $\bar{Y}_1$ 'den oldukça uzaktır ve doğru kestirim yapıldığına dair güven muhtemelen düşüktür. Diğer taraftan; $\bar{Y}'_2$ için yapılan kestirmede kabul edilebilir düzeyde bir güven vardır. Sadece bu tür bir düşünce, hata terimine sürekli olarak düzeltme uygulanmasına neden olmaktadır.

Burada kullanılan örnek, kovaryans analizinin önemli bir özelliğine işaret etmektedir - gerçek şu ki ortak değişken sadece öncelik alan başka bir değişkendir. Çalışmayı desenlerken, öncelikli olarak içiciliğin etkileri ile ilgilenmekteyiz. Bununla birlikte, görevin türü ve dikkat dağınıklığı olmak üzere, kontrol altına almamız gerektiğine inandığımız iki değişken bulunmaktadır. Bunlardan ilki (Görev), çalışmaya bağımsız değişken olarak dâhil edilmiştir. Kontrol altında alınması gerekli olan ikinci değişken (Dikkat dağınıklığı) de ortak değişken olarak araştırmaya eklenmiştir. Birçok bakımdan, bunlar aynı problemi ele almanın iki farklı yoludur. Bu iki değişkenin ele alınış biçiminde önemli farklılıklar bulunmasına rağmen; önemli benzerlikler de söz konusudur. $KT_{\text{grup}}$ değerinin elde edilmesinde, *hem* Görev *hem de* ortak değişken kontrol altında tutulmaktadır. Örneklem büyüklükleri eşit olduğunda Görev değişkeni Grup değişkenine dik olmakta ve kontrol altına alınması gereken bir şey bırakmamaktadır. Ancak bu sadece teknik bir özelliktir. Eşit olmayan örneklem büyüklüklerinde, her iki değişkenin kontrol altına alınması gerçek bir işlemdir. Pek çok ek varsayımı beraberinde getirmesine rağmen kovaryans analizinin ayrı bir istatistiksel teknik olduğunun gözden kaçırılmaması önemli olmakla birlikte, ortak değişkenin yalnızca başka bir değişken olduğuna dikkat edilmesi de eşit derecede önemlidir.

## 16.10 Çoklu Ortak Değişkenlerin Kullanılması

Şimdiye kadar tek bir ortak değişkenin kullanılmasıyla ilgilendik. Ancak araştırmanın tek bir ortak değişken ile kısıtlanmasına ilişkin mantıklı bir teorik veya pratik gerekçe yoktur. Örneğin, farklı öğretim yöntemlerinin etkililiği üzerine yapılan bir araştırmada, IQ, Yaş ve

Okul Türü (gelişen veya geleneksel) ortak değişkenler olarak alınabilir. Çoklu regresyon çerçevesinden bakıldığında bu özel bir problem oluşturmaz, buna karşılık geleneksel varyans analizi açısından bakıldığında birkaç ortak değişken kullanılması hesaplama karmaşıklığına boğar.

$R^2_{c,\alpha,\beta,\alpha\beta}$ ifadesinde, $\beta$ çok sayıda yordayıcı değişkeni temsil etmenin kısa bir yoludur (örneğin, $\beta_1, \beta_2, \ldots, \beta_b$). Aynı şekilde, $c$ bir dizi ortak değişkeni ($C_1, C_2, \ldots, C_k$) göstermek için kullanılır. Böylece daha özel bir gösterim açısından $R^2_{c,\alpha,\beta,\alpha\beta}$ aşağıdaki gibi bir temsile karşılık gelir:

$$R^2_{0.\text{IQ, Yaş, Okul, } A_1,\ B_1,\ B_2,\ AB_{11},\ AB_{12}}$$

Bu açıdan bakıldığında, çok sayıda ortak değişken kullanılmasının tek bir ortak değişken kullanılmasından bir farkı bulunmamaktadır. Eğer $C$, IQ, Yaş ve Okul Türü ortak değişkenlerini temsil ederse, $KT_{AB(\text{düzeltilmiş})}$ aşağıdaki gibi bir eşitlik ile elde edilir:

$$KT_{AB(\text{düzeltilmiş})} = KT_{\text{regresyon(IQ, Yaş, Okul, } A_1,\ B_1,\ B_2, AB_{11},\ AB_{12})} - KT_{\text{regresyon(IQ, Yaş Okul } A_1,\ B_1,\ B_2)}$$

Bir önceki örnekten, ortak değişken ile bağımlı değişken arasında doğrusal bir ilişki bulunması varsayımı dışında, ortak değişkenin doğasına ilişkin herhangi bir kısıtlama bulunmadığı hale gelmiş olmalıdır. Ortak değişken, Yaş ve IQ'de olduğu gibi sürekli veya okulun gelişimci ve geleneksel olması gibi iki kategorili süreksiz bir değişken olabilir.

**Bir uyarı sözü:** Sadece mümkün olmasından dolayı, çoklu ortak değişkenin kullanılması (ve gerçekte kolay), bu yöntemi benimsemek için iyi bir gerekçe değildir. Tek bir ortak değişken olduğunda kovaryans analizi sonuçlarını yorumlamak (imkânsız değilse) zordur. Ek ortak değişkenler analize eklendiğinde sonuçların yorumlanmasına ilişkin bu problem çok hızlı bir şekilde büyümektedir. Bundan dolayı eğer uygulamanın yapıldığı grup yaş, IQ, okul türü, ebeveyn mesleği vb. açıdan eşit ise, İngilizce öğretiminde kullanılan çeşitli yöntemlerin karşılaştırıldığı bir araştırmada, etkililiği test edilen yöntemler arasında karşılaştırma yapmanın kolay olduğu *söylenebilir*. Ancak, araştırmacının bu şekilde eşit grupların evrende gerçekten var olup olmadığını sorgulaması gerekir. Eşit grupların evrende bulunması söz konusu değilse, araştırmacı asla karşılaşılmayacak bir probleme cevap bulmaya çalışacak ve problemi cevaplamak için ortaya koyduğu çabanın karşılığını alamayacaktır. Ayrıca bu tür grupları oluşturmak mümkün olsa bile, bu grupların beklenen şekilde davranıp davranmayacağı cevaplanması gereken oldukça önemli bir sorudur. Öğrencilerin homojen sınıflarda yer alması, bağımlı değişken üzerinde tahmin edilemeyen çeşitli etkiler oluşturabilir.

## 16.11 Alternatif Deneysel Desenler

**tabakalara ayırma**

Ortak değişkenin önemli olduğu veri kümelerini ele almanın tek yolu kovaryans analizi değildir. **Tabakalara ayırma** (eşleştirilmiş örneklemler) ve fark puanları şeklinde iki alternatif yol daha vardır.

Ortak değişkene ilişkin ölçümler elimizde varsa ve katılımcıları gruplara ayırma konusunda özgürsek, ortak değişken açısından homojen olan alt gruplar oluşturabiliriz. Daha sonra bu homojen alt gruplardaki üyelerden her biri farklı bir gruba atanır. Böylelikle, varyans analizi sonucunda alt gruplardan dolayı oluşan etki, hata teriminden arındırılmış olur.

$Y$'nin $C$ üzerindeki regresyonu doğrusal olduğunda, eşleştirilmiş örneklemler ve kovaryans analizi neredeyse eşit derecede etkili olmaktadır. Eğer $p$ evrende $Y$ ile $C$ arasındaki korelasyon ve $\sigma^2_{\text{hata}}$ $Y$ üzerinden gerçekleştirilen varyans analizindeki hata terimi ise, eşleştirilmiş örneklemlerin kullanılması hata varyansında aşağıdaki gibi azalma sağlamaktadır.

$$\sigma^2_{\text{hata}}(1 - \rho^2)$$

Bu durumda, kovaryans analizinden dolayı oluşan azalma aşağıdaki gibi olur:

$$\sigma^2_{\text{hata}}(1-\rho^2)\frac{(f_{\text{hata}})}{(f_{\text{hata}}-1)}$$

Burada $f_{\text{hata}}$, hata varyansına ait serbestlik derecesidir. Açık bir şekilde görüldüğü gibi, herhangi bir makul $f_{\text{hata}}$ değeri için regresyonun doğrusallığı varsayımı altında iki yöntem hemen hemen aynı derecede etkilidir. Eğer *Y* ile *C* arasındaki ilişki doğrusal değilse, eşleştirilmiş örneklemler, kovaryans analizine göre daha etkili bir yol olur.

**fark puanları**

Kovaryans analizine ikinci bir alternatif de **fark puanlarının** kullanılmasıdır. Ortak değişken (*C*) denel işlem uygulanmadan önceki ölçümleri ve *Y* denel işlemin uygulanmasının ardından elde edilen son ölçümleri temsil etmek üzere, ön ölçümlerden kaynaklanan farklılıkları kontrol etmek amacıyla bazen $C - Y$ değeri varyans analizinde bağımlı değişken olarak kullanılır. Ancak bu yaklaşım sadece *C* ve *Y*'ye ilişkin ölçümlerin karşılaştırılabilir olduğu durumlarda uygundur. Örneğin sürüş testine ilişkin puanların (*Y*) IQ puanlarından (*C*) çıkarılması pek de doğru bir yaklaşım değildir. Nadiren karşılaşılan bir durum olsa da, *Y* ile *C* arasındaki ilişkinin doğrusal olduğu ve $b_{\text{CY}} = 1{,}00$ değerine karşılık geldiği durumlarda, grup etkisine ilişkin *fark* puanları üzerinden gerçekleştirilen varyans analizi ile kovaryans analizinden elde edilen sonuçlar birbirine eşit olmaktadır. Gerçekten, Conti ve Musty'nin (1984) çalışmasında, THC enjeksiyonu öncesi ve sonrası ölçümler arasındaki farkın bağımlı değişken olarak işlem gördüğü bir varyans analizi uygulandığında, sonuçlar kesinlikle değişmektedir [$F_{(4,42)} = 0{,}197$]. Bu durumda, kovaryans analizi açıkça daha güçlü bir yöntemdir. Bu bölümün sonunda yer alan Alıştırma 16.24, kovaryans analizinin bu yönünü ortaya koymaktadır. Harris (1963) ile Huitema (1980, 2005), bu problemin daha kapsamlı bir şekilde ele alındığı kaynaklardır.

Burada akılda tutulması gereken; ön test ve son test puanları arasındaki ilişkiye ait eğimin 1,0 olması, bireylerin hangi noktada başladıklarından bağımsız olarak uygulanan deneysel işlemin, puanlarda benzer bir artışa yol açmakta olmasıdır. Ancak değişim katılımcıların başladıkları nokta ile orantılı olabilir. Örneğin matematikte çok başarısız olan bir kişi, deneysel işlem sonrasında matematikte daha başarılı olan birine kıyasla daha fazla bir gelişme gösterebilir. Bu durumda, bireyin matematik başarısındaki artış puanları ön test puanları ile negatif ilişkili olacaktır. Conti ve Musty'nin çalışmasında, daha hareketli hayvanların değişmesi olasılığı daha az hareketli hayvanların değişmesi olasılığına göre daha yüksektir. Conti and Musty'nin, iki puan arasındaki farkı değil, son test puanlarının ön test puanlarına göre yüzdesini bağımlı değişken olarak almalarına kaynaklık eden faktör de budur.

## Anahtar Terimler

Genel Doğrusal Model (16.1)
Tasarım Matrisi (16.1)
Yöntem III (16.4)
Tip III KT (16.4)
Metot II (16.4)
Tip II KT (16.4)
Metod I (16.4)
Tip I KT (16.4)
Hiyerarşik kareler toplamları (16.4)
Ardışık Kareler toplamları (16.4)
Eşit ağırlıklandırılmış ortalamalar (16.4)
Kovaryans analizi (16.5)
Ortak değişken (16.5)
Düzeltilmiş *Y* ortalamaları (16.5)
Regresyonun homojenliği (16.5)
$KO'_{\text{hata}}$ (16.5)
Gerçek-puan kovaryans analizi (16.6)
Eşit olmayan grup deseni (16.6)
Tabakalara ayırma (16.11)
Fark Puanları (16.11)

## Alıştırmalar

16.1 Aşağıdaki hipotetik veri kümesi metnin her bir satırında birim zamandaki göz odaklanmalarının sayısına göre zayıf, orta ve iyi okuyuculardan elde edilmiştir.

| Zayıf | Orta | İyi |
|---|---|---|
| 10 | 5 | 3 |
| 7 | 8 | 5 |
| 8 | 4 | 2 |
| 11 | 6 | 3 |
| 5 | 5 | 4 |

© Cengage Learning 2013

a. Bu veri kümesi için tasarım matrisini oluşturunuz.

b. En-küçük kareler varyans analizini hesaplamak için herhangi bir standart regresyon programını kullanınız.

c. Geleneksel yolla varyans analizini gerçekleştirerek cevaplarınızı karşılaştırınız.

16.2 Alıştırma 16.1'deki veri kümesi için

a. İşlem etkisini hesaplayınız ve regresyon modelinin bu işlem etkisini (yeniden ürettiğimi ortaya koyduğunu) gösteriniz.

b. Regresyon analizindeki $R^2$ değerinin varyans analizindeki $\eta^2$ değerine eşit olduğunu gösteriniz.

16.3 Alıştırma 16.1'deki veri kümesi için, Orta gruba 5 ile 8 ve İyi gruba 2, 3, 3 ile 5 puanlarını ekleyiniz. Alıştırma 16.1'e göre daha fazla gözlemden oluşan bu veri kümesi için varyans analizini tekrar gerçekleştiriniz.

16.4 Alıştırma 16.3'te elde edilen veri kümesi için Alıştırma 16.2'deki analizleri tekrar yapınız.

16.5 Bir psikolog Cinsiyet, Sosyo Ekonomik Düzey (SED) ve algılanan Kontrol Odağı arasındaki ilişkilerin belirlenmesiyle ilgilenmektedir. Araştırma kapsamında her bir cinsiyet ve SED kombinasyonu için 25 ile 30 yaş arasındaki 8 yetişkin örnekleme alınmıştır. Araştırmanın katılımcılarını oluşturan bireylere kontrol odağı ölçeği uygulanmıştır. (Kontrol odağı ölçeğindeki yüksek puanlar bireyin günlük yaşamda kendini kontrol altında hissettiğine işaret etmektedir.)

| | Sosyo Ekonomik Düzey | | |
|---|---|---|---|
| | Düşük | Orta | Yüksek |
| Erkek | 10 | 16 | 18 |
| | 12 | 12 | 14 |
| | 8 | 19 | 17 |
| | 14 | 17 | 13 |
| | 10 | 15 | 19 |
| | 16 | 11 | 15 |
| | 15 | 14 | 22 |
| | 13 | 10 | 20 |
| Kadın | 8 | 14 | 12 |
| | 10 | 10 | 18 |
| | 7 | 13 | 14 |
| | 9 | 9 | 21 |
| | 12 | 17 | 19 |
| | 5 | 15 | 17 |
| | 8 | 12 | 13 |
| | 7 | 8 | 16 |

© Cengage Learning 2013

a. Bu veri kümesine geleneksel varyans analizini uygulayınız.

b. Uygun tasarım matrisi kullanarak bu veri üzerinden aşağıdaki kareler toplamı değerleri hesaplanmıştır ($a$ = Cinsiyet ve $\beta$ = SED).

$KT_{Y} = 777{,}6667$ $\quad KT_{reg(a,\ \beta,\ a\beta)} = 422{,}6667$

$KT_{reg(a,\ \beta)} = 404{,}0000$ $\quad KT_{reg(\beta,\ a\beta)} = 357{,}333$

$KT_{reg(a,\ a\beta)} = 84{,}000$

Verilen kareler toplamı değerlerinden yararlanarak varyans analizi özet tablosunu yapınız.

16.6 Yordayıcı olarak tasarım matrisinin SED kısmı kullanıldığında, $KT_{reg(\beta)} = 338{,}6667$ olarak bulunmuştur.

a. Bu değerin Alıştırma 16.5'te hesaplanan $KT_{SED}$ ile aynı olmasının sebebi nedir?

b. Böyle bir durum tüm varyans analizlerinde gözlenir mi?

16.7 Alıştırma 16.5'teki veri kümesi için erkek ile düşük-SED kesişimindeki son iki gözlem, erkek ile orta-SED kesişimindeki son üç gözlem ve kadın ile yüksek-SED kesişimindeki son üç gözlem veri kümesinden çıkarıldığında aşağıdaki kareler toplamı değerleri elde edilmektedir.

$$KT_{Y} = 750{,}1951 \qquad KT_{reg(a,\ \beta,\ a\beta)} = 458{,}7285$$
$$KT_{reg(a,\ \beta)} = 437{,}6338 \qquad KT_{reg(\beta,\ a\beta)} = 398{,}7135$$
$$KT_{reg(a,\ a\beta)} = 112{,}3392 \qquad KT_{reg(a)} = 95{,}4511$$
$$KT_{reg(\beta)} = 379{,}3325$$
$$KT_{reg(a\beta)} = 15{,}8132$$

Bu kareler toplamı değerlerinden yararlanarak varyans analizini hesaplayınız.

16.8 Alıştırma 16.7'deki veri kümesi için yalnızca SED yordayıcısı kullanıldığında, $KT_{reg(\beta)} = 379{,}3325$ bulunmuştur. Bu değerin alıştırma 16.7'deki $KT_{SED}$ değerine eşit olmamasının sebebi nedir?

16.9 Alıştırma 16.5'teki veri kümesi için tam model aşağıdaki gibidir.

$$1{,}1167A_1 - 3{,}1667B_1 - 0{,}1667B_2 + 0{,}8333AB_{11} - 0{,}1667AB_{12} + 13{,}4167$$

a. Bu modelin Tablo 16.2'de sunulan yöntem ile hesaplandığı gibi, grup ve etkileşim etkisini ürettiğini gösteriniz.

16.10 Alıştırma 16.7'deki veri kümesi için tam model aşağıdaki gibidir.

$$1{,}2306A_1 - 3{,}7167B_1 - 0{,}3500B_2 + 0{,}4778AB_{11} - 0{,}5444AB_{12} + 13{,}6750$$

a. Bu modelin Tablo 16.3'de hesaplandığı gibi, grup ve etkileşim etkisini ürettiğini gösteriniz.

16.11 Aşağıdaki veri kümesini kullanarak, bu bölümde savunulan Yöntem III'ün ağırlıklandırılmamış ortalamalar ile ilgilendiğini gösteriniz.

| | $B_1$ | $B_2$ |
|---|---|---|
| $A_1$ | 5 | 11 |
| | 3 | 9 |
| | | 14 |
| | | 6 |
| | | 11 |
| | | 9 |
| $A_2$ | 10 | 6 |
| | 11 | 2 |
| | 12 | |
| | 7 | |

© Cengage Learning 2013

16.12 Alıştırma 16.5'teki kareler toplamını temsil eden bir Venn diyagramı çiziniz.

16.13 Alıştırma 16.7'deki kareler toplamını temsil eden bir Venn diyagramı çiziniz.

16.14 Eğer SAS programına erişebiliyorsanız, SAS programından yararlanarak Alıştırma 16.7'deki veri kümesini analiz ediniz. Model komutunun sonuna Add/SS1 SS2 SS3 SS4 ekleyerek aşağıdaki açıklamaların doğruluğunu gösteriniz.

a. Tip I kareler toplamı modeldeki her bir terime model açıklamasında yalnızca kendisinden daha önce gelen terimler için düzeltme uygulamaktadır.

b. Tip II kareler toplamı, etkileşim terimine temel etkilerin her biri için düzeltme uygularken temel etkilere yalnızca diğer temel etkiler için düzeltme uygulamaktadır.

c. Tip III kareler toplamı, modeldeki diğer tüm terimler için her bir terime düzeltme uygulamaktadır.

d. Bu durumda, Tip IV kareler toplamı Tip II kareler toplamına eşittir.

16.15 Ailelerin enerji tüketimi ile ilgili bir çalışmada, aileler üç gruba ayrılmıştır. Grup 1 günlük elektrik oranı sistemine kayıtlı kişilerden oluşmaktadır (elektriğin kilovat/saat ücreti gün içerisinde elektrik tüketiminin en yoğun olduğu saatlerde daha fazladır). Grup 2 bu sisteme kayıtlı olan ancak kullanmayan kişilerden oluşmaktadır. Grup 3 ise bu sisteme ilgi göstermeyen kişilerden oluşmaktadır. Her bir ev için her ay kaydedilen elektrik faturası araştırmada bağımlı değişken ($Y$) olarak alınmıştır. Ortak değişken ($C$) olarak bu evlerin geçen yılın aynı ayındaki elektrik faturaları alınmıştır. Elde edilen veri aşağıdaki gibidir.

| Grup 1 | | Grup 2 | | Grup 3 | |
|---|---|---|---|---|---|
| Y | C | Y | C | Y | C |
| 58 | 75 | 60 | 70 | 75 | 80 |
| 25 | 40 | 30 | 25 | 60 | 55 |
| 50 | 68 | 55 | 65 | 70 | 73 |
| 40 | 62 | 50 | 50 | 65 | 61 |
| 55 | 67 | 45 | 55 | 55 | 65 |

© Cengage Learning 2013

a. Tasarım matrisini oluşturunuz.
b. Kovaryans analizini uygulayınız.

16.16 Alıştırma 16.15'te açıklanan deneyi geliştirmek için bir psikolog her bir gruba bir dizi ev eklemiştir. Bu evlere elektrik faturalarının ne kadar hızlı arttığını göstermek için özel bir sayaç monte edilmiştir. (Günlük miktar sayaçta görüntülenmektedir.) Veri aşağıdaki gibidir. Aşağıdaki tablo ve Alıştırma 16.15 incelendiğinde görüleceği gibi, sayaçsız evlere ait veri Alıştırma 16.15'teki veri ile aynıdır.

| | *Y* | *C* | *Y* | *C* | *Y* | *C* |
|---|---|---|---|---|---|---|
| Sayaçsız | 58 | 75 | 60 | 70 | 75 | 80 |
| | 25 | 40 | 30 | 25 | 60 | 55 |
| | 50 | 68 | 55 | 65 | 70 | 73 |
| | 40 | 62 | 50 | 50 | 65 | 61 |
| | 55 | 67 | 45 | 55 | 55 | 65 |
| Sayaçlı | 25 | 42 | 40 | 55 | 55 | 56 |
| | 38 | 64 | 47 | 52 | 62 | 74 |
| | 46 | 70 | 56 | 68 | 57 | 60 |
| | 50 | 67 | 28 | 30 | 50 | 68 |
| | 55 | 75 | 55 | 72 | 70 | 76 |

© Cengage Learning 2013

a. Regresyonların homojenliği varsayımını test ettikten sonra bu veri için kovaryans analizi uygulayınız.
b. Sonuçlar hakkında uygun çıkarımlarda bulununuz.

16.17 Alıştırma 16.16'daki veri için düzeltilmiş ortalamaları hesaplayınız.

16.18 Bu yıla ait elektrik faturasını geçen yılki elektrik faturasından çıkararak, Alıştırma 16.16'daki veri için her bir eve ilişkin enerji tasarrufunu hesaplayınız. Daha sonra, tasarruf puanları üzerinden varyans analizi uygulayarak elde ettiğiniz sonuçları kovaryans analizi sonuçları ile karşılaştırınız.

16.19 Klemchuk, Bond ve Howell (1990) tarafından çocuklarda rol alma davranışının incelendiği bir araştırma yürütülmüştür. Araştırma kapsamında, çocuklara bir dizi rol alma görevi uygulanmıştır. Çocuklar kreşe gidip gitmeme durumuna göre gruplandırılmış ve yaş açısından da 2-3 ve 4-5 olacak şekilde iki gruba ayrılmıştır. Araştırmada, kreşe giden çocukların rol alma performanslarının daha iyi olacağı şeklinde bir hipotez kurulmuştur. Bu alıştırmaya ilişkin veri bu kitabın internet sitesinde Ex16-9.dat adı ile kayıtlıdır. Uygun analizleri gerçekleştiriniz.

## Bilgisayar Alıştırmaları

16.20 Introini-Collison ve McGaugh (1986) tarafından gerçekleştirilen ve Alıştırma 11.29 ve 11.31 öncesinde açıklanan çalışmayı incelemek için internet sitesinde Epinuneq. dat adı ile kayıtlı olan veri dosyasını kullanınız. Bir istatistik paket programından yararlanarak, eşit olmayan örneklem büyüklükleri için iki yönlü varyans analizi uygulayınız. Bu analizden ne tür bir sonuca varılabilir?

16.21 Mireault ve Bond'un (1992) çalışmasına ait olan ve Alıştırma 7.6'da belirtilen Mireault.dat adlı veri dosyasını kullanarak, Genel Semptom İndeksi *T* (Global Symptom Index T-GSIT) puanları üzerinden Cinsiyet ve Grup değişkenlerinin bağımsız değişken olarak alındığı iki yönlü varyans analizini uygulayınız.

16.22 Alıştırma 16.21'deki verinin aynısını kullanarak, üniversitede geçirilen yıl'ın (ÜnivYıl) ortak değişken olarak alındığı bir kovaryans analizi uygulayınız.

a. ÜnivYıl neden ortak değişken olarak alınmıştır?

b. Elde edilen sonuçlar nasıl yorumlanmalıdır?

16.23 Alıştırma 16.22'de, ÜnivYıl ortak değişken olarak kullanılmıştır. Cinsiyet ve grup değişkenlerini bağımsız değişken olarak kullanarak ÜnivYıl üzerinden varyans analizi uygulayınız. Bu varyans analizinden elde edilen sonuçlar, önceki kovaryans analizi ile ilgili ne tür bilgiler verir?

16.24 Everitt, üç farklı deneysel işlemin genç kızlarda anoreksi (iştahsızlık) üzerindeki etkisinin incelendiği bir araştırmanın veri kümesini rapor etmiştir. Birinci grupta bilişsel terapiye yer verilmiştir. Kontrol grubu olan ikinci grupta herhangi bir işlem uygulanmamıştır. Üçüncü grupta ise, aile terapisi uygulanmıştır. Buradaki veri kümesi Bölüm 14'te incelenen veri ile aynıdır. Veri kümesi aşağıda sunulmuştur. Bu veriye kitabın internet sitesinden erişilebilmektedir.

a. Kazanç puanları açısından grup farklılıklarına ilişkin varyans analizi uygulayınız. (Bu işlem, Bölüm 14'te zaten yapılmış olabilir.)

b. Ön test puanlarını göz ardı ederek son test puanlarına ilişkin varyans analizi uygulayınız.

c. Son olarak, ön test puanlarını ortak değişken olarak alıp son test puanları için kovaryans analizi uygulayınız.

d. Elde edilen üç cevap arasında nasıl bir ilişki bulunmaktadır? Bu cevaplar, kazanç (değişim) puanları işlemi ile kovaryans analizi arasındaki benzerlik ve farklılıklar hakkında ne gibi bilgiler vermektedir?

e. Kovaryans analizi için Grup değişkeni üzerinde $\eta^2$ değerini hesaplayınız.

f. Düzeltilmiş ortalamaları kullanarak, iki terapi grubunun (kontrol grubunu dikkate almadan) karşılaştırılmasına ilişkin *d* değerini hesaplayınız.

| Grup | Ön test | Son Test | Kazanç | Grup | Ön test | Son Test | Kazanç |
|---|---|---|---|---|---|---|---|
| 1 | 80,5 | 82,2 | 1,7 | 2 | 75,1 | 86,7 | 11,6 |
| 1 | 84,9 | 85,6 | ,7 | 2 | 80,6 | 73,5 | -7,1 |
| 1 | 81,5 | 81,4 | -,1 | 2 | 78,4 | 84,6 | 6,2 |
| 1 | 82,6 | 81,9 | -,7 | 2 | 77,6 | 77,4 | -0,2 |
| 1 | 79,9 | 76,4 | -3,5 | 2 | 88,7 | 79,5 | -9,2 |
| 1 | 88,7 | 103,6 | 14,9 | 2 | 81,3 | 89,6 | 8,3 |
| 1 | 94,9 | 98,4 | 3,5 | 2 | 78,1 | 81,4 | 3,3 |
| 1 | 76,3 | 93,4 | 17,1 | 2 | 70,5 | 81,8 | 11,3 |
| 1 | 81,0 | 73,4 | -7,6 | 2 | 77,3 | 77,3 | 0,0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 80,5 | 82,1 | 1,6 | 2 | 85,2 | 84,2 | -1,0 |
| 1 | 85,0 | 96,7 | 11,7 | 2 | 86,0 | 75,4 | -10,6 |
| 1 | 89,2 | 95,3 | 6,1 | 2 | 84,1 | 79,5 | -4,6 |
| 1 | 81,3 | 82,4 | 1,1 | 2 | 79,7 | 73,0 | -6,7 |
| 1 | 76,5 | 72,5 | -4,0 | 2 | 85,5 | 88,3 | 2,8 |
| 1 | 70,0 | 90,9 | 20,9 | 2 | 84,4 | 84,7 | 0,3 |
| 1 | 80,4 | 71,3 | -9,1 | 2 | 79,6 | 81,4 | 1,8 |
| 1 | 83,3 | 85,4 | 2,1 | 2 | 77,5 | 81,2 | 3,7 |
| 1 | 83,0 | 81,6 | -1,4 | 2 | 72,3 | 88,2 | 15,9 |
| 1 | 87,7 | 89,1 | 1,4 | 2 | 89,0 | 78,8 | -10,2 |
| 1 | 84,2 | 83,9 | -0,3 | 3 | 83,8 | 95,2 | 11,4 |
| 1 | 86,4 | 82,7 | -3,7 | 3 | 83,3 | 94,3 | 11,0 |
| 1 | 76,5 | 75,7 | -0,8 | 3 | 86,0 | 91,5 | 5,5 |
| 1 | 80,2 | 82,6 | 2,4 | 3 | 82,5 | 91,9 | 9,4 |
| 1 | 87,8 | 100,4 | 12,6 | 3 | 86,7 | 100,3 | 13,6 |
| 1 | 83,3 | 85,2 | 1,9 | 3 | 79,6 | 76,7 | -2,9 |
| 1 | 79,7 | 83,6 | 3,9 | 3 | 76,9 | 76,8 | -0,1 |
| 1 | 84,5 | 84,6 | ,1 | 3 | 94,2 | 101,6 | 7,4 |
| 1 | 80,8 | 96,2 | 15,4 | 3 | 73,4 | 94,9 | 21,5 |
| 1 | 87,4 | 86,7 | -,7 | 3 | 80,5 | 75,2 | -5,3 |
| 2 | 80,7 | 80,2 | -0,5 | 3 | 81,6 | 77,8 | -3,8 |
| 2 | 89,4 | 80,1 | -9,3 | 3 | 82,1 | 95,5 | 13,4 |
| 2 | 91,8 | 86,4 | -5,4 | 3 | 77,6 | 90,7 | 13,1 |
| 2 | 74,0 | 86,3 | 12,3 | 3 | 83,5 | 92,5 | 9,0 |
| 2 | 78,1 | 76,1 | -2,0 | 3 | 89,9 | 93,8 | 3,9 |
| 2 | 88,3 | 78,1 | -10,2 | 3 | 86,0 | 91,7 | 5,7 |
| 2 | 87,3 | 75,1 | -12,2 | 3 | 87,3 | 98,0 | 10,7 |

16.25 Etki büyüklüğünü de içerecek şekilde, Everitt'in deneyine ilişkin sonuçları yazınız.

## Tartışma Soruları

16.26 İlk olarak Bölüm 14'ün 14.7. Kısmı'ndaki örnekten bir varyans analizi örneği oluşturulması düşünülmüştür. Cinsiyet ve Grup bağımsız değişken, son test puanları bağımlı değişken ve ön test puanları ortak değişken olarak alınmıştır. (İzleyen 6. ve 12. ayda elde edilen ölçümler dikkate alınmamıştır.) Böyle bir araştırma kovaryans analizi için kötü bir örnek olarak düşünülebilir. Bu durum neden kaynaklanıyor olabilir? Sözü geçen çalışmayı kovaryans analizi için uygun hale getirecek bir yol var mıdır?

16.27 Ön test ve son test puanlarına sahip olduğumuz herhangi bir çalışmada, fark puanlarına bakılabileceği (gruplar arasında karşılaştırılarak) ya da ön test puanlarının ortak değişken olarak alınabileceği ifade edilmişti. Bu iki analiz ancak son test ile ön test puanlarına ilişkin eğimin 1,00'a eşit olması halinde eşit sonuçlar üretmektedir. Bu koşulun karşılanması (ya da en azından eğimin böyle bir değere yaklaşması) olasılığı hakkında ne söylenebilir? $b$ = 1,00 olması gerçekte ne anlama gelmektedir?

16.28 Eğimin 1,0 değerine yakın olmadığı Alıştırma 16.25 ile ilgili ilgili bir örnek bulunuz ya da oluşturunuz. Bu örnek için hem kovaryans analizini hem de fark puanları üzerinde gerçekleştirilen bir $t$-testi uygulayınız. Bu analizlerin her ikisi de mantıklı mıdır?



Bölüm 17

# Meta Analizi ve Tek-Durum Desenleri

## Hedefler

Meta-analizi, birtakım farklı çalışmalardan mantıklı bir çıkarımda bulunma yöntemi olarak sunmak ve meta-analize ilişkin tekniklerin tek-durum desenlerinde nasıl kullanılacağını göstermek.

## Konular


### Meta-Analiz

**17.1** Etki Büyüklüğü Ölçülerinin Kısa Özeti

**17.2** Bir Örnek - Çocuk ve Ergenlik Depresyonu

**17.3** İkinci Bir Örnek - Nikotin Sakızı ve Sigarayı Bırakma

### Tek Durum Desenleri

**17.4** Standartlaştırılmış Ortalama Farkların Araştırıldığı Analizler

**17.5** Depresyon İçin Bir Durum Çalışması

**17.6** Tek-Durum Desenine İkinci Bir Yaklaşım - Parça Parça Regresyon Kullanımı

ÖNCEKİ 16 BÖLÜMDE, grup ortalamalarının karşılaştırıldığı, korelasyon ve regresyon modellerine bakıldığı, olasılık tablolarının analiz edildiği vb. bireysel çalışmalara odaklanmıştık. Bunlar, çoğunlukla davranış bilimlerindeki araştırmacıların yaptığı çalışmalardır. Bununla birlikte, aynı araştırma sorusuna odaklanmış birtakım benzer çalışmalara ulaştıktan sonra, bu çalışmaları nasıl bir araya getirebiliriz sorusunu sormalıyız ve genel sonuçlara varmalıyız. Bir araştırmacı, bir deney faresine şok uygulandıktan sonra kaçınma davranışını çalışmış ve şokun şiddetiyle kaçınmada artış olduğunu bulmuş olabilir. Bir başka araştırmacı da benzer bir çalışma yapmış ve örneğin *p* değeri 0,17 ile manidar olmayan bir etki bulmuş olabilir. Bu durumda, biz nasıl bir sonuca varacağız? Şok düzeyi gerçekten kaçınmayı etkilemekte midir, kaçınma davranışı sadece belirli durumlarda mı şoktan etkilenmektedir veya ikinci çalışma birinci çalışma ile tamamen uyumlu mudur? Bu bölümde, aynı anda ikiden fazla çalışmayı ve bunların yanında yukarıdaki türden soruları da dikkate alacağız.

Bu bölümün ikinci yarısında, yukarıdakiyle ilgili olarak tek durum veya tek denek üzerindeki davranışsal müdahaleler uygulayan bir araştırmacının ilgilendiği sorular ele alınacaktır. Bu durumları en iyi şekilde nasıl pekiştirebiliriz? İki farklı konunun bir araya getirilmesi tuhaf görülebilir, ancak gerçekte, bir çalışma diğerine yol gösterir ve bu konuların birlikte ele alınmasına ihtiyaç olabilir.

# META ANALİZ

Borenstein, Higgins ve Ronthstein'ın (2009) meta-analizi ile ilgili mükemmel kitabı, Dr. Benjamin Spock ve diğer pediyatristlerin ailelere bebeklerinin karınları üzerinde yatmalarını tavsiyelerini işaret ettikleri gerçeği ile başlar. (Eğer Dr. Spock'ı tanımıyorsanız, hayır, O *Yıldız Savaşları*'nda oynamıyordu, anneniz kesin tanır. 20'nci yüzyılın ikinci yarısı boyunca bu ülkede çocuk yetiştirme konusunda başkalarından çok daha etkili olduğu muhtemeldir.) Bu süre boyunca 100.00 bebek Ani Bebek Ölümü Sendromu'ndan (Sudden Infant Death Syndrome) hayatını kaybemişti - çoğu karınları üzerinde uyurken. Bu süreçte, bu uygulamanın tehlikeli olduğuna ilişkin kanıtlar toplanabilirdi ancak çok geç olana kadar hiç kimse bu çalışmaları bir araya getirmedi. Nihayet bu yapıldığında, ailelere bebeklerini sırt üstü yatırmaları tavsiye edildiğinde ABÖS ölümleri ciddi bir şekilde düşüşe geçti. Bu örneği, bir alanda var olan tüm literatürün karılaştırılmasına ve mantıklı sonuçların birleştirilmesine ihtiyaç olduğuna ilişkin basit ancak dramatik bir örnek olarak kullandım.

**meta-analiz**

**Meta-analiz**, bu amaçla tasarlanır. Meta-analiz, Kanıt Temelli İlaç'ın (Evidence Based Medicine) ana özelliğidir ve Cochrane Coalition, medikalin hemen hemen tüm alanlarındaki analizlere ilişkin binlerce çalışma yayınlar, bunların çoğu da doğrudan davranış bilimleri ile ilgilidir. Bu bölümde bu çalışmalardaki veri kümelerini kullanacağız.

Kanıtı pekiştirecek ilk girişim öyküsel çalışmalara dayanır. Burada araştırmacı, ilgilendiği konu ile ilgili literatürden çok sayıda çalışma okuyacak, okuduğu çalışmalarla ilgili öznel (subjektif) yargılara varacak ve nihayet bir sonuca ulaşacaktır. Bu şekilde elde edilen sonuçlar da oldukça özneldir, araştırmacının seçtiği şekilde farklı çalışmalar farklı ağırlıklandırılır. Ayrıca, yaklaşım, öncelikli olarak literatürde yayımlanmış çalışmalar olur, ve doğal olarak, istatistiksel manidar sonuçlara odaklanılır. Rosenthal'ın "**çekmeceye kaldırılmış dosya (file drawer problem) problemi**" adını verdiği yayımlanmadan dosyalara kaldırılmış olumsuz sonuçlar göz ardı edilir.

**çekmeceye kaldırılmış dosya (file drawer problem) problemi**

En basit şekilde ifade edildiğinde meta-analiz, tek bir konudaki birden çok çalışmanın sonuçlarının karmaşık bir ortalamasıdır. Sonuçların kesinliğine göre çalışmaların ağırlıklandırılması ve etki büyüklükleri ile başlanır. Burada, ergenlik depresyonu yaşayan çocuklardan oluşan deney ve kontrol gruplarının karşılaştırıldığı çalışmaları örnek alarak baş-

layacağım. Fakat, akılda tutulması gereken önemli bir nokta, etki büyüklüğünü bir kez *d*, risk oranı veya korelasyon ölçüsü olarak aldığımızda etki büyüklüklerinin test edilmesi ve birleştirilmesi için eşitliklerin tümünün aynı olacağıdır. Böylece, eşitlikte *d* gördüğümüzde *r*'nin Fisher dönüşümüi log(RO) veya diğer etki büyüklüğü ölçüsü yerini alabilir ve tüm hesaplmalar bu değerler ile sürdürülür.

# 17.1 Etki Büyüklüğü Ölçülerinin Kısa Özeti

Devam etmeden önce, asıl konudan biraz uzaklaşalım ve kısaca etki büyüklüğü ölçülerini yeniden inceleyelim. Bunları farklı bölümlerde gördük ve birleştirmek mantıklı olur. Bir sonraki kesim kapsamlı değil ama en önemli ölçüleri içerir.

## Ortalamalar Üzerine Ölçüler

Ortalamalar üzerine birkaç ölçüyle ilgileneceğim. Bunların hepsi, Hedge *g* olarak bilinen Hedgs'in düzeltmesi uygulanmakla birlikte Cohen'ın *d* formundadır.

- **İki Bağımsız Örneklem – Deney ve Kontrol**
  - Varyansları Birleştirme

$$d = \frac{\overline{X}_1 \quad - \overline{X}_2}{s_{\text{birleşik}}}$$

$$s_d = \sqrt{\frac{n_1 + n_2}{n_1 x n_2} + \frac{d^2}{2(n_1 + n_2)}}$$

  - Kontrol grubunun standart sapmasının standartlaştırılması

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s_K}$$

$$s_d = \sqrt{\frac{n_1 + n_2}{n_1 x n_2} + \frac{d^2}{2n_1}}$$

- **Eşleştirilmiş Ölçümler**
  - Ortalama standart sapmanın standartlaştırılması

$$d = \frac{\overline{X}_{\text{son}} - \overline{X}_{\text{ön}}}{s_{\text{birleşik}}} = \frac{\overline{X}_{\text{son}} - \overline{X}_{\text{ön}}}{s_{\text{fark}} \Big/ \sqrt{2(1-r)}}$$

$$s_d = \sqrt{\left(\frac{1}{n_1} + \frac{d^2}{2n}\right) \times 2(1-r)}$$

burada *r*, ön ve son-işlemler arasındaki korelasyondur.

$s_{\text{fark}}$ üzerine standartlaştırma nadiren önerilir.

$$d = \frac{\overline{X}_{\text{son}} - \overline{X}_{\text{ön}}}{s_{\text{fark}}}$$

- **Çoklu Bağımsız Gruplar**
  Deney (işlem) etkilerini iki veya daha fazla grupta hesaplamak genelde mantıklıdır; böylece çoklu grupların etkileri için formülleri atlayacağım.

- **Hedges'in *g'si***
  Bu noktada standartlaştırılmış ortalama fark olarak *d*'yi kullanacağız ve bu bölümde bu şekilde devam edeceğiz. Ortak standart sapma yerine kontrol grubunun standart sapmasına bölünmesi Glass tarafından önerilmiş olmasına rağmen buradaki *d* aslında Cohen'ın *d*'sidir ve bu şekilde kullanmaya devam edeceğiz. *d* ile ilgili problem, örneklem büyüklüğü küçük olduğunda, bir miktar yanlıdır ve $\delta$ değeri olduğundan büyük çıkma eğilimindedir.

  Hedges (1981), düzeltme faktörü için çözüm içeren küçük bir düzenleme önermiştir:

  $$J = 1 = \frac{3}{4sd - 1}$$

  ve *d* ve $s_d$ *J* ile çarpılarak.

  $$g = d \times J$$

  $$s_g = s_d \times J$$

  bulunur. Bu, bize Hedges'in *g*'sini verir ve bundan sonra *d* ile nasıl çalışıyorsak aynı şekilde *g* ile çalışmaya devam edebiliriz. Bu örnek için *d*'yi bırakmayacağız, *d* ile çalışacağız, ancak bu nokta önemlidir. Etki büyüklüklerini bir kez elde ettiğimizde, bunların *d*, *g* veya log(risk oranı) gibi diğer etki büyüklüğü olması önemli değildir. Matematiksel işlemler aynıdır.

## Olasılık Tablolarına Dayalı Etki Büyüklükleri

Aşağıdaki özet tablonun olduğunu varsayalım:

| Durum | Olay | Olay-değil | N |
|---|---|---|---|
| Deney | A | B | A+B |
| Kontrol | C | D | C+D |

© Cengage Learning 2013

- Bağıl Risk = Risk Oranı

  $$RO = \frac{A}{A + B} \Big/ \frac{C}{C + D} = \frac{A(C + D)}{B(A + B)}$$

  Bununla birlikte, risk oranını genellikle logaritmik ölçekte ele alırız. Böylece;

  $$\text{LogRiskOranı} = \ln(RO)$$

  $$Var_{\text{LogRiskOranı}} = \frac{1}{A} - \frac{1}{A+B} + \frac{1}{C} - \frac{1}{C+D}$$

  $$S_{\text{LogRiskOranı}} = \sqrt{Var_{\text{LogRiskOranı}}}$$

  $$GA_{\ln(\text{RO})} = \text{LogRiskOranı} \pm 1{,}96 \times \text{S}_{\text{LogOddsOranı}}$$

Etki büyüklüklerini ve bu istatistikleri kullanarak güven aralıklarını hesaplamayı gerçekleştiririz, ancak daha sonra ortalama ve güven aralıklarını aşağıdaki şekilde orjinal ölçülerine dönüştürürüz:

$$\text{Ortalama} = e^{\ln(RO)}$$

$$\text{Alt sınır} = e^{\text{alt sınırLog}_{\text{RiskOranı}}}$$

$$\text{Üst sınır} = e^{\text{üst sınırLog}_{\text{RiskOranı}}}$$

Burada e, doğal sayı sisteminde 2,71828 değerini ifade eder.

- Olasılık Oranı (OR)

  $$OR = \frac{A/B}{C/D} = \frac{A \times D}{B \times C}$$

Tekrar belirtelim ki normal olarak log odds oranı (Ln(OO)) ile çalışırız, bu durumda, LogOddsOranı = ln(*OO*)

$$\text{Var}_{\text{Logoddsoranı}} = \frac{1}{A} - \frac{1}{A} + \frac{1}{C} - \frac{1}{D}$$

$$S_{\text{LogRiskOranı}} = \sqrt{Var_{\text{LogRiskOranı}}}$$

$$GA\ln(OO) = \text{LogOddsOranı} \pm 1{,}96 \text{ x } S\text{LogOddsOranı}$$

Risk oranlarında olduğu gibi, bütün hesaplamaları log değerleriyle yaparız, geri dönerek sonuç değerlerini aşağıdaki ifadeler yardımıyla odds oranları cinsinden yazarız.

$$\text{Ortalama} = e^{\text{Ln}(OO)}$$
$$\text{Alt sınır} = e^{\text{alt sınırLog}_{\text{OddsOranı}}}$$
$$\text{Üst sınır} = e^{\text{üst sınırLog}_{\text{RiskOranı}}}$$

Hem RO hem de OO için niçin logaritmik ölçekte çalıştığımızı sorabilirsiniz. Bunun nedenlerinden biri her birinin dağılımının alt sınırı "0" ile sınırlanmış ve üst sınırı sınırsız olup pozitif kayışa sahip olmalarıdır. Her bir oranın log değerinin alınmasıyla yaklaşık normal dağılıma sahip bir istatistik elde ederiz. İkinci bir sebep de olasılık oranlarında olumlu ve olumsuz sonuçlardan biri diğerinin basitçe çarpmaya göre tersi olmakla birlikte, risk oranının olumlu ve olumsuz sonuçları arasında basit bir ilişkisidir.

## Etki Büyüklüklerinin Korelasyonlar Cinsinden İfadesi

Korelasyon katsayıları, sıklıkla etki büyüklükleri ölçüsü olarak kullanılır, fakat daha önce gördüğümüz gibi, evren parametresi "0" olmadıkça dağılımları kayışlıdır. Bu nedenle, $r'$ ile gösterdiğim Fisher'ın dönüşümünü kullanırız.

$$r' = 0.5 \times \ln\left(\frac{1 + r}{1 - r}\right)$$

$$s_{r'}^2 = \frac{1}{n - 3}$$

$$s_{r'} = \sqrt{\frac{1}{n - 3}}$$

Burada yine dönüştürülmüş değerleri meta-analizi hesaplamalarında temel alırız, daha sonra sonuçları orjinal kullanımlarına dönüştürürüz.

$$r = \frac{e^{2r'} - 1}{e^{2r'} + 1}$$

## Etki Büyüklüğü Ölçüleri Arasında Dönüştürmeler

Varsayalım ki siz, ben ve diğer bazı araştırmacılar yaklaşık denk çalışmalar gerçekleştirdik. Bunların, kesinlikle birbirinin tekrarı olmaları şart değil, ancak tümünün aynı genel soruya işaret eder nitelikte olması gereklidir. Sizin çalışmanız olasılık oranı ile sonlanmış, benimki iki grup ortalamaları arasındaki *t* testini içermiş ve diğer çalışma aynı şekilde korelasyon katsayısının hesaplanması olabilir. İdeal dünyada, tümü manidar sonuçlar oluşturabilir ve biz mutlu bir kişi olarak farklı şekillerde elde edilmiş tüm çalışmaların tutarlı ve birbirini destekler nitelikte olduğunu söyleyebiliriz. Fakat ideal bir dünyada yaşamıyoruz ve tutarsız,

**Tablo 17.1** Etki Büyüklüğü Ölçülerinin Dönüşümleri

LogOlasılık Oranının *d*'ye dönüştürülmesi

$$d = \text{LogOlasılıkOranı} \times \frac{\sqrt{3}}{\pi}$$

$$\text{var}(d) - \text{var}(\text{LogOlasılıkOranı}) \times \frac{3}{\pi^2}$$

*r* to *d*

$$d = \frac{2 \times r}{\sqrt{1 - r^2}}$$

$$\text{var}(d) = \frac{4 \times \text{var}(r)}{(1 - r^2)^3}$$

LogRiskOranının *d*'ye dönüştürülmesi

$$d = \ln\left(\frac{RR}{1 - RR}\right) \times \frac{\sqrt{3}}{\pi}$$

$$\text{var(d)} = \text{var(LogOlasılıkOranı)} \times \frac{\sqrt{3}}{\pi}$$

© Cengage Learning 2013

genellikle manidar olmayan sonuçlarla karşılaşırız Fakat, *t* istatistiğini, $\chi^2$'yi ve korelasyon katsayısını nasıl bir araya getireceğiz ve daha sonra 3'e mi böleceğiz? En basit cevabı, "bunu yapamayız" olur. Çalışmaları aynı ölçekte toplayacak ortak bir metriğe ihtiyacımız vardır. Bu kitap boyunca gördüğümüz etki büyüklükleri ölçülerinin bu ihtiyaca ideal bir şekilde uygun olduğu karşımıza çıkar. Bunun başlıca nedeni, bir etki büyüklüğü ölçüsünün diğerine nasıl dönüştürüleceğini bilmemizdir.

Bir etki büyüklüğü ölçüsünün diğerine dönüştürülmesi oldukça kolaydır. Tablo 17.1 bazı dönüştürmelerin nasıl yapıldığını göstermektedir. Eğer ihtiyaç duyduğunuz dönüşüm bu tabloda yer almıyorsa kısa bir internet aramasıyla istediğiniz dönüşüm bulunabilir. (Bu online hesaplayıcı P. D Ellis'te (2009) vardır. "Effect size equations," Web site: http://www.polyu.edu.hk/mm/effectsizefaqs/calculator.html erişim 1/23/2011.)

## 17.2 Bir Örnek – Çocuk ve Ergen Depresyonu

Çocuklarda ve ergenlerde depresyon, uzun süredir toplumumuzu ilgilendirmektedir ve ortaya çıkma sıklığını azaltmak üzere önlem programlarına ilişkin kuruluşlar tarafından oldukça fazla çaba gösterilmektedir. Bunlardan biri, bir hedef evrendeki tüm bireylere uygulanan "evrensel" programlardır. Örneğin, bölge okulu bu konuya ilişkin dersler düzenleyebilir veya bölgedeki tüm öğrencilere etki edecek şekilde öğretim programında değişikliklere gidebilirler. Bu programların bir kısmı "seçmeli"dir; bir başka deyişle, depresyon riski altında olduğu varsayılan daha küçük öğrenci gruplarını amaçlayacak şekildedir. Programlarda üçüncü bir kategori de "işaret eden" müdahaleler olarak sınıflanan, halihazırda klinik semptomlar gösteren öğrencileri amaçlar. Horowitz ve Garber (2006), bu şekildeki programlara ilişkin bir meta-analiz çalışması düzenlemiş, çalışmaların genel anlamda etkili olup olmadığını ve etkililiğin program sınıflamalarına göre farklılaşıp farklılaşmadığını araştırmışlardır. İlk araştırma, PsychINFO veri tabanından "depresyon (depression)" ve "önleme (prevention)" terimeleriyle yapılmıştır. Bu araştırma, yanlılığı en aza indirmek için tezleri de içermiştir. Daha sonra, yanlılığı en aza indirmek amacıyla 30-yıllık bir zaman aralığına ait 15 bilimsel dergiyi el ile taramışlardır. Kullanabilecekleri işlem ve kontrol grupları bulunan 30 çalışma bulmuşlardır. Bu çalışmaların hepsinde işlem

ve kontrol şartları vardır. Burada, veri toplama sürecine odaklanmayacak olmama rağmen, mekânikte meta analizin istatistiğine karşı olan geniş bir literatür vardır ve bu tür bir araştırma yapmayı planlayan biri, bu çalışmaları araştırmaya ihtiyaç duyar. [Cooper (2009), süreci kapsamlı bir şekilde kapsar; meta-analiz planlayanlar, başlamadan önce bu referansa veya benzerlerine başvurmalıdır. Yararlı diğer referanslar şunlardır: Borenstein ve ark. (2009) ve Cooper, Hedges ve Valentine (2009).]

Mevcut çalışmaları topladıktan sonra, yazarlarına, amaçlarına (evrenselliklerine, seçiciliklerine ve belirtilenlere), örneklem büyüklüklerine, yaş ortalamalarına, kadın yüzdelerine, uzunluklarına, müdahele sonrası etki büyüklüklerine, izleme süresine ve müdahele sonrası özetine göre sınıflandırırlar. Böyle bir sınıflama, çoğu meta-analizinde bir rutindir. Amacımız için biz, etki büyüklüklerine odaklanacağız. İki grubun yer aldığı çalışmaların etki büyüklüğü ölçülerini Bölüm 7'de tartışmıştık ve standartlaştırılmış ölçüler (standart sapma) olarak tanımlanan ölçülere alternatif bir yol olarak dikkate almıştık. Horowitz ve Garber, deney ve kontrol gruplarının son-test ortalama depresyon puanları arasındaki farkı, kontrol grubunun standart sapmasına bölerek etki büyüklüğünü hesaplamayı tercih etmişlerdir.

$$d = \frac{\overline{X}_{\text{kontrol}} - \overline{X}_{\text{deney}}}{s_{kontrol}}$$

Bu ölçü, daha önce Cohen'ın *d*'si olarak tanımladığımız değerdir (gerçekte öneren Glass'e özürlerimizle). Analiz sonuçları, Tablo 17.2'de verilmiştir. Her bir çalışma, bir deney ve bir kontrol grubu içerir. (Etki büyüklüğü ölçüsünü hesaplamak mümkün olmadığından, bu tablodan bir çalışma çıkarılmıştır.) Her bir grup için örneklem büyüklüklerini

**Tablo 17.2** Horrowitz ve Garber (2006)'dan sonuçlar

| Yazar | Hedef | *N* | $d_1$ | $d_2$ | $d_3$ |
|---|---|---|---|---|---|
| Clarke1 | U | 662 | 0,06 | -,06 | -,06 |
| Clarke2 | U | 380 | 0,09 | 0,14 | 0,14 |
| Kellam | U | 575 | -,01 | NA | NA |
| HainsEllman | U | 21 | 0,36 | -,04 | -,04 |
| Cecchini | U | 100 | 0,11 | -,15 | -,15 |
| Peterson | U | 335 | -,12 | NA | NA |
| Pattison | U | 66 | -,01 | 0,40 | 0,40 |
| Lowry-Web | U | 594 | 0,17 | NA | NA |
| Shochet | U | 260 | 0,39 | 0,25 | 0,25 |
| Spence | U | 1500 | 029 | 0,03 | 0,03 |
| Merry | U | 364 | 0,02 | -,13 | 0,05 |
| Gwynn | S | 60 | 1,37 | NA | NA |
| Roosa | S | 81 | 0,41 | NA | NA |
| Sandler | S | 72 | 0,24 | NA | NA |
| Wolchik | S | 94 | -,06 | NA | NA |
| Beardslee | S | 52 | 0,20 | 0,42 | 0,42 |
| Seligman | S | 235 | 0,32 | 0,12 | 0,25 |
| Quayle | S | 47 | -,62 | 0,62 | 0,62 |
| Cardemil1 | S | 49 | 0,99 | 1,24 | 1,24 |
| Cardemil2 | S | 106 | 0,16 | 0,31 | 0,31 |
| Jaycox | I | 143 | 0,18 | 0,32 | 0,20 |
| Clarke3 | I | 150 | 0,31 | 0,07 | 0,01 |
| Reivich | I | 152 | 0,12 | 0,40 | 0,22 |
| Lamb | I | 41 | 0,70 | NA | NA |
| Forsyth | I | 59 | 1,51 | 1,95 | 1,95 |
| Clarke4 | I | 94 | 0,41 | 0,47 | 0,04 |
| Yu-Selig | I | 220 | 0,23 | 0,30 | 0,30 |
| Freres1 | I | 268 | -,06 | 0,16 | 0,03 |
| Freres2 | I | 74 | 0,07 | 0,56 | 0,56 |

© Cengage Learning 2013

vermemişlerdir, bu sebeple her iki grup için eşitmiş gibi ayrıştırılmıştır. (Veri girişinde, kayıp veri NA olarak girilmiştir ancak bilgisayar programınıza bağlı olarak kayıp veriyi farklı bir sembolle girmeniz gerekebilir.) İlk etki büyüklüğü ölçüsü ($d_1$ ), müdahele bitiminde alınmıştır. İkinci etki büyüklüğü ölçüsü ($d_2$ ), olabildiğince 6 aylık müdahelenin bitiminde elde edilmiştir. Son sütundaki etki büyüklüğü ölçüsü ise ($d_3$ ), 6 aylık izleme sürecidir. 6-aylık izleme sürecindeki etki büyüklüğüne ($d_3$ ) ait kayıp veridir.

Programın uzun süreli etki sağlayıp sağlamadığıyla ilgilendiğimden, programın uzun dönemde etkili olup olmadığını öğrenmek istiyorum. Böylece, 21 çalışmadaki verinin bağımlı değişkenine ($d_2$ ) ilişkin sonuçları analiz edeceğiz.

Tablo 17.1'den, 6 aydaki etki büyüklüklerinin çoğunun pozitif olduğu görülebilir, ancak bazıları oldukça küçüktür. Tablo manidarlık testlerini göstermez, fakat etki büyüklüklerine ait güven aralıklarını hesapladığımızda bunları elde edebiliriz. Gerçekten, çoğunun pozitif olması, müdahelenin genelde pozitif etkisi olduğuna dair şüphelenmemize yol açar. Ancak bazı müdahale türlerinin diğerlerinden daha etkili olup olmadığını görebilmek için beklememiz gerekecektir.

## Orman Grafiği

**orman grafiği**

Bu sonuçların araştırılmasında kullanılan en basit yollardan birine **orman grafiği** adı verilir. Bu grafikler, her bir çalışmayı, çalışmanın etki büyüklüklerini ve etki büyüklüklerinin güven aralığını belirten ayrı bir çizgi ile gösterir. Bu grafik, Şekil 17.1'de görülmektedir. Kesikli dik çizgi "0" etki büyüklüğünü belirtir ki etki büyüklüklerini bu çizginin sağ tarafında görmek isteriz. Etki büyüklüklerinden uzakta sağda yer alan değerler, her bir çalışma için etki büyüklükleri ve etki büyüklüklerinin güven aralıklarıdır.[1] Bu grafik, R'nin metafor kütüphanesi kullanılarak oluşturulmuştur, fakat meta-analiz için bazı bilgisayar programları da benzer grafikleri verir. (Orman grafiğini sunan koda, Web'ten Horowitz.R. bölümünden ulaşılabilir.)

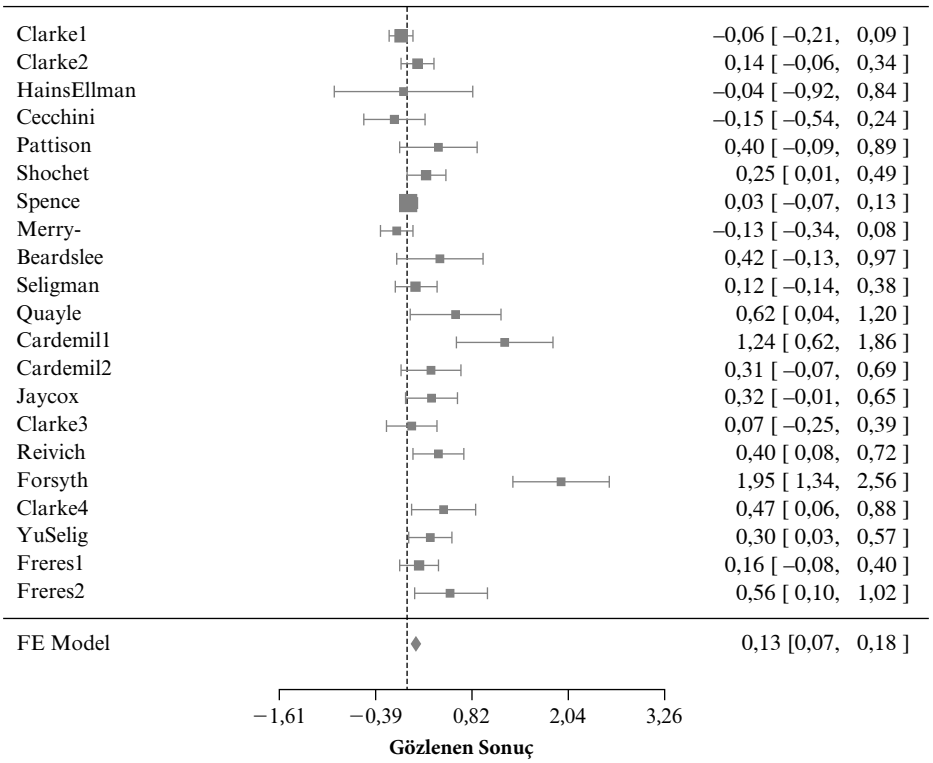
Kare kutular, çalışma sorusunun nokta kestirimlerini gösterir. Dikkat edilirse, her kutunun kendisiyle ilişkili güven sınırları vardır. Güven aralığı daraldıkça kestirimlerin de daha güvenilir olduğu açıkça görülmektedir. Genel olarak, aralığın genişliği ayrıca örneklem büyüklüğü ile de ilişkilidir; geniş örneklem daha kesin sonuçlara yönlendirir. Kutuların büyüklüklerinin de değiştiği fark edilmektedir. Bu değişim, doğrudan kestirimlerin kesinliğiyle de ilişkilidir.

## Sabit ve Tesadüfî Modeller

Devam etmeden önce sabit ve tesadüfî modeller arasındaki farkı da ortaya koymamız gerekir. Sabit etki modellerinde, gerçek bir etki büyüklüğü olduğunu varsayarız ve çoklu deneylerin sonuçlarına bakarak etki büyüklüğünü kestirmeye çalışırız.[2] Eğer bir astronom olsaydınız ve belirli bir yıldızın parlaklığını ölçme girişiminde bulunsaydınız, bir gerçek parlaklığı olduğunu ve sizin ölçümleriniz ile meslektaşlarınızın ölçümleri arasındaki farkın, hata varyansı olduğunu düşünmek mantıklı olurdu, bu farklılık belki ölçme aletinden, gökyüzünün gece karanlığının yoğunluğundan ve diğer faktörlerden kaynaklanabilirdi. Sabit etki modeline ilişkin önemli bir nokta da ölçümlerdeki değişkenlik için tek sebebin tesa-

1 Güven aralıklarının hesaplanmasını, bölümün ilerleyen kısmında açıklayacağım.

2 Viechtbauer (2010), kestirimleri tartışma konusu olan çalışma ile sınırlayacak olursak, evrendeki bütün etki büyüklüklerinin eşit olduğunu varsayma ihtiyacımız olmayacaktır. Ancak, böyle bir duruma çok ender rastlanır - hemen hemen daima evrendeki etkilerin daha geniş olduğunu düşünürüz.




© Cengage Learning 2013

**Şekil 17.1** Horrowitz ve Gabler'ın veri kümesine ait orman grafiği

düfî örneklem hatası olduğunu varsaymamızdır. Eğer çalışmalarımızın her birinde sonsuz çoklukta katılımcı olsaydı tüm çalışmalarda aynı sonuca ulaşırdık; çünkü tüm ölçümler aynı şeyi ölçüyor olurdu.

Davranış bilimlerindeki çoğu çalışmada yapılan ölçümler basit değildir. Örneğin depresyon; cinsiyete, yaşa, aile ile ilgili özelliklere ve diğer değişkenlere bağlı olabilir. Böylece, siz ve ben iki farklı örneklemden depresyon ölçümleri aldığımızda, gerçekten, depresyondan daha fazla bir şey ölçebiliriz. Eğer sizin örnekleminizde benimkinden daha fazla kadın katılımcı varsa sizin kestirdiğiniz gerçek etki büyüklüğü ile benim kestirdiğim farklı olabilir. Bu demektir ki ölçümlerimizdeki farklılıklar tesadüfî örneklem hatalarından daha fazladır. Tesadüfî etki modellerinde, gerçek etkilerin tesadüfî olarak ve bir değer etrafında normal dağıldığını varsayacağız.

Tesadüfî modelde, ölçümleri iki parça olarak aldığımızı varsayarız. Birinde, ölçümlerimiz tesadüfî hata içerir. Diğeri ise çalışmadan çalışmaya değişebilecek ve genel ortalama etkiye eşit olmayan gerçek etkilerdir. Bu model, varyans analizine oldukça benzer bir model verir.

$$Y_i = \mu + \tau_j + \varepsilon_{ij}$$

Burada $Y_i$, deney grubu $_i$'nin etki büyüklüğünü, $\mu$ genel etkiyi, $\tau_j$ gerçek etki ile çalışmada bizim ölçtüğümüz etkinin farkını ve genel ortalama gerçek etkiyi ve $\varepsilon_{ij}$ örneklem hatasını ifade eder. Bu demektir ki hem $\tau_j$'nin hem de $\varepsilon$'nın varyansını göz önüne almamamız ge-

rekir. Sabit modelde, $\tau$'yu kolayca göz ardı edebiliriz, çünkü tüm çalışmalar aynı nitelikte olduğu varsayıldığından herhangi bir anlamı yoktur.

Şimdi sabit ve tesadüfî modelleri birbirinden ayıralım, hangisini kullanmalıyız? Kararın açık olduğunu düşünebilirsiniz. Eğer depresyonu etkileyen çok sayıda değişkenimiz varsa ve farklı çalışmalar farklı katılımcılar ile gerçekleştirilmişse; örneklemin ve müdahalemizin özelliklerine bağlı olarak çok sayıda gerçek etki olduğunu düşünebiliriz. Bu durum, açıkça tesadüfî desene sahip olduğumuzu söyleyecektir. Bununla birlikte, Hedges (2009), kararın gerçekte bu kadar açık olmadığını savunur. Eğer üç farklı belirli psikoterapi türünü varyans analizi ile karşılaştırıyorsak; terapileri tesadüfî olarak almadığımızdan desenimizin sabit olacağını düşünebiliriz. Bununla birlikte, desen sabit olsa da olmasa da araştırmacı sonuçlarını, benzer diğer terapilere genellemek isteyebilir ve sabit desen kullansa da hiç kimse yüksek sesle bu duruma karşı çıkmayacaktır. Benzer problem, tesadüfî desende de karşımıza çıkar; çünkü bu desen, bağımsız değişkenlerimizin örneklemlendiği evrene genelleme yapmamıza izin verir, ancak bu evren nedir? Bu evren, aldığımız örnekleme ne derece yakındır? Tablo 17.2'de verilen çalışmada, her bir hedefin sabit fakat diğer hedeflerden farklı olduğuna dair mantıklı bir örnek sunar. Buradan yola çıkacağız. Eğer oldukça az sayıda çalışmamız varsa, desenimizin sabit olması olasıdır, çünkü gerçek farkları doğru şekilde kestireceğimiz yeterli güce (power) sahip değiliz.

## Genel Etkinin Hesaplanması

Tablo 17.3'teki sonuçlar, Tablo 17.2'deki veriden hesaplanan istatistiklerin bir özetidir. Sütunların her biri doğrudan, aşağıda verilen formüller ile elde edilen değerleri gösterir. Sütun 6'da, $W_i = 1/s^2_{d_i}$ olarak tanımlanan ağırlıkları görüyorsunuz. Bu ağırlıkları isteriz, çünkü her bir etki büyüklüğünü kesinliği oranında ağırlıklandırmak isteriz ki aslında bu, varyans kestiriminin çarpmaya göre tersidir. Analize, sabit model ile başlayacağız. Şimdilik, müdahele hedeflerine karşılık gelen tablodaki üç bölümü göz ardı edeceğiz. 21 çalışmanın tümünü inceleyerek başlayacağız.

*d*'yi, daha önce aşağıdaki gibi tanımlamıştık:

$$d = \frac{\overline{\mathrm{X}}_{kontrol} - \overline{\mathrm{X}}_{işlem}}{\mathrm{s}_{kontrol}}$$

Daha sonra d'nin varyansını ve standart sapmasını da aşağıdaki gibi ifade etmiştik:

$$s^2_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1)}$$

$$s_d = \sqrt{s^2_d}$$

$\mathrm{s}_d$ ve $s^2_d$ istatistikleri, *d*'nin varyans ve standart sapma kestirimleridir. (Varyanslar, Tablo 17.3'ün 5'inci sütununda gösterilmiştir.) Her çalışma bir *d* değerine sahiptir, ancak biz bu istatistikteki değişkenliği, daima yapacağımız tek tekrar olmasına rağmen, belirli bir çalışmanın sonsuz sayıda tekrarı üzerinden kestiririz. Bu, doğrudan bir çalışmanın ortalamasının standart hatasının kestirimini hesaplamaya benzer.

Bir sonraki adımda, her bir çalışma için ayrı ayrı kestirilen *d*'lerin güven aralıklarını hesaplarız:

$$GA_{,95} = d \pm 1{,}96 s_d$$

**Tablo 17.3** Tablo 17.2'deki veri kümesine ait istatistiklerin özeti

| Çalışma | Hedef | $n_1 = n_2$ | $d_i$ | $sd^2$ | $W_i$ | $W_i^* d$ | $W_i d^2$ | $W_i^2$ |
|---|---|---|---|---|---|---|---|---|
| Clarke1 | U | 331 | -0,06 | 0,006 | 165,426 | -9,9255 | 0,5955 | 27365,6154 |
| Clarke2 | U | 190 | 0,14 | 0,010 | 94,7678 | 13,2675 | 1,8574 | 8980,9395 |
| HainsEllm | U | 10 | -0,04 | 0,200 | 4,9990 | -0,2000 | 0,0080 | 24,9900 |
| Cecchini | U | 50 | -0,15 | 0,040 | 24,9299 | -3,7395 | 0,5609 | 621,4992 |
| Pattison | U | 33 | 0,40 | 0,061 | 16,1765 | 6,4706 | 2,5882 | 261,6782 |
| Shochet | U | 130 | 0,25 | 0,015 | 64,4961 | 16,1240 | 4,0310 | 4159,7500 |
| Spence | U | 750 | 0,03 | 0,002 | 374,958 | 11,2487 | 0,3375 | 140593,3647 |
| Merry- | U | 182 | -0,13 | 0,011 | 90,8082 | 11,8051 | 1,5347 | 8246,133 |
| TOPLAM | | | | | 836,561 | 21,4408 | 11,5133 | 190253,9603 |
| Beards1 | S | 26 | 0,42 | 0,078 | 12,7195 | 5,3422 | 2,2437 | 161,7866 |
| Seligma | S | 118 | 0,12 | 0,017 | 58,8940 | 7,0673 | 0,8481 | 3468,5022 |
| Quayle | S | 24 | 0,62 | 0,087 | 11,4498 | 7,0989 | 4,4013 | 131,0987 |
| Cardem | S | 24 | 1,24 | 0,099 | 10,0654 | 12,4811 | 15,4766 | 101,3128 |
| Cardem | S | 53 | 0,31 | 0,038 | 6,1854 | 8,1175 | 2,5164 | 685,6777 |
| TOPLAM | | | | | 119,314 | 40,1070 | 25,4861 | 4548,3779 |
| Jaycox | I | 72 | 0,32 | 0,028 | 5,5450 | 11,3744 | 3,6398 | 1263,4487 |
| Clarke3 | I | 75 | 0,07 | 0,026 | 37,4770 | 2,6234 | 0,1836 | 1404,5289 |
| Reivich | I | 76 | 0,40 | 0,026 | 37,2549 | 14,9020 | 5,9608 | 1387,9277 |
| Forsyth | I | 30 | 1,95 | 0,098 | 10,1673 | 19,8263 | 3,6613 | 103,3748 |
| Clarke4 | I | 47 | 0,47 | 0,043 | 22,8685 | 10,7482 | 5,0517 | 522,9702 |
| YuSelig | I | 110 | 0,30 | 0,018 | 54,3881 | 16,3164 | 4,8949 | 2958,0691 |
| Freres1 | I | 134 | 0,16 | 0,015 | 66,763 | 10,6858 | 1,7097 | 4460,4077 |
| Freres2 | I | 37 | 0,56 | 0,056 | 17,8022 | 9,9692 | 5,5828 | 316,9167 |
| TOPLAM | | | | | 282,239 | 96,4457 | 65,6846 | 12417,6439 |
| GENEL TOPLAM | | | | | 1238,164 | 157,9935 | 102,6840 | 207219,9982 |

© Cengage Learning 2013

Hatırlayacağınız gibi, kitabın önceki bölümlerinde, merkezî olmayan $t$ dağılımı kullanarak etki büyüklüklerinin güven aralığını hesaplamıştık. Bu hesaplama, en uygun yoldur. Bununla birlikte, meta-analizinde güven aralığının hesaplanmasında görülen standart yaklaşım $\overline{d} \pm s_d \times (z_{a/2})$'dir. Bir araya getirdiğimiz çok sayıda çalışma verildiğinde, tüm etkilerin güven aralıkları için bu yaklaşım muhtemelen yeterince doğrudur. Bununla birlikte, küçük örneklemli bireysel çalışmalar için $d$'nin güven aralıklarının hesaplanmasında bu yaklaşımın kullanılması sorgulanabilir (Hedges ve Olkin (1985), Hedges (2009)).

Horrowitz ve Gabler, her bir çalışma için $d$'yi ve $d$'nin standart hatasını hesaplamıştır. Bu istatistikleri elde ettikten sonra, şimdi de kesinliği daha fazla olan ektilere daha fazla ağırlık veren, her bir etki büyüklüğünün varyanslarının çarpmaya göre tersini alırız. Daha sonra, ortalama deney etkisini ($\overline{d}_i$) ve bunun standart hatasını ($s_{\overline{d}}$) tüm 21 çalışma için hesaplarız. Tablonun en alt satırındaki toplamları kullanarak, aşağıdaki değerleri elde ederiz:

$$W_i = \frac{1}{s^2_{d_i}} = \frac{1}{0,0006}, \frac{1}{0,010}, \cdots, \frac{1}{0,056}$$

$$\overline{d} = \frac{\sum W_i d_i}{\sum W_i} = \frac{157,9935}{1238,164} = 0,128$$

$$s_{\overline{d}} = \sqrt{\frac{1}{\sum W_i}} = \sqrt{\frac{1}{1238,164}} = 0,028$$

$$GA = d \pm 1,96(s_{\overline{d}})$$

$$GA_{\text{altsınır}} = 0,128 - 1,96(0,028) = 0,073$$

$$GA_{\text{altsınır}} = 0,128 + 1,96(0,028) = 0,183$$

$$0,073 \leq \delta \leq 0,183$$

Dikkat edilirse, bu güven aralığı "0"ı içermemektedir, böylece tüm etki manidardır. Şekil 17.1'e tekrar bakarsak, şeklin alt kısmına yakın bir yerde elmasa benzer bir şekil dikkatimizi çekecektir. Bu elmasın merkezi, kestirilen tüm etki büyüklüğüdür ve genişliği, tartıştığımız güven aralığıdır. SE (sabit etki - FE /fixed effect) ile isimlendirilmiştir çünkü sabit-etki modeli için hesaplanmıştır.

## Etki Büyüklüklerinin Heterojenliği

Şimdi, tüm çalışmalar üzerinden *d*'nin güven aralığına ve ortalamasına sahibiz ve tüm çalışmalarda kestirilen gerçek etkinin aynı olmadığını gösteren etki büyüklükleri kestirimlerinin, heterojen olup olmadığını belirlemek üzere veriyi daha yakından incelemek istiyoruz. Veri, iki sebepten dolayı heterojen olabilir. Üç hedef katılımcı, müdaheleye farklı tepkiler vermiş olabilir ki çalışma farklılıkları gerçekte Hedef farklılıklarıdır. Diğeri, aynı hedefli çalışmalar arasında dahî kestirilen parametrelerde farklılıklar olabilir veya her ikisinin bir karışımı da olabilir.

Heterojenliğin ölçüsünü Q istatistiği olarak alacağız. Bu istatistik aşağıdaki gibi tanımlanır:

$$Q = \sum_{i=1}^{k} W_i(d_i - \overline{d})^2 = \sum W_i d_i^2 - \frac{(\sum W_i d_i)^2}{\sum W_i}$$

Bu bölümde gördüğümüz pek çok formüle benzer şekilde, bu formül de oldukça karmaşık görünür. Formülün sonuçları, çoğunlukla istediğimiz gibi, *d*'nin daha kesin kestirimlerine daha fazla ağırlık verir. Q basitçe, ortalama etki büyüklüğünden her bir çalışma için etki büyüklüklerinin kareler sapmasının ağırlıklı toplamlarıdır. Etki büyüklükleri değişkenliği arttıkça Q'nun değeri de artar. Varyans analizindeki $SS_{\text{gruplar-arası}}$ ile benzerdir. Ağırlıkları (varyansların çarpmaya göre terslerini) kullandığımız için aynı zamanda standartlaştırılmış bir ölçüdür. Bu örnekte Q, *d* kullanılarak hesaplanmasına rağmen diğer etki büyüklükleri ölçüleri de alınabilir ve aynı formül kullanılabilir. Örneğimiz için Q aşağıdaki gibi elde edilir:

$$Q = \sum W_i d_i^2 - \frac{(\sum W_i d_i)^2}{\sum W_i}$$

$$= 102,6840 - \frac{157,9935^2}{1238,164} = 82,5235$$

*Q* ile ilgili bir güzel durum da, *Q*'nün etki büyüklüklerinde farklılık olmaması, yokluk hipotezi altında, çalışma sayısı *k* olmak üzere *k*-1 serbestlik derecesinde $\chi^2$ dağılımına sahip olmasıdır. 21 çalışma ile ilgili $\chi^2(20) = 31,41$'dir. 82,52 > 31,41 olduğundan yokluk hipotezini reddederiz ve kestirilen gerçek etkinin tüm çalışmalarda aynı olmadığı sonucuna

varırız. Belirtmeliyim ki meta-analizde genel görüş, kararın, tesadüfî modellere karşı sabit modellere göre verilmemesidir -karar, çalışmanın kendisine dayanan esas bilgiye göre verilmelidir- fakat genel bir test hâlâ değerlidir. Meta-analizini, tek-yönlü varyans analizindeki genel *F*'e benzetebiliriz.

Sonuçlarımıza göre tesadüfî modele sahipsek, açık olarak tesadüfî olduğunu belirtmesek de, gerçek etkilerin varyansını ($\sigma_\tau^2$) hesaplayabiliriz. Sabit modelde sadece bir gerçek etki vardır, fakat tesadüfî modeller için gerçek etkilerin varyansı manidar bir istatistiktir. $\sigma_\tau^2$'yu kestirmek için DerSimonian ve Laird metodu olarak bilinen bir yöntem kulanırız. Bu, optimum bir ölçü vermemekle birlikte hesaplaması oldukça kolaydır. Eğer $\tau$'yu çalışmanın gerçek etkisinin, ortalama gerçek etkiden sapması olarak ifade edersek, ki bu varyans analizinde yapılana kesinlikle benzerdir, $\tau$'nun varyansının kestirimi aşağıdaki gibi tanımlanır:

$$T^2 = \frac{Q - sd}{C}$$

Burada;

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

Örneğimiz için,

$$Q = 82{,}5235$$

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i} = 1238{,}164 - \frac{207219.9982}{1238{,}164} = 1070{,}8033$$

$$T^2 = \frac{Q - df}{C} = \frac{82.5235 - 20}{1070.8033} = 0{,}0584$$

$$T = \sqrt{T^2} = \sqrt{0.0584} = 0.2416$$

bulunur.

$T^2$'nin tanımı göründüğü kadar garip değildir. *sd* yokluk hipotezi doğru olduğunda beklenen değişkenlik olmak üzere *Q*, etki büyüklükleri ölçüleri arasındaki farklılığı ölçer. Böylece pay, çalışmalar arası tesadüfî farklılıklara atfedilemeyecek aşırı değişkenliktir. *C* terimini, varyans analizindeki grup-içi'ne benzer olarak düşünebilirsiniz, fakat aynı zamanda C'nin, $X - \mu$'nun standart sapmayı standartlaştıran *d*'ye bölünmesiyle olduğu kadar standartlaştırılmış olduğunu da düşünmelisiniz. Borenstein ve ark.nın (2009) ifade ettiği gibi, çalışmalar arası ağırlıklandırılmış kareler toplamı ölçüsü olan *Q* ile başlarız ve meta-analizdeki çalışmalarımızın sayısıyla uyumlu olan *sd*'yi çıkarır ve orjinal metriği geride bırakarak standartlaştırmayı sağlayan *C*'ye böleriz.

## Alt Grup Farkları

Farklı çalışmalardan farklı gerçek etki büyüklükleri kestirileceğine inanmak için iyi bir sebebimiz olduğu ve Hedef katılımcının bu etki farklılıkları ile ilişkili olup olmadığını sormanın mantıklı olduğu sonucuna vardık. Hedef katılımcılar-içi etki büyüklükleri için, ancak katılımcılar-arası değil, sabit modeli tutmaya devam ederiz. Bu durumu araştırmak üzere, hedefin her bir düzeyi için ayrı ayrı *Q* hesaplayacağız ve daha sonra bu *Q* değerlerinin birbirinden manidar derecede farklı olup olmadığını soracağız.

Her bir katılımcı türü için $Q$'yu hesaplamada gerekli olan bilgi, Tablo 17.2'deki satırlarda belirtilmiştir. $Q$'nun alt indisleri Hedefleri göstermek üzere, bu hesaplamalar aşağıda gösterilmiştir:

$$Q = \sum W_i d_i^2 - \frac{(\sum W_i d_i)^2}{\sum W_i}$$

$$Q_1 = 11{,}5133 - \frac{21{,}4408^2}{836{,}561} = 10{,}9638$$

$$Q_2 = 25{,}4861 - \frac{40{,}1070^2}{119{,}314} = 12{,}0043$$

$$Q_3 = 65{,}6846 - \frac{96{,}4457^2}{282{,}289} = 32{,}7334$$

Şimdi de katılımcı türleri için $Q_i$ grup-içi toplamları ifade eden $Q_{\text{grup-içi}}$ değerini bulalım:

$$Q_{\text{grup-içi}} = Q_1 + Q_2 + Q_3 = 10{,}9638 + 12{,}0043 + 32{,}7334$$
$$= 55{,}7015$$

Daha önceki hesaplamalardan, 21 çalışma için $Q$ değerini $Q_{\text{toplam}} = 82{,}5235$ bulmuştuk. Buradan;

$$Q_{\text{gruplar-arası}} = Q_{\text{toplam}} - Q_{\text{grup-içi}}$$
$$= 82{,}5235 - 55{,}7015$$
$$= 26{,}822$$

olarak bulunur.

Tüm bu yaptıklarımız oldukça tanıdık gelebilir, çünkü bu hesaplamalar varyans analizine doğrudan benzerdir. Tüm *gruplar* için etki büyüklüğünün aynı olduğu yokluk hipotezi altında, $Q_{\text{gruplar-arası}}$, $g - 1$ ($g$: grup sayısı) serbestlik derecesinde ki-kare dağılımı gösterir. Örneğimiz için $\chi^2(2) = 5{,}99$'dur, böylece yokluk hipotezini reddederiz ve hedef katılımcıların, deneylerin etkililiğinde bir faktör olduğu sonucuna varırız. Ayrıca, $Q_i$'lerin ayrı ayrı değerleri, bir kümedeki çalışmaların sayısının bir eksiğine eşit serbestlik derecesinde ki-kare dağılımı gösterirler. Böylece kritik değerler, $\chi^2(7) = 14{,}07$ ve $\chi^2(4) = 9{,}49$ olarak elde edilir ve hem Seçici hem de Belirtilen durumlarda homojenlik varsayımını reddederiz. Açıktır ki müdahelenin etkileri, müdahelenin katılımcıya farklı yollarda verilmesi etkisinden daha fazladır.

## Tesadüfî Model

Biraz da tesdüfî modellerden söz edeceğim ve bu konu ile ilgilenen okuyuculara da Borenstein ve ark.nın (2009) çalışmasından bahsedeceğim. Bilindiği gibi tesadüfî modellerdeki başlıca farklılık, sonuçlarımızdaki değişkenliğin iki kaynağının olmasıdır. Bunlardan biri, sabit modellerdeki gibi, çalışmada yer alan katılımcılar arası farklılıklar; diğeri de farklı çalışmaların farklı gerçek etki kestirimlerinin olacağı varsayımımıza bağlı farklılıklardır. Her bir çalışma için ağırlıklandırmaları, her iki değişkenliği içerecek şekilde hesaplayacağız. Geri kalan hesaplamalar tamamen aynı olacak.

Sabit model için tek değişkenlik kaynağının hata varyansı olduğunu varsaydık. Log($RO$) gibi diğer etki büyüklükleri de olmasına rağmen, etki büyüklükleri ölçüsü için $d_i$'yi kullanarak aşağıdaki eşitliği elde ettik:

$$W_i = \frac{1}{s_d^2}$$

$$\overline{d} = \frac{W_j d_i}{\sum W_i}$$

Tesadüfî model için değişkenlik, $s_d^2$ olarak kestirilen hata varyansınızdan oluşturulan ve her çalışmada kestirilen ($T^2$) gerçek etkilerdeki değişkenliktir. Böylece çalışmaların ağırlıklandırılması farklı bir şekilde gerçekleştirilir ve aşağıdaki eşitlik kullanılır:

$$W_i^* = \frac{1}{s_d^2 + T^2}$$

"W" üzerindeki "*" işareti, sabit modeldeki ağırlıklandırmadan ayrılması için kullanılır. Şimdi de ağırlıklandırılmış ortalamayı ve etki büyüklükleri kestirimlerinin standart hatasını aşağıdaki şekilde tanımlayacağız:

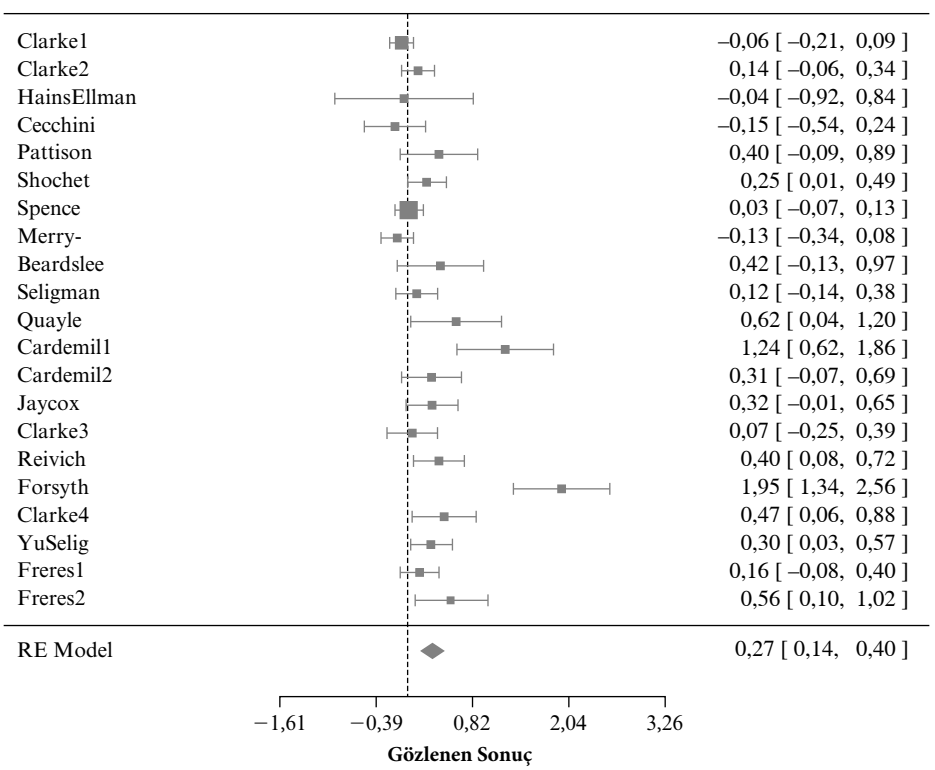
$$\overline{d} = \frac{\sum W_i^* d_i}{\sum W_i^*}$$

$$s_{\overline{d}} = \sqrt{\frac{1}{\sum W_i^*}}$$

Sonraki hesaplamaların tamamında, sabit model için hesaplanan değerler yerine bu değerleri kullanırız. $\overline{d}$'nın güven aralığını hesaplarken yaklaşım olarak merkezî-olmayan *t* yerine *z* kullanırız ve böylece aşağıdaki eşitliği elde ederiz:

$$\overline{d} \pm 1.96(s_{\overline{d}})$$

Biraz önce yaptığımız hesaplamaları tekrarlamak yerine, sadece ağırlıkları farklı olan aynı hesapları yaparak Şekil 17.2'deki sonuçları elde ederiz. Bu şekil ile Şekil 17.1 arasındaki tek farklılık, tablonun altında yer alan güven aralıklarının tesadüfî model için kesti-




© Cengage Learning 2013

**Şekil 17.2** Tesadüfî modeli için orman grafiği

rilmiş olmasıdır. Dikkat edilirse, kestirilen özet etki, 0,13 yerine burada 0,27'dir, ancak güven aralığı dikkate değer ölçüde geniştir (0,14; 0,40) ki bu durum belirsiz bir kaynağın eklendiğini yansıtır.

## 17.3 İkinci Bir Örnek – Nikotin Sakızı ve Sigara Bırakma

Psikologlar, çok uzun süredir bireylerin sigarayı bırakabilmelerinde yardımcı olmakla ilgilenmektedir ve son 20 yıldır bu amaçla yaygın olarak kullanılan nikotin sakızları ve petlerini görmekteyiz. Bu tür desteklerin etkililiğine ilişkin çok sayıda meta-analiz çalışması Cochrane veri-tabanında yayımlanmış olup bunlardan biri Stead, Perera, Bullen, Mant ve Lancaster T. (2008) çalışmasıdır. Bu araştırmacılar, 132 farklı çalışma üzerinden yapılmıştır. Bunlardan 53'ü nikotin sakızı kullananlar ve kontrol gruplarını içerir. Çok yer kaplamaması bakımından, bu çalışmaların sadece yarısını kullanacağız, ancak sonuçlarımız kesinlikle araştırmacıların sonuçları ile aynı olacaktır. (Stead ve diğerleri, çalışmaların ağırlıklandırılması için Mantel-Haenzsel metodu olarak bilinen bir yöntem kullanmışlardır, böylece bizim sonuçlarımız bir miktar farklılık gösterecektir, fakat bu farklılık önemli ölçüde değildir.) Nikotin sakızı kullananlar ile kontrol grubunun karşılaştırıldığı 26 çalışmaya ait veri Tablo 17.4'te gösterilmiştir.

**Tablo 17.4** Sigara bırakmaya yardımcı olarak nikotin sakızı verileri
(BaşarıT ve ToplamT ifadeleri, deney grubundaki başarı sayısını ve toplam durum sayısını temsil etmektedir.)

| Çalışma | Yıl | BaşarıT | ToplamT | BaşarıC | ToplamC | RO | $Log_{RO}$ |
|---|---|---|---|---|---|---|---|
| Ahluwalia | 2006 | 53 | 378 | 42 | 377 | 1,23 | 0,20 |
| Areechon | 1988 | 56 | 99 | 37 | 101 | 1,35 | 0,33 |
| Blondal | 1989 | 30 | 92 | 22 | 90 | 1,25 | 0,22 |
| BrThorSociety | 1983 | 39 | 40 | 111 | 1208 | 1,03 | 0,03 |
| Campbell | 1987 | 13 | 424 | 9 | 412 | 1,39 | 0,33 |
| Campbell | 1991 | 21 | 107 | 21 | 105 | 0,98 | -,02 |
| Clavel | 1985 | 24 | 205 | 6 | 222 | 3,98 | 1,38 |
| Clavel-Chapelon | 1992 | 47 | 481 | 42 | 515 | 1,18 | 0,17 |
| Cooper | 2005 | 17 | 146 | 15 | 147 | 1,13 | 0,12 |
| Fagerstorm | 1982 | 30 | 50 | 23 | 50 | 1,19 | 0,17 |
| Fagerstorm | 1984 | 28 | 96 | 5 | 49 | 2,44 | 0,89 |
| Fee1982 | 1982 | 23 | 180 | 15 | 12 | 1,41 | 0,35 |
| Fortmann | 1995 | 110 | 552 | 84 | 522 | 1,20 | 0,18 |
| Garcia | 1989 | 21 | 68 | 5 | 38 | 2,03 | 0,71 |
| Garvey | 2000 | 75 | 405 | 17 | 203 | 2,02 | 0,70 |
| Gilbert | 1989 | 11 | 112 | 9 | 111 | 1,19 | 0,18 |
| Gross95 | 1995 | 37 | 131 | 6 | 46 | 1,91 | 0,65 |
| Hall85 | 1985 | 18 | 41 | 10 | 36 | 1,40 | 0,34 |
| Hall87 | 1987 | 30 | 71 | 14 | 68 | 1,74 | 0,55 |
| Hall96 | 1996 | 24 | 98 | 28 | 103 | 0,92 | -,08 |
| Harackiewicz | 1988 | 12 | 99 | 7 | 52 | 0,91 | -,09 |
| Herrera | 1995 | 30 | 76 | 13 | 78 | 1,98 | 0,68 |
| Hjalmarson | 1984 | 31 | 106 | 16 | 100 | 1,64 | 0,50 |
| Huber88 | 1988 | 13 | 54 | 11 | 60 | 1,25 | 0,23 |
| Hughes | 1989 | 23 | 210 | 6 | 105 | 1,83 | 0,60 |
| Hughes | 1990 | 15 | 59 | 5 | 19 | 0,97 | -,03 |

Yukarıdaki veriye ait orman grafiği aşağıda gösterilmiştir. (Dikkat edilirse, sağ tarafta verilen etki büyüklükleri, risk oranlarının kendileri yerine log risk oranlarıdır.) Dikkat edeceğiniz gibi, nerdeyse tüm çalışmaların etki büyüklükleri risk oranı için 1,0'den veya log risk oranı için 0'dan büyüktür. Bu sonuç, nikotin sakızının sigarayı bırakmaya yardımcı olduğu izlenimi vermektedir.

Standartlaştırılmış ortalama farklar yerine risk oranları ile çalışıyor olmamıza rağmen, hesaplamalarımızda sadece küçük değişiklikler olacak. İlk olarak, hemen risk oranlarını log risk oranlarına dönüştüreceğiz. Bundan sonra da standart metriğe çevirene dek sonuna kadar logaritmik değerler ile çalışacağız.

Risk oranı aşağıdaki şekilde tanımlanır:

$$\text{RiskOranı} = \frac{\text{BaşarıT/ToplamT}}{\text{BaşarıC/ToplamC}}$$

$$\text{LogRiskOranı} = \text{In}(\text{RiskOranı})$$

Yakın bir kestirimle, varyans ve standart hata aşağıdaki gibi verilir:

$$\text{Var}_{LogRiskOranı} = \frac{1}{\text{BaşarıT}} - \frac{1}{\text{ToplamT}} + \frac{1}{\text{BaşarıC}} - \frac{1}{\text{ToplamC}}$$

$$s_{\text{LogRiskOran}} = \sqrt{\text{Var}_{\text{LogRiskOran}}}$$

Risk oranı ile bunun varyans ve standart sapmasının hesaplanmasıyla uğraşmayacağız, çünkü bunlara ihtiyacımız olmayacak.

Ancak, ortalama risk oranı için özet istatistikleri ve güven aralığını hesaplamaya ihtiyacımız olacak. Bunu yapabilmek için logaritmik değerlerle çalışacağız. *d* için log risk oranı kullanmanın dışında, daha önce kullandığımız temel formülleri kullanacağız.

$$\text{Ortalama}_{\text{LogRR}} = \frac{\sum W_i \text{LogRR}_\text{i}}{\sum W_i} = \frac{73{,}766}{267{,}416} = 0{,}278$$

$$\text{Var}_{\text{LogRR}} = \frac{1}{\sum W_i} = ,0037$$

$$\text{Shata}_{\text{LogRR}} = \sqrt{Var_{LogRR}} = \sqrt{,0037} = ,06$$

$$GA = \text{ortalama}_{\text{Log}RR} \pm 1{,}96(Shata_{\text{Log}RR}) = ,276 \pm 1{,}96(,06) = ,276 \pm 0{,}118$$

$$GA_{\text{alt}} = 0{,}158$$

$$GA_{\text{üst}} = 0{,}394$$

$$e^{\text{ortalama}} = e^{,276} = 1{,}32$$

$$e^{GA_{\text{Alt}}} = e^{,158} = 1{,}14$$

$$e^{GA_{\text{üst}}} = e^{,294} = 1{,}478$$

Yukarıdaki en son adımda, risk oranının logaritmik değerleri orjinal değerlerine dönüştürüldü. Böylece, risk oranı için güven aralığını

$$1{,}17 \leq \text{RiskOranı} \leq 1{,}48$$

eşitsizlikleriyle yazabiliriz. Risk oranımız 1'den manidar olarak büyüktür ve bu durum, nikotin sakızı kullanan gruptaki başarı olasılığının daha yüksek olduğunu gösterir ve aralığın alt sınırı, bu grup için başarı şansının %25 daha yüksek olduğunu yansıtır. Yapacaklarımız

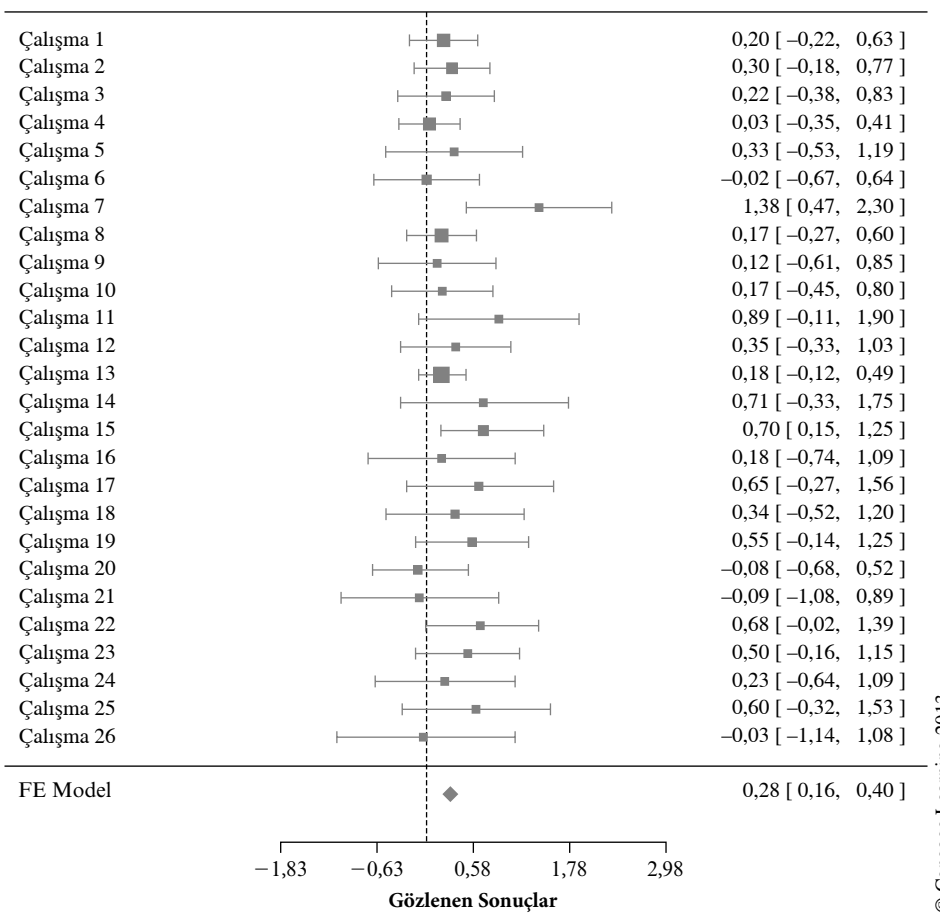
henüz bitmedi, çünkü bir de çalışma etkileri için heterojenlik testi yapmayı isteriz. Tüm çalışmalar yeterince birbirine benzer olduğundan sabit model olduğunu varsaymak mantıklıdır. Daha önceki çalışmada yaptığımız gibi $Q$ değerini hesaplayacağız ve sonra bu değeri, $k - 1 = 25$ serbestlik derecesinde $\chi^2$ ile karşılaştıracağız.

$$Q = \sum W_i(\log RR)_i^2 - \frac{(\sum W_i \log RR_i)^2}{\sum W_i} = 39{,}570 - \frac{5441{,}402}{276{,}4161} = 19{,}22$$

$$\chi^2_{(25,.05)} = 37{,}65$$

Etki büyüklüklerinin homojenliği varsayımını reddetmeyeceğiz. Bu sonucu, görsel bir izlenim sunan Şekil 17.3 büyük ölçüde onaylar.

Buraya kadar yapılanlar meta-analizini kapsar ancak burada yapacağımız literatürde rahatlamamızı sağlayacaktır. Tesadüfî ve sabit modeller arasındaki temel farkı biliyoruz, ortalama etki büyüklüğünü, standart hatasını hesaplayabiliriz ve etki büyüklüğü için güven aralığını oluşturabiliriz. Meta-analiz için gerçekten en zor olan kısım tüm verinin toplanmasıdır. Geri kalanının üstesinden kolayca geliriz. Kesinlikle çok fazla formül var ancak sistematik bir şekilde çalışılırsa bunlar karmaşık gelmeyecektir ve hiç kimse de sizden bu formüllerin tümünü hatırlamanızı beklemeyecektir. Yukarıda sözü edilen kaynaklar, çok




© Cengage Learning 2013

**Şekil 17.3** Nikotin sakızı çalışması için orman grafiği

iyi çalışmalardır ve ihtiyacımız olan bilgiyi sağlar. Ayrıca, bilgisayar programları sadece hesaplamaları yapmakla kalmaz; her bir çalışmaya ilişkin verinin organize edilmesine ve etki büyüklüklerinin, cinsiyet veya yaş gibi diğer değişkenlere bağlı olup olmadığının incelenmesine imkan tanır. En iyi paket programlardan biri; Borenstein, Hedges, Higgins ve Rothstein tarafından geliştirilen Comprehensive Meta-Analysis (CMA) programıdır. Bu program, ücretsiz bir yazılım değildir, ancak üniversitenizin bunun lisansını alacağını ümit edebilirsiniz. Buna ek olarak Cochrane Collaboration, RevMan adında ücretsiz bir bilgisayar programı geliştirmiştir ve bu program istediğimiz ve daha fazla her şeyi yapacaktır. William Shadish, meta-analiz için uygun olan bilgisayar programlarının bir listesini http://faculty.ucmerced.edu/wshadish/Meta-AnalysisSoftware.htm adresinde vermiştir. Bu liste, pek çok programı kapsar. Son olarak R, meta-analiz için en az üç bilgisayar programına sahiptir ve bunlardan biri olan "metafor" benim favorimdir. Örneklerde kullandığımız iki analizin yapıldığı programlara, kitabın web sayfasından ulaşılabilir.

# TEK-DURUM DESENLERİ

## 17.4 Standartlaştırılmış Ortalama Farkların Araştırıldığı Analizler

Davranış bilimlerindeki bazı alanlarda araştırma desenlerinden "tek-denek desenleri" veya "tek-durum desenleri" adı verilen desenlere sıkça rastlanır. Örneğin; davranışsal problemlerle, öğrenme veya dikkat bozuklukları ile ilgili çalışmalarda, tek durumun ele alınması oldukça yaygındır, dikkat dağınıklığı gibi davranışlar belirli periyotlar boyunca ölçülür, istenilen davranış için pekiştireç gibi bazı müdahalelerde bulunulur ve izleyen müdahele süreçleri boyunca istenmeyen davranışın ölçülmesine devam edilir. Davranış probleminin gözlendiği birkaç birey bulunabilir ancak bu problemler tam olarak aynı değildir. Bir çocuk, gözünü gökyüzene dikerek bakarken bir diğeri yanındakini dürter ve bir diğeri de sırasını karalar. Bu sebeple, farklı davranışları basitçe bir araya getiremeyiz ve çocuklar üzerinden bir ortalama alamayız - her çocuğu kendi durumunda değerlendirmemiz gerekir. Bir anlamda, "tek-durum deseni" yanıltıcıdır çünkü çalışmaların çoğu, birkaç farklı durumla ilgilidir. Ancak biz, her bir çocukla kendi durumunda ilgileniriz. Böylece sorunumuz, birkaç farklı çalışmaya ait veriyi tek bir sonuca ulaşacak şekilde nasıl analiz edebilirize dönüşür. Bu konu, yıllardır üzerinde anlaşmazlıkların olduğu bir alandır ve muhtemelen böyle olmaya da devam edecektir, ancak bildiğimiz bazı noktalar vardır ki literatürü bir araya getirmemize izin verir ve mantıklı bazı tavsiyelerde bulunabiliriz. Ortaya çıkan bu konulardan bazıları geleneksel gruplar-arası desenlerde de ortaya çıkar, ancak bu durumda, araştırmacılar arasında analizlerin nasıl yapılacağına ilişkin genel bir hemfikir bulunur. Örneğin, anoreksiyayı araştırıyorsak, kontrol ve deney gruplarına tesadüfî atama yapabilir, bir müdahelede bulunabilir ve daha sonra deneyin sonucunda katılımcılarımızın her birini ağırlıklandırabiliriz. İki grubun ortalamaları için *t* testi uygulamak mantıklı bir yaklaşım olarak görülecektir ve bu analize ilişkin araştırmacılar arasında çok fazla bir anlaşmazlık olmayacaktır. Ancak tek-durum çalışmalarında analiz, bu kadar belirgin değildir.

**A-B desenleri**

Bu bölümde, daha karmaşık desenlerin sonuçlarının analizinde temel bir biçim, düzen oluşturacağından, **A-B desenleri** adı verilen desenlere odaklanacağız. A-B deseninde, esas alınan bir noktadan itibaren çok sayıda denemelerden (A) veri toplanır ve sonra bir müdahalede bulunulur ve ardından da geri kalan denemelerden (B) veri toplanır. A-B-A deseninde, A-B desenine bir üçüncü adım daha eklenmiştir, bu adımda müdahele uygulanmaz ve

performansın esas alınan noktaya yakın bir düzeye dönmesi beklenir. Bu desene gerileme (withdrawal) deseni adı da verilir. A-B-A-B deseninde, dördüncü adımda müdahele tekrar uygulanır. Bu bölümde ele alacağımız analizler bu tür daha karmaşık desenlere uzanacaktır. Bu tür desenlerin avantajı, müdahelenin etkili olduğuna dair sonuca zamanından önce varılmasını önlemesidir. Tüm tek-durum araştırma desenlerine ilişkin çok mükemmel bilgi Franklin, Allison ve Gorman'ın (1996) çalışmasında bulunabilir.

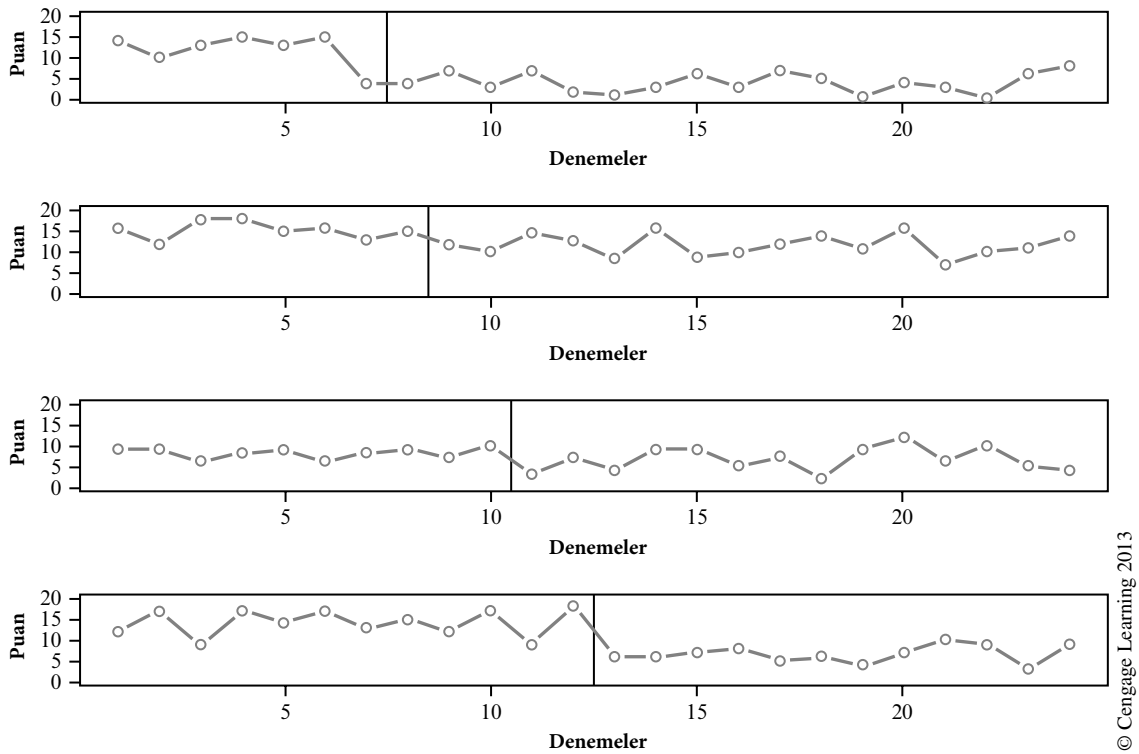
Tek-durum çalışmalarındaki en büyük konulardan biri performans değişimlerindeki ölçümler için kullanılacak istatistiği seçmektir. Örneğin, çalışmanın her bir adımındaki ortalama performansları karşılaştırabiliriz. Aynı zamanda, denemelerdeki bağımlı değişkenle ilişkili doğruların eğimlerini veya esas alınan adımdaki süreçte denemelerin ortancasını aşan müdahele sürecindeki denemelerin yüzdesini karşılaştırabiliriz veya daha başka pek çok istatistiği hesaplayabiliriz. Veriyi nasıl ele alacağımızı belirlemek önemli bir iştir. Bu bölümde, birbirinden oldukça farklı iki yaklaşımı göstereceğim. Bunlardan ilki, ortalama farklarla, diğeri eğim ve kesim noktasındaki değişimlerle ilgilidir. Burada, iki analiz birbirinden oldukça farklı bulunacaktır. Bunlar arasından seçim yapmak, zaman sürecindeki performansta nasıl bir değişim beklendiğine bağlıdır ve her bir araştırıcı ayrı kararlara varır.

## 17.5 Depresyon İçin Bir Durum Çalışması

Depresyon geçiren hastalara uygulanan yeni bir ilâcın etkilerini araştırmakla ilgilenen bir psikiyatrist örneğiyle başlayacağız. Araştırmacı, deneyin kontrol altında gerçekleştirilmesine inanıyorsa, müdaheleden önce altı veya fazla hafta boyunca her bir hastanın depresyon düzeylerine ilişkin müdahele-öncesi veri toplar. Daha sonra hastalara, araştırılan ilâcı vermeye başlar ve izleyen en az altı hafta boyunca elde ettiği veriyi kaydeder. Alınan ilâcın vücutta etki göstermesi biraz zaman alacağından ve bu nedenle etkileri zamanla artacağından, ilâca başlandıktan sonra ilk üç hafta veri toplanmaz. Bildiğimiz gibi, insanların nasıl hissettikleri zaman içinde değiştiğinden ne ilk ölçümlerin tümünün ne de müdaheleden sonraki ölçümlerin aynı olması beklenmez. Bu durum, evre-içi (within-phase) değişkenliktir. Bununla birlikte, ölçümleri haftalara ayırırsak, elde edilen değerler arasında ciddi bir oto-korelasyon olması ön görülmez. Diğer bir deyişle, adım-içi veride bir eğilim yoktur ve ardışık gözlemler en azından büyük ölçüde bağımsızdır. Oto-korelasyonun niçin sıfıra yakın olmasını istediğimizi kısaca açıklayacağım. Örneğimizdeki psikiyatristin çalışmaya gönüllü gelen dört hastası vardır, ancak bunların her birine farklı zamanlarda müdaheleye başlamış; her birinden müdaheleden önce farklı sayılarda ölçümler alınmış ve belki de ilâca tepkilerine göre farklı dozajlarda ilâç uygulanmıştır. Bu çalışmanın deseni, adım A esas süreç, adım B müdaheleden sonraki süreç olmak üzere A-B desenidir. Bu hipotetik çalışmaya ait veri Şekil 17.4'te gösterilmiştir.

Buradaki gibi, bir veriye en genel yaklaşım yollarından biri, her bir durumdaki müdahele öncesi ve sonrası ölçümler arasındaki standartlaştırılmış ortalama farkın ($d$) hesaplanmasıdır. Ham ortalama farka göre standartlaştırılmış ortalama fark ($d$) daha avantajlıdır; çünkü her bir hasta, ortak bir ölçeğe yerleştirilir. Böylece, hastaların depresyon düzeyleri arasındaki farklılıklar, analizi etkilemez. Bu yaklaşım oldukça yaygın olarak ve uygun bir şekilde kullanılıyor olsa da bu konuya daha sonra döneceğiz ve pek çok çalışmada bu yaklaşımın niçin en iyi yaklaşım olamayabileceğini tartışacağız.

Bu çalışmadaki bağımlı değişkenimiz haftalık depresyon düzeyleridir ve sonuç olarak elde ettiğimiz istatistik, her bir hasta için müdaheleden önceki ve sonraki adımlar arasındaki etki büyüklüğüdür ($d$). $d$'nin hesaplanması için standart sapmayı kullanırız. Esas adımdaki (müdahele öncesi) standart sapmayı alabiliriz, iki adım (müdahele öncesi ve son-




© Cengage Learning 2013

**Şekil 17.4** Dört hasta için depresyon puanları. Dikey çizgiler müdahale noktalarını gösterir.

rası) için ayrı ayrı standart sapmaları birleştirmeden veya adımlar üzerinden birleştirilmiş standart sapmayı alabiliriz. Biz sonrakini kullanacağız, fakat, eğer, müdahele boyunca fark edilir ölçüde daha fazla değişkenlik bekliyorsak esas adımdaki (müdahele öncesi) standart sapmanın kullanılması daha iyi olabilir. Sonuçta, her bir hastayı ayrı bir durum olarak alacağız ve sadece en sonunda tüm durumları birleştireceğiz.

Tablo 17.5'te, her bir hasta için *d*'yi görüyoruz. Daha önce söylediğim gibi, her bir hasta için *d*'nin hesaplanmasında birleşik standart sapmayı kullandım. Aynı zamanda *d*'nin standart sapmasını ve güven sınırlarını da hesapladım. Bu hesaplamalardaki formüller aşağıdaki gibidir:

$$d = \frac{\overline{X}_{\text{son}} - \overline{X}_{\text{ön}}}{s_{\text{birlesik}}}$$

$$s_{\text{d}} = \sqrt{\frac{1}{n_{\text{ön}}} + \frac{1}{n_{\text{son}}} + \frac{d^2}{2(n_{\text{ön}} + n_{\text{son}})}}$$

(d'yi hesaplamak için kontrol grubunun standart sapmasını kullanmış olsaydım, paydanın sağındaki değer $2(n_{\text{önceki}})$ olacaktı. Bakınız, Gleser ve Olkin (2009).)

*d*'nin güven aralığından, 1., 2. ve 4. hastanın performansının manidar bir şekilde arttığını görebilirken 3. hastanınki bu şekilde değildir – bu hasta için *d*'nin güven aralığı 0'ı içerir. Dikkat edilirse güven aralıklarının çoğu oldukça geniştir.

**Tablo 17.5** Müdahale öncesi ve sonrası depresyon puanları
($A_i$, Adım$_i$ müdahale öncesi ya da sonrasını; $Y_i$, i. durumdaki puan$_i$'yi gösterir.)

| $P_1$ | $Y_1$ | $P_2$ | $Y_2$ | $P_3$ | $Y_3$ | $P_4$ | $Y_4$ |
|---|---|---|---|---|---|---|---|
| A | 14 | A | 16 | A | 9 | A | 12 |
| A | 10 | A | 12 | A | 9 | A | 17 |
| A | 13 | A | 18 | A | 6 | A | 9 |
| A | 15 | A | 18 | A | 8 | A | 17 |
| A | 13 | A | 15 | A | 9 | A | 14 |
| A | 15 | A | 16 | A | 6 | A | 17 |
| B | 4 | A | 13 | A | 8 | A | 13 |
| B | 4 | A | 15 | A | 9 | A | 15 |
| B | 7 | B | 12 | A | 7 | A | 12 |
| B | 3 | B | 10 | A | 10 | A | 17 |
| B | 7 | B | 15 | B | 3 | A | 9 |
| B | 2 | B | 13 | B | 7 | A | 18 |
| B | 1 | B | 9 | B | 4 | B | 6 |
| B | 3 | B | 16 | B | 9 | B | 6 |
| B | 6 | B | 9 | B | 9 | B | 7 |
| B | 3 | B | 10 | B | 5 | B | 8 |
| B | 7 | B | 12 | B | 7 | B | 5 |
| B | 5 | B | 14 | B | 2 | B | 6 |
| B | 0 | B | 11 | B | 9 | B | 4 |
| B | 4 | B | 16 | B | 12 | B | 7 |
| B | 3 | B | 7 | B | 6 | B | 10 |
| B | 0 | B | 10 | B | 10 | B | 9 |
| B | 6 | B | 11 | B | 5 | B | 3 |
| B | 8 | B | 14 | B | 4 | B | 9 |
| Ortalama (A) | 13,333 | | 15,375 | | 8,100 | | 14,167 |
| SD (A) | 1,862 | | 2,134 | | 1,370 | | 3,186 |
| Ortalama( )B | 4,056 | | 11,812 | | 6,571 | | 6,667 |
| SD (B) | 2,437 | | 2,664 | | 2,928 | | 2,103 |
| $s_{ortak}$ | 2,319 | | 2,507 | | 2,415 | | 2,700 |
| $d$ | 4,001 | | 1,421 | | 0,633 | | 2,778 |
| $s_d$ | 1,248 | | 0,560 | | 0,438 | | 0,699 |
| $CI_{alt}$ | 1,556 | | 0,323 | | −0,225 | | 1,408 |
| $CI_{üst}$ | 6,447 | | 2,519 | | 1,490 | | 4,147 |
| *Ağırlık* | 0,642 | | 3,188 | | 5,223 | | 2,048 |

© Cengage Learning 2013

Şimdi, dört hasta üzerinden $d$'nin ortalamasını hesaplayalım. Bu durumda, meta-analizde incelediğimiz gibi, dört $d$ değerinin sadece ortalamasını almayız. Bunun yerine, daha küçük standart hataya sahip olan $d$'ler daha fazla ağırlıklı olacak şekilde ağırlıklandırılmış ortalamayı kullanacağız. Amacımız için

$$W_i = \frac{1}{s_{d_i}^2} = \frac{1}{1{,}248^2}; \frac{1}{0{,}560^2}; \frac{1}{0{,}438^2}; \frac{1}{0{,}669^2}$$

$$\overline{d} = \frac{\sum W_i d_i}{\sum W_i} = 1{,}450$$

$$s_{\overline{d}} = \sqrt{\frac{1}{\sum W_i}} = 0,300$$

$$GA(s_{\overline{d}}) = d \pm 1,96_{s_{\overline{d}}}$$

$$= 1,450 \pm 1,96 \times 0,300$$

$$0,861 \leq \delta \leq 2,038$$

Dikkat edilirse, birleştirilmiş veri kümesi için özet istatikte $\overline{d}$ = 1,450 ve güven aralığı 0,861 ≤ $\delta$ ≤ 2,038'dir. Diğer bir deyişle, depresyon puanlarının ortalama etki büyüklüğü, 1,5 standart sapma düşmüştür. Aralığın alt sınırı dahi bir standart sapmanın 8/10. değeri üzerindedir. Araştırmada kullanılan ilâcın, depresyonu dikkate değer ölçüde düşürdüğü yorumunu yapabiliriz.

Bu bölümün, ilk yarısı boyunca devam ediyor ve sabit modele karşı tesadüfî modelleri tartışıyor, $\tau$'nun kestirimini hesaplıyor olsak da burada bunları daha fazla yapmayacağız. Genellikle tek-durum desenleri oldukça az olaya sahiptir ve tipik bir meta-analizde uygulanan daha incelikli sonuçların çok azı elde edilir.

## Dikkat Edilmesi Gereken Noktalar

Bu veri için, meta-analizinin yapmak istediğimizi yapacağına ve bu tür bireysel durum çalışmaları için doğru bir şekilde sonuçları yansıtacağına eminim. Mevcut örneği de bu sebeple verdim. Bununla birlikte, eldeki örnek tamamen iyi bir çalışmayı yansıtıyorken diğer iyi çalışmalar bu tür bir analiz için uygun olmayabilir.

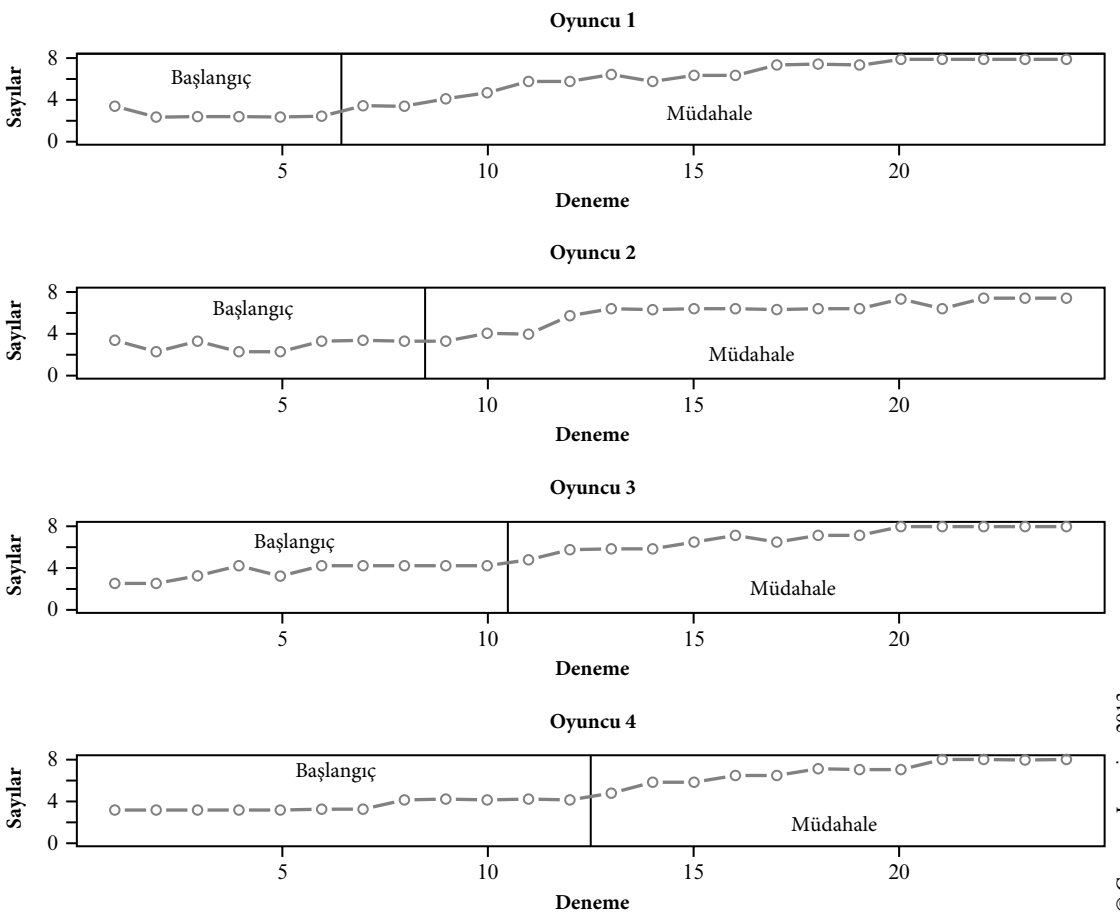
Herşeyden önce, çoğu kez oto-korelasyon problemi vardır. Bu çalışma için adımlar içindeki tepkileri birbirinden bağımsız varsaydık. Bu varsayım, burada mantıklı görünmektedir. Ancak pek çok A-B deseninde, bir bireyin deneme$_i$'deki performansı, bu bireyin$_{i-1}$'deki performansıyla ilişkilidir. Diğer bir deyişle, gözlemlere (adım-içi) ilişkin oto-korelasyonla ilgili endişe duymamız gerekir. Örneğin; eğer birey, adımlar boyunca genel bir artış veya azalış eğilimi gösteriyorsa oto-korelasyonun bulunması olasılığı yüksektir. Oto-korelasyon ile ilgilenmek bile tartışmalı bir konudur. Manolov ve Solanas (2008), bir simulasyon çalışması yapmışlar, kullandıkları analizin, oto-korelasyonu ele alan alternatif analizler kadar iyi olduğu sonucuna ulaşmışlardır. Huitema, birkaç çalışmasında [örneğin; Huitema ve McKean (1991)], oto-korelasyonun düşündüğümüz gibi bir problem olmadığını savunmuştur. Analizimizi gerçekleştirdikten sonra asıl endişe duymamız gereken artıkların (residuals) oto-korelasyonudur. Eğer analizimize, korelasyonları özümseyen veya korelasyonlar için açıklama getiren terimler ekleyebilirsek artıklar mâkul derecede bağımsız olacaktır. Örneğin; eğer oto-korelasyon, verideki genel bir artış eğiliminden dolayı ortaya çıktıysa ve denemeleri, analize bir değişken olarak dahil ederek, denemelerdeki oto-korelasyon düzeltilmiş olur. Bununla birlikte, vurgulamak gerekir ki bu işlem evrensel bir görüş değildir.

Düşünülmesi gereken ikinci bir konu, gerçekten tasarlanan ve ciddi bir şekilde ele alınan tek-denekli verinin alternatif yollarla incelenmesidir. Bir sonraki örnekte, Center, Skiba ve Casey (1985-1986) tarafından önerilen ve Huitema ve McKean (2000) tarafından geliştirilen regresyon yaklaşımını ele alacağız. Van den Noortgate ve Onghena (2003a ve b), hiyerarşik lineer modellerin kullanılmasını önermiştir, ancak bu yaklaşıma burada yer verilmeyecektir. Bu araştırmacılar, davranışların modellenmesine ilişkin çok iyi bir çalışma da yapmışlardır. Son olarak, Bulté ve Onghena (2008), oto-korelasyona çözüm getiren güçlü tekrar örnekleme süreçleri geliştirmişlerdir, ancak özellikle ortalama farkların manidarlık testine odaklanmışlardır.

## 17.6 Tek-Durum Desenine İkinci Bir Yaklaşım – Parça Parça Regresyon Kullanımı

Bir önceki örnekten oldukça farklı olan ve tek-durum desenlerinin analizlerindeki karmaşıklığı göstermek üzere Zeigler'dan (1994) bir örnek ele alacağız. Zeigler, Amerikan futbolu oyuncularının uygun çevresel işaretlere çabucak tepki vermelerine ilişkin antremanlarıyla ilgilenmiştir. Performansı düşük olan dört oyuncu seçmiş ve birkaç seri deneme boyunca bu oyuncuların top ile hedefi vurma becerilerini gözlemiştir. Her bir oyuncu için başlangıç deneme sayıları farklıdır. Daha sonra, dikkat toplama eğitimi (attentional shift-training) uygulamış ve performanslarındaki hatasızlığı sürekli olarak kaydetmiştir. Elde edilen veri Şekil 17.5'teki grafikle gösterilmiş ve bireysel puanlar Tablo 17.6'da verilmiştir.

Şekil 17.5'ten, uygulamanın etkiliymiş gibi olduğu görünüyor. Üç oyuncunun performansları, müdaheleden sonra artarken, üç numaralı oyuncu ile ilgili şüphe duyulabilir. Az bir dikkatle, bu oyuncunun performansının sürekli olarak az da olsa bir artış gösterdiğini ve bu artışın aynı şekilde uygulamadan sonra da devam ettiğini hissedebilirsiniz. Diğer bir deyişle, tüm veri boyunca bir regresyon doğrusu gayet uygun olur ve uygulama-sonrası performans ile başlangıçtaki performansın eğilimi bir süreklilik gösterir.




© Cengage Learning 2013

**Şekil 17.5** Zeigler'dan (1994) veri

**Şekil 17.6** Dördüncü oyuncu için futbol testindeki performans
($A_i$ deney aşamasını ve $Y_i$ oyuncu için bağımlı değişkeni göstermektedir.)

| $A_1$ | $Y_1$ | $A_2$ | $Y_2$ | $A_3$ | $Y_3$ | $A_4$ | $Y_4$ |
|---|---|---|---|---|---|---|---|
| A | 4 | A | 4 | A | 3 | A | 4 |
| A | 3 | A | 3 | A | 3 | A | 4 |
| A | 3 | A | 4 | A | 4 | A | 4 |
| A | 3 | A | 3 | A | 5 | A | 4 |
| A | 3 | A | 3 | A | 4 | A | 4 |
| A | 3 | A | 4 | A | 5 | A | 4 |
| B | 4 | A | 4 | A | 5 | A | 4 |
| B | 4 | A | 4 | A | 5 | A | 5 |
| B | 5 | B | 4 | A | 5 | A | 5 |
| B | 6 | B | 5 | A | 5 | A | 5 |
| B | 7 | B | 5 | B | 6 | A | 5 |
| B | 7 | B | 7 | B | 7 | A | 5 |
| B | 8 | B | 8 | B | 7 | B | 6 |
| B | 7 | B | 8 | B | 7 | B | 7 |
| B | 8 | B | 8 | B | 8 | B | 7 |
| B | 8 | B | 8 | B | 9 | B | 8 |
| B | 9 | B | 8 | B | 8 | B | 8 |
| B | 9 | B | 8 | B | 9 | B | 9 |
| B | 9 | B | 8 | B | 9 | B | 9 |
| B | 10 | B | 9 | B | 10 | B | 9 |
| B | 10 | B | 8 | B | 10 | B | 10 |
| B | 10 | B | 9 | B | 10 | B | 10 |
| B | 10 | B | 9 | B | 10 | B | 10 |
| B | 10 | B | 9 | B | 10 | B | 10 |

© Cengage Learning 2013

Daha önce belirttiğim gibi, performansın ölçülmesinde kullanacağımız istatistiksel seçeneklerimiz vardır. Önceki örnekte, standartlaştırılmış ortalama farkı kullanmıştık. Bununla birlikte, savunulan farklı istatistikler de bulunur. Veride süreklilik gösteren bir eğilim varsa ve bu eğilim sabit tutularak geri kalanlarla ilgileniliyorsa regresyon yaklaşımını önerenler vardır. Bunlar aynı zamanda, örtüşmeyen verinin yüzdesinin ölçülmesini de önerirler. Örtüşmeyen veri, müdahaleden önceki denemelerdeki en yüksek (veya en düşük) gözlemin üzerindeki (veya altındaki) müdahele boyunca yapılan denemelerin sayısının belirlenmesi demektir. Diğer bazı araştırmacılar da, müdahele sırasında, başlangıçtaki ortanca sayısını geçene kadar olan gözlemlerin sayılmasını önerirler. Maalesef, bu ölçümler genelde birbirleriyle çok ilişkili değildir (Owens, Farmer, Ferron ve Allsopp, 2010).

Center, Skiba ve Casey (1985-1986), Huitema ve McKean (1998 ve 2000) yaptıkları çalışmalarda, bir adım daha ileri gidilmesi gerektiğini önerirler. Çalışmaları, basitçe standartlaştırılmış ortalama farka bakılmasından daha fazla çalışmayı içerir; ancak daha fazla bilgi verir ve genelde elimizdeki veri kümesiyle daha çok ilişkilidir. Bu yaklaşım, **parçalı regresyon** yaklaşımı olarak bilinir ve zaman, aşama ve bunların etkileşiminin kullanıldığı uygun regresyon doğruları ile çalışılır. (Center ve diğerleri (1985-1986) ile Hutiema ve McKean'in (2000) önerileri arasında bir miktar farklılık vardır.)

parçalı regresyon

Bir an için dördüncü oyuncuya ait ölçümlerle ilgilenelim. Bu oyuncuyu, ölçümlerinin eğimi açık bir değişim gösterdiğinden seçiyoruz. Denemeleri 1'den 24'e kadar olan sayılarla kodlayalım. Diğer bir değişkenimizi de Aşama olarak adlandıralım ve müdahale öncesi aşama için 0, müdahale sonrası aşama için 1 diyelim. Aşama sayısı ile [Deneme sayısı (*nb*-1)]'i çarparız; burada *nb*, müdahale öncesi deneme sayısını gösterir. Bu, Zaman

**Tablo 17.7** Zeigler çalışması için dördünci duruma ait veri

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Y* | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 6 | 7 | 7 |
| *A* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *D* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| *AD* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Y* | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| *A* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *D* | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| *AD* | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |

© Cengage Learning 2013

ve Aşamanın etkileşimini kodlar. Bu son kodlamanın etkisiyle tüm müdahele öncesi denemeler 0 ve müdahaledeki denemeler 0, 1, 2, 3,... olarak kodlanır. Dördüncü durum için veri, Tablo 17.7'de gösterilmiştir. Bu durumun gösterimi tercih edilmiştir, çünkü hem eğimdeki hem de başlangıçtan müdaheleye kadar olan düzeylerdeki farklar çok açık olarak görülür.

Sadece göstermek amacıyla, dördüncü durumun müdahale öncesi aşama ölçümleri için $\text{Deneme}_i$'deki $Y_i$'nin regresyonunu inceleyeceğiz. Aynı süreci, müdahele sonrası aşama için de yapacağız. Son olarak, her bir eşitliği, müdahale sonrası ilk deneme olan $\text{Deneme}_{13}$'teki puanın kestirimi için kullanacağız. Sonuçlar aşağıda verilmiştir:

| Müdahale öncesi verinin regresyonu | Müdahaledeki verinin regresyonu |
|---|---|
| $\hat{Y}_i = 3{,}6212 + {,}1224 \times \text{Deneme}_i$ | $\hat{Y}_i = 1{,}7914 + {,}3671 \times \text{Deneme}_i$ |
| $\text{Deneme}_{13}$ için kestirim | |
| $\hat{Y}_i = 3{,}6212 + {,}1224 \times 13 = 5{,}2124$ | $\hat{Y}_i = 1{,}7914 + {,}3671 \times 13 = 6{,}5637$ |

$\text{Deneme}_{13}$ için kestirilen değerin; müdahale etkisi ile 5,2124'ten 6,5637'ye değiştiğini, 1,3513 değerinde bir fark olduğunu görebiliyoruz.

Burada, neye sahip olduğumuzu -ya da daha önemlisi neye sahip olmadığımızı- farketmek önemlidir. Parçalı regresyon yaklaşımını kullanarak, "düzey"deki bir değişim, ilk müdahaleye ait denemedeki kestirilen puanda görülen değişimdir. Değişim için başlangıçtan müdahaleye kadar olan süredeki ortalama değişim hesaplanmaz. Eğer Şekil 17.5'teki üçüncü oyuncuya bakarsak ve hayalî regresyon doğruları çizersek, kestirimin müdahale öncesi ilk denemeye mi, müdahale aşamasındaki veriye mi dayalı olduğundan şüphe duyabiliriz, çünkü kestirimler büyük ölçüde aynı olacaktır. Gerçekte, iki aşama için ortalamalar oldukça farklı görünür. (Kestirimler, yaklaşık 0,85 farklıdır ve bu farklılık manidar değildir. Diğer taraftan, ortalamalar arası farklılık 4,17'dir.)

Sonuçta, $\text{Deneme}_i$'deki $Y_i$'yi kestirimi için üç kestirici (Deneme, Aşama ve DA) kullanabiliriz. Bu sonuç 4 numaralı durum için aşağıda gösterilmiştir.

$\text{Model} = Y = b_0 + b_1(\text{Deneme}) + b_2(\text{Aşama}) + b_3(\text{D}\times\text{A})$

| | $b_i$ | $s_b$ | $t$ | $p$ | $r_{sp}$ |
|---|---|---|---|---|---|
| Kesim noktası | 3,6212 | 0,2140 | 16,924 | ,000* | |
| Aşama | 1,3519 | 0,2853 | 4,738 | ,000* | ,146 |
| Deneme | 0,1224 | 0,0291 | 4,209 | ,000* | ,129 |
| A X D | 0,2448 | 0,0411 | 5,953 | ,000* | ,183 |

© Cengage Learning 2013

Bu sonucun regresyon katsayılarına bakarsak, bazı ilginç değerler görürüz. Müdahale öncesi veri üzerinden hesaplanan tam model için kesim noktası 3,6212'dir. Denemeler için katsayı (0,1224) başlangıç verideki denemeler için regresyon katsayısıdır. Daha ilginç olan, $\text{Deneme}_{13}$ için iki kestirim arasındaki fark başlangıç ve müdahaledeki veriye dayanmaktadır ve 1,3519'dur, tam modeldeki Aşama için regresyon katsayısıdır. Eğer iki eğim (müdahale öncesi ve sonrası veri için ayrı ayrı) arasındaki farka bakarsak, bu fark 0,3671 – 0,1224 = ,2448'dir ve tam modeldeki etkileşim için regresyon katsayısıdır. Diğer bir deyişle, veri, buradaki gibi, ki ben öyle yaptım, yapılandırıldığında; tam model, düzeydeki değişim, eğimdeki değişim ve her ikisinden hangisini istiyorsak onun için katsayıları bize verir. Ancak, dikkatli olalım: Düzeydeki değişim, ilk müdahale denemesinde kestirilen performansı ifade eder; iki aşama arasındaki ortalama fark demek değildir. Buna ek olarak Deneme için etki, olasılıkla çok ilgilenilmeyen, sadece müdahale öncesi verinin eğimidir. Tek-durum çalışmaları ile ilgilenildiğinde, genelde aşama-içi (within-phase) varyans, özellikle de müdahale öncesi için, küçüktür ve böylece manidarlık testleri ve etki büyüklüklerinin denekler-arası desenlerde normalde beklenenden daha büyük bulunması muhtemeldir.

Her defasında bir kestiricinin modelden çıkarıldığı bu sistemin avantajı, her bir yordayıcıyının etkisini anlamak için tüm modele ait *t* testine bakmanın yeterli olmasıdır. Buradaki *t* testi katsayılar üzerinden yapılan *t* testidir. Dördüncü oyuncu için, Denemelerde olduğu kadar, düzey değişiminde ve eğim değişiminde manidar farklılık vardır.

Gerçekten istediğimiz, etki büyüklüğü ölçüsüne dönmek için tüm durumlar üzerine veriyi birleştirebilmemizdir. Bu dönüş tahmin etiğimizden çok daha kolay olacaktır. Center ve ark.nın (1985-1986) belirttiği gibi, her bir katsayıdaki test için aşağıdaki eşitlik yardımıyla *t* değerlerinden *d*'ye dönüşüm yapabiliriz:

$$d = \frac{2|t|}{\sqrt{sd_{\text{hata}}}}$$

(Dikkat edilirse, bu hesaplama için *t*'nin mutlak değerini kullanıyoruz.) Normal olarak, standartlaştırılmış ortalama farklarla ilişkilendirdiğimiz bir ölçüyü özellikle regresyon modeliyle birlikte hesaplamak garip görünebilir. Center ve ark. "regresyon yaklaşımından türetilen etki büyüklüklerini, deneyin bir fonksiyonu olarak ortaya çıkan varyans miktarının, kestirimin standart hatasının birimi olarak açıklanan göstergesi olarak yorumlamayı" (s. 398) tartışmışlardır. Huitema'nın da belirttiği gibi, önemli bir nokta da; tek-durum çalışmalarındaki etki büyüklüklerinin grup çalışmalarındaki etki büyüklükleri kadar açık olmadığıdır. Gerçekten, göreceğimiz gibi, etki büyüklüklerinin önemli ölçüde olduğundan en azından bir miktar büyük hesaplanması muhtemeldir, çünkü standart hatalar, grup çalışmalarında beklediğimizden çok daha küçüktür.

Her bir oyuncu için *t*'yi *d*'ye dönüştürmeye devam edersek, aşağıdaki sonuçları elde ederiz:

| Oyuncu | $d_{\text{Aşama}}$ | $d_{\text{Deneme}}$ | $d_{\text{AxD}}$ |
|---|---|---|---|
| 1 | 1,88* | 0,46 | 1,61* |
| 2 | 0,97* | 0,22 | 0,76 |
| 3 | 0,86 | 1,94* | 0,62 |
| 4 | 1,74* | 1,88* | 2,66* |

* Tam modeldeki manidar *t*.

© Cengage Learning 2013

Dikkat edilirse, sonuçlarda dikkate değer bir değişkenlik vardır. Tüm dört durum, en az yaklaşık bir standart sapma düzey değişikliği gösterir. (Bu değişikliklerden üçü manidardır.) İki durumda, eğimde manidar değişiklik görülür, aynı zamanda güçlü etkilerdir. (Daha önce de sözü edildiği gibi, Durum 3 için veri, müdahale öncesinden başlayarak, genel olarak sürekli bir performans artışı gösterir. Bu sonuç da bu tavsiyeyi destekler.) Fakat dikkat edilirse bu

**Tablo 17.8** Etki büyüklükleri ve *d*'nin ortalama ve standart sapmasının hesaplanması

| Durum | $d_{Aşama}$ | $s_{bAşama}$ | $W_{Aşama}$ | $d_{zaman}$ | $s_{bzaman}$ | $W_{zaman}$ | $d_{etk.}$ | $s_{betk.}$ | $W_{etk.}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,88* | ,60 | 2,78 | 0,46 | ,14 | 51,02 | 1,61* | ,14 | 51,02 |
| 2 | 0,97* | ,71 | 1,98 | 0,22 | ,12 | 69,44 | 0,76 | ,13 | 59,17 |
| 3 | 0,86 | ,41 | 5,95 | 1,94* | ,05 | 400,00 | 0,62 | ,06 | 277,78 |
| 4 | 1,74* | ,29 | 11,89 | 1,88* | ,03 | 1111,11 | 2,66* | ,04 | 625,00 |
| $\bar{d}$ | 1,46 | | | 1,78 | | | 1,94 | | |
| $s_{\bar{d}}$ | 0,044 | | | 0,001 | | | 0,001 | | |

© Cengage Learning 2013

paragrafta, değişimleri bireysel olarak incelemekteyim. Meta-analizinde, durumlar (veya çalışmalar) üzerinden inceleme yapılır. Varsayalım ki genel gücü ,70 olan bir çalışmamız var. Bu demektir ki farklılığın %70'inin manidar olması, %30'unun olmaması beklenir. Bu durumda, yokluk hipotezini "ispatlar"mış gibi manidar olmayan kısma odaklanmak mantıklı olmaz. Her ne zaman güç 1,00'den küçükse bir miktar manidar olmayan farklılıklar beklenir. Böylece, doğru yöndeki bir manidar olmayan farklılık, deneysel hipotezi destekler şekilde ve sonuçların birleştirilmesini gerektiren mâkul bir test olarak görülebilir.

Her bir durum için *d* değerlerini hesapladıktan sonra, daha önce yaptığımız gibi, tüm durumlar üzerinden bunları birleştiririz. Bu durumda, birleştirilmiş durumlar için ağırlıklandırma seçimine dair bir soru ortaya çıkar. Daha önce, kestirimin kesinliğine göre ve böylece bir metrik olarak, denekler-içi varyansların çarpmaya göre terslerini kullanmıştık. Bu durumda, yarı-kısmî regresyon katsayılarını ve etki büyüklükleri için bunların *t* değerlerini kullanırız ve böylece ağırlıklandırma olarak regresyon katsayılarının standart hatalarının karelerinin çarpmaya göre terslerini kullanmak mantıklıdır. Bununla birlikte, bunun etkisi, *d* için, şüpheli görülen, oldukça küçük standart hataların ortaya çıkması olacaktır.

Dört durum üzerinden meta-analiz için yapılan hesaplamalar aşağıda gösterilmiştir. Ortalama *d*'nin hesaplanması için önce aşağıda görüldüğü gibi tanımlanan ağırlıklandırmalara ihtiyacımız olacak:

$$W_i = \frac{1}{s_d^2}$$

İlk duruma ait aşamalardaki düzey farklılığı için

$$W_{devre} = \frac{1}{,60^2} = 2,78$$

elde edilir.

Benzer hesaplamalar, geri kalan değerleri, Tablo 17.8'de görülen, yarı-kısmi regresyon katsayılarının standart hataları ile birlikte verir.

Buna ek olarak, her bir sonuç için, *d* değerlerinin ağırlıklandırılmış toplamlarının ağırlıkların toplamlarına bölümü olan *d*'nin ortalama etkisini tanımlarız. Aşama etkisi için;

$$\bar{d} = \frac{\sum W_i d_i}{\sum W_i} = \frac{2,78 \times 1,88 + 1,98 \times 0,97 + 5,95 \times 0,86 + 11,89 \times 1,74}{2,78 + 1,98 + 5,95 + 11,89}$$

$$= \frac{32,9526}{22,62} = 1,46$$

$$s_{\bar{d}} = \sqrt{\frac{1}{\sum W_i}} = \sqrt{\frac{1}{22,62}} = ,044$$

bulunur. Geri kalan hesaplamalar, Tablo 17.8'in altında gösterilmiştir. Dikkat edilirse, standart hataların küçüklüğü, tek-durum desenlerinde elde ettiğimiz kesinliği yansıtır.

## Anahtar Terimler

| | | |
|---|---|---|
| Meta-analiz (Giriş) | Orman grafiği (17.2) | Parçalı regresyon (17.5) |
| Çekmeceye kaldırılmış dosya problemi (Giriş) | A-B Deseni (17.4) | |

## Alıştırmalar

Mazucchelli, Kane ve Rees (2010), subjektif mutluluk duygularını geliştirme yolu olarak davranışsal etkinlik çalışmalarını incelemişlerdir. (Araştırmacılar, davranışsal etkinlikleri; "amaçlı davranışsal, bilişsel veya bilinçli davranışlar ... kazanılması istenilen pozitif duygular, davranışlar veya bilişler" olarak tanımlamışlardır.) Deney ve kontrol gruplarının karşılaştırıldığı 11 çalışmanın veri kümelerini raporlaştırmışlardır. Minimal ve Yüksek düzeyde depresyon semptomu gösterenler olarak bölünmüş olan bu veri aşağıda gösterilmiştir.

| Yazar | Altgrup | $n_E$ (kest.) | $n_C$ (kest.) | Hedges g | Standart hata |
|---|---|---|---|---|---|
| Barlow86a | Yüksek | 12 | 12 | -,134 | 0,523 |
| Besyner79 | Yüksek | 14 | 16 | 0,675 | 0,423 |
| Lovett88 | Yüksek | 33 | 27 | 0,204 | 0305 |
| Stark | Yüksek | 10 | 9 | 0,043 | 0,439 |
| Van den Hout | Yüksek | 15 | 14 | 0,644 | 0,371 |
| Weinberg | Yüksek | 10 | 9 | 0,976 | 0,467 |
| Wilson | Yüksek | 9 | 11 | 1,466 | 0,524 |
| Barlow86b | Minimal | 12 | 13 | 0,133 | 0,352 |
| Fordyce77 | Minimal | 50 | 60 | 0,609 | 0,195 |
| Fordyce83 | Minimal | 40 | 13 | 1,410 | 0,483 |
| Reich81 | Minimal | 49 | 49 | 0,378 | 0,179 |

© Cengage Learning 2013

17.1 Bu veri kümesi için orman grafiğini oluşturunuz.

17.2 Tesadüfî model olduğunu varsayarak ortalama etki büyüklüğünü ve standart sapmasını hesaplayınız.

17.3 Önceki alıştırmada hesapladığınız ortalama etki büyüklüğünün güven sınırlarını hesaplayınız.

17.4 Bu tablodaki iki alt grup arasında fark olmadığı hipotezini test ediniz.

17.5 Modelin sabit olduğunu varsayarak, ortalama etki büyüklüğünü ve standart sapmasını tekrar hesaplayınız.

---

Bloch ve ark. (2009), dikkat-eksikliği/hiperaktivite bozukluğunun (ADHD) iyileştirilmesi üzerine bir meta-analiz çalışması gerçekleştirmişlerdir. Tourette sendromu geçmişine sahip ADHD'li çocuklara verilen psiko-uyarıcı meditasyonun, tiklerin (istemsiz hareket) şiddetini arttırabildiğini rapor etmişlerdir. Araştırmacılar, tik ve ADHD şiddeti için sıralama ölçeği ile methylphenidate türevine karşı plasebo-kontrol durumlarını karşılaştıran dört çalışma (diğer çalışmalar arasından) bulmuşlardır. Bu dört çalışmadaki tiklerin varlığı ve sıklığına ilişkin sonuçlar aşağıda gösterilmiştir. (*d*'nin standart hataları, yayımlanmış sonuçlardan kestirilmiştir.)

| Çalışma | *n* | *d* | *Güven Aralığı* | | $s_d$ |
|---|---|---|---|---|---|
| Gadow92 | 11 | 0,11 | -0,55 | -0,77 | 0,336 |
| Castellanos | 20 | 0,29 | -0,18 | -0,76 | 0,240 |
| TSSG | 103 | 0,64 | 0,27 | -1,00 | 0,184 |
| Gadow07 | 71 | 0,06 | -0,17 | -0,29 | 0,117 |

© Cengage Learning 2013

17.6 Bu tablodaki veri için kestirilen ortalama etki büyüklüğünü ve güven aralığını hesaplayınız.

17.7 Bu meta-analizini tesadüfî model olarak ele almak niçin mantıklı olacaktır?

17.8 Bloch'un çalışmasındaki veri için orman grafiğini oluşturunuz.

17.9 Bloch ve diğerlerinin çalışması için methylphenidate kullanımının tik davranışını arttırma riski ile ilgili olarak nasıl bir sonuca varırsınız?

17.10 Önceki alıştırmalarda kullanılan üç çalışmadaki ADHD şiddeti ölçümlerine de bakılmıştır. Buna ilişkin sonuçlar aşağıda gösterilmiştir.

| Çalışma | *n* | *d* | *Güven Aralığı* | $s_d$ |
|---|---|---|---|---|
| Gadow92 | 11 | 1,11 | 0,44 – 1,77 | 0,341 |
| TSSG | 103 | 0,56 | 0,19 – 0,92 | 0,189 |
| Gadow07 | 71 | 0,76 | 0,53 – 0,99 | 0,117 |

© Cengage Learning 2013

a. Bu çalimalara ait ortalama büyüklük etkisini hesaplayınız.

b. Ortalama büyüklük etkisine ait güven aralığını bulunuz.

17.11 Bu çalışmalara ait orman grafiğini oluşturunuz.

17.12 Bu örnekteki etki büyüklüklerinin heterojenliğine bakmak niçin çok az anlam taşır?

---

Bauer ve Dópfmer (1999), geleneksel anti-depresanların artışında lityumun etkisini incelemişlerdir. Diğer anti-depresanlara tepki vermeyenler için lityumun test edildiği, plasebo ve kontrol gruplu 9 çalışma bulmuşlardır. Bu çalışmalardan kestirilen veri aşağıda verilmiştir.

| Çalışma | *n* | *d* | $s_d$ | *Güven Aralığı* | |
|---|---|---|---|---|---|
| Stein | 34 | -0,30 | 0,82 | -1,10 | -0,50 |
| Zusky | 16 | 0,19 | 0,86 | -0,65 | -1,03 |
| Katona | 61 | 0,50 | 0,51 | 0,02 | -1,02 |
| Schopf | 27 | 1,36 | 1,39 | 0,05 | -2,77 |
| Baumann | 24 | 0,97 | 0,99 | 0,09 | -1,85 |
| Browne | 1 | 0,51 | 0,87 | -0,34 | -1,36 |
| Kantor | 7 | 0,51 | 1,55 | -1,01 | -2,03 |
| Heninger | 15 | 1,33 | 1,38 | -0,02 | -2,68 |
| Joffe | 33 | 0,76 | 0,70 | -0,02 | -1,36 |

© Cengage Learning 2013

17.13 Bu sonuçlar için orman grafiğini oluşturunuz.

17.14 Ortalama etki büyüklüğünü ve güven sınırlarını hesaplayınız.

---

Kapoor, Rajkumar ve diğerleri (2011), kan kanserinin (miyelom) tedavisiyle ilgili veri toplamışlardır. (Bu hastalığın, psikoloji ile çok az ilgili olduğunu düşünebilirsiniz, ancak ben kan kanseriyim ve sonuçlara özel bir ilgim var.) Kapoor, Rajkumar ve diğerleri, bu hastalık için standart kemoterapi yoluyla tedavi ile aynı tedaviye thalidomide'in dahil olduğu tedavinin karşılaştırıldığı, 2007'den beri yapılmış olan dört çalışma bulmuşlardır. (Thalidomide, 1950'lerde hamilelerde sakinleştirici bir ilâç olarak kullanılmış ve sonrasında ciddî doğum özürlerine sebep olduğu bulunmuş, piyasadan çekilmiştir. Fakat bu süre içinde dünya çapında 10.000-20.000 özürlü doğuma da neden olmuştur. Son zamanlarda, bu ilâcın bazı kanser türleri için çok iyi bir tedavi olduğu bulunmuştur.) Aşağıdaki tabloda, dört çalışmanın sonuçları verilmiştir.

| Çalışma | Başarılı Thalidomide | Toplam Thalidomide | Başarılı Kontrol | Toplam Kontrol |
|---|---|---|---|---|
| Palumbo | 21 | 129 | 5 | 126 |
| Facon | 16 | 125 | 4 | 198 |
| Hulin | 8 | 113 | 1 | 116 |
| Hovon | 3 | 165 | 1 | 108 |

© Cengage Learning 2013

17.15 Her bir çalışma için log risk oranlarını ve risk oranlarını hesaplayınız.

17.16 Dört çalışma üzerinden ağırlıklandırılmış ortalama risk oranını hesaplayınız.

17.17 Ortalama risk oranının güven sınırlarını hesaplayınız ve uygun yorumu yapınız.

17.18 Thalidomide'in geçmişteki kötü sonuçlarını temize çıkardığı yorumunu yapabilir miyiz?

17.19 Niçin etkilerin heterojenliği için sonuçları incelemek çok anlamlı olmayabilir?

---

Bisson ve Andrew (2007), post-travmatik stres bozukluğu (PTSD) için tedavi olarak bilişsel davranış terapisine (CBT) meta-analiz uygulamışlardır. CBT odaklı ve genel tedavi uygulamanın yapıldığı PTSD semptomlarının, klinik tedavi uzmanları tarafından puanlandığı 14 çalışma bulmuşlardır. Sonuçlar, aşağıda verilmiştir.

| | DBT | | | Kontrol | | |
|---|---|---|---|---|---|---|
| Çalışma | N | Ort. | Std. Sap. | N | Ort. | Std. Sap. |
| Kubany 1 | 45 | 15,80 | 14,40 | 40 | 71,90 | 23,80 |
| Foa1 | 45 | 12,60 | 8,37 | 15 | 26,93 | 8,47 |
| Kubany2 | 18 | 10,10 | 19,30 | 14 | 76,10 | 25,20 |
| Resick | 81 | 23,00 | 19,92 | 40 | 69,73 | 19,19 |
| Cloitre | 22 | 31,00 | 25,20 | 24 | 62,00 | 22,70 |
| Foa2 | 10 | 1540 | 11,09 | 10 | 19,50 | 7,18 |
| Keane | 11 | 28,80 | 10,05 | 13 | 31,90 | 31,90 |
| Ehlers | 14 | 21,58 | 28,56 | 14 | 74,55 | 19,12 |
| Vaughan | 13 | 23,00 | 10,20 | 17 | 28,50 | 8,90 |
| Brom | 27 | 56,20 | 24,10 | 23 | 66,40 | 24,30 |
| Blanchard | 27 | 23,70 | 26,20 | 24 | 54,00 | 25,90 |
| Fecteau | 10 | 37,50 | 30,40 | 10 | 74,60 | 24,70 |
| Gersons | 22 | 3,00 | 10,00 | 20 | 9,00 | 13,00 |
| Rothbaum | 20 | 21,25 | 22,50 | 20 | 64,55 | 19,87 |

© Cengage Learning 2013

17.20 Bu veriye meta-analizi uygulayınız ve uygun olan yorumu yapınız.

---

Sosyal hikâyeler, otizm spektrumlu çocuklara, istenmeyen davranışların azaltılarak uygun sosyal becerilerin öğrenilmesine yardım eden eğitim olarak kullanılır. Eckelberry (2007), dikkat ve hiperaktivite bozukluğu olan üç çocuğa yardımcı olmak üzere sosyal hikâyelerin kullanılmasını incelemiştir. Araştırmacı, müdahale öncesinde bir hafta, müdahale sürecinde iki hafta gözlemlerde bulunarak A-B deseni kullanmıştır. Çalışmanın bağımlı değişkeni, her gün kaydedilen yıkıcı davranışların sayısıdır. Sonuçlar, aşağıdaki tabloda görülmektedir.

| Öğrenci | Müdahale öncesi | Müdahale süreci |
|---|---|---|
| A | 13 12 12 15 16 | 14 12 08 10 06 03 03 02 00 00 |
| B | 10 12 13 08 12 | 11 10 07 06 02 05 00 06 00 03 |
| C | 22 20 25 20 25 | 20 18 22 20 17 24 19 22 20 20 |

© Cengage Learning 2013

17.21 Şekil 17.4'tekine benzer şekilde, yukarıdaki veriye uygun grafiği çiziniz.

17.22 Her öğrenci için standartlaştırılmış ortalama fark olarak *d*'yi hesaplayınız ve standart hatasını bulunuz.

17.23 Özet istatistikleri (ortalama (*d*)) ve güven aralığı hesaplayınız.

17.24 Çalışmadan çıkan sonuçları yorumlayınız.

---

Howard ve Kendall (1996), kaygı-rahatsızlığı yaşayan altı çocuk ile bilişsel-davranış aile terapisini test etmişlerdir. Her bir çocuktan, terapi başlamadan önce 2, 4 veya 6 ölçüm alınmıştır. Takip eden birkaç hafta boyunca 18 aile terapi seansı düzenlenmiştir. Diğer yapılanların yanında, çocuklar her hafta kendi başa çıkma becerilerini puanlamışlardır. Sonuçlar, grafiksel gösterimden uyarlanarak aşağıda verilmiştir. Yanında "*" işareti olan değerler, müdahale öncesi ölçümleri belirtmektedir. (Veriye, Ex17-25.dat. web adresinden ulaşılabilir.)

| Hafta | Çocuk 1 | Çocuk 2 | Çocuk 3 | Çocuk 4 | Çocuk 5 | Çocuk 6 |
|---|---|---|---|---|---|---|
| 1 | 2,0* | 6,0* | 6,3* | 4,0* | 4,4* | 2,3* |
| 2 | 2,5* | 6,7* | 3,4* | 3,9* | 3,8* | 2,3* |
| 3 | 2,5* | 6,7* | 3,2* | 4,0 | 3,4* | 2,0* |
| 4 | 2,8* | 6,3* | 3,4* | 2,6 | 4,0* | 1,9* |
| 5 | 2,3* | 6,7* | 3,2* | 3,8 | 4,5 | 2,6 |
| 6 | 2,3* | 6,9* | 2,8* | 3,9 | 4,9 | 3,1 |
| 7 | 2,6 | 6,2 | 7,1 | 2,3 | 3,9 | 3,0 |
| 8 | 2,2 | 6,6 | 2,9 | 4,0 | 4,0 | 3,3 |
| 9 | 1,7 | 6,6 | 3,2 | 3,2 | 4,3 | 4,2 |
| 10 | 3,0 | 6,9 | 3,4 | 4,0 | 4,2 | 3,3 |
| 11 | 3,0 | 6,9 | 2,9 | 3,7 | 4,2 | 5,4 |
| 12 | 5,1 | 6,8 | 3,8 | 7,1 | 4,2 | 4,9 |
| 13 | 3,0 | 6,3 | 2,6 | 4,0 | 4,2 | 6,2 |
| 14 | 3,6 | 6,9 | 3,4 | 6,9 | 4,2 | 6,7 |
| 15 | 2,7 | 6,9 | 3,2 | 7,0 | 4,3 | 6,2 |
| 16 | 2,7 | 6,9 | 7,1 | 7,0 | 4,0 | 6,7 |
| 17 | 4,0 | 6,9 | 4,1 | 7,0 | 4,7 | 6,7 |
| 18 | 3,0 | 6,9 | 3,0 | 7,1 | 4,9 | 6,5 |
| 19 | 4,9 | 6,9 | 3,3 | 7,1 | 4,9 | 6,8 |
| 20 | 4,0 | 6,9 | 3,6 | 7,1 | 4,9 | 6,8 |
| 21 | 6,0 | 6,9 | 3,7 | 7,1 | 4,8 | 6,8 |
| 22 | 5,4 | 6,9 | 6,7 | 7,1 | 4,9 | 6,2 |
| 23 | 5,1 | 6,9 | 3,3 | 6,3 | 4,8 | 6,6 |
| 24 | 5,3 | 6,1 | 4,7 | 5,9 | 5,1 | 6,6 |
| 25 | 5,3 | 6,9 | 4,8 | 6,5 | 5,7 | 6,9 |
| 26 | 5,0 | 6,9 | 4,1 | 6,9 | 5,3 | 6,9 |
| 27 | 5,6 | 7,0 | 4,1 | 6,6 | 5,9 | 6,9 |
| 28 | 5,6 | 6,9 | 4,8 | 5,6 | 5,1 | 6,2 |
| 29 | 5,3 | 6,9 | 4,1 | 6,6 | 5,7 | 6,2 |
| 30 | 5,6 | 7,0 | 5,2 | 5,9 | 53 | 6,2 |
| 31 | 5,1 | 7,0 | 4,7 | 5,9 | 5,4 | 5,9 |
| 32 | 4,9 | 6,9 | 4,7 | 6,6 | 5,4 | 6,7 |
| 33 | 6,9 | 6,9 | 4,7 | 5,5 | 5,4 | 5,8 |
| 34 | 5,3 | 6,9 | 5,0 | 5,5 | 5,4 | 6,4 |
| 35 | 5,0 | 7,0 | 5,1 | 5,6 | 5,5 | 6,3 |
| 36 | 4,1 | 7,0 | 5,0 | 5,7 | 5,5 | 5,6 |

17.25 Her çocuğa ait veri için ayrı ayrı grafik çiziniz ve aşamaları gösteriniz.

17.26 Altı grafik için uygun olan sonuç (outcome) ölçümlerine karar veriniz.

17.27 Her çocuk için standartlaştırılmış ortalama farkları hesaplayınız.

17.28 Veri için genel ortalama etkiyi hesaplayınız ve güven sınırlarını bulunuz.

---

Dunlap ve Fox (1999), otizmli altı çocuk için davranışsal destek programının etkilerini göstermişlerdir. Bağımlı değişken, her oturumdaki problemli davranışların süresinin oranıdır. Bu çocuklardan üçüne ait veri aşağıda verilmiştir. Yanında "*" işareti olan değerler, müdahale öncesi ölçümleri belirtmektedir. (Veriye, Ex17-29.dat. web adresinden ulaşılabilir.)

| Oturum | Çocuk 1 | Çocuk 2 | Çocuk 3 |
|---|---|---|---|
| 1 | 50* | 45* | 35* |
| 2 | 50* | 50* | 30* |
| 3 | 50* | 30* | 24* |
| 4 | 37 | 23 | 61* |
| 5 | 61 | 24 | 42* |
| 6 | 16 | 27 | 38* |
| 7 | 1 | 14 | 8 |
| 8 | 8 | 14 | 19 |
| 9 | 0 | 18 | 9 |
| 10 | 11 | 23 | 6 |
| 11 | 5 | 20 | 14 |
| 12 | 21 | 10 | 21 |
| 13 | 6 | 13 | 1 |
| 14 | 36 | 2 | 8 |
| 15 | 1 | 1 | 0 |
| 16 | 0 | 0 | 2 |
| 17 | 3 | 2 | 8 |
| 18 | 0 | | 6 |
| 19 | 16 | | 62 |
| 20 | 0 | | 19 |
| 21 | 9 | | 1 |
| 22 | 0 | | 28 |
| 23 | | | 6 |
| 24 | | | 2 |
| 25 | | | 0 |
| 26 | | | 6 |
| 27 | | | 4 |
| 28 | | | 1 |
| 29 | | | 0 |

©Cengage Learning

17. 29 Her çocuk için sonuçların ayrı ayrı grafiğini çiziniz, müdahale öncesi/sonrası geçişi gösteriniz.

17.30 Her çocuk için veriye uygun parçalı regresyon modelini oluşturunuz ve modeli yorumlayınız.

17.31 Çocuklar üzerinden parçalı regresyon çözümlerini birleştiriniz ve sonuçları yorumlayınız.



Bölüm 18

# Yeniden Örnekleme ve Parametrik Olmayan Testler

## Hedefler

Daha önce tartışılan testlerdeki evrenlere ilişkin varsayımlardan daha az sınırlayıcı varsayımlara dayalı hipotezlerin test edilmesinde kullanılabilen yeniden örnekleme ve parametrik olamayan (dağılımdan bağımsız) süreçleri tanıtma.

## Konular


**18.1** Genel Bir Yaklaşım Olarak Özyükleme
**18.2** Bir Örneklem ile Özyükleme
**18.3** Bir Korelasyon Katsayısında Özyükleme Güvenlik Sınırları
**18.4** Eşleştirilmiş İki Örneklem ile Yeniden Örnekleme Testleri
**18.5** Bağımsız İki Örneklem ile Yeniden Örnekleme Testleri
**18.6** Wilcoxon Sıra-Toplamları Testi
**18.7** Wilcoxon Eşleştirilmiş-Çiftler İşaretli-Sıralar Testi
**18.8** İşaret Testi
**18.9** Kruskall-Wallis Tek-Yönlü Varyans Analizi
**18.10** *k* Sayıda İlişkili Örneklem İçin Friedman Sıra Testi

BU KİTAPTA BURAYA KADAR verinin örneklendirildiği evrendeki (veya evrenlerdeki) puanlar dağılımının bir veya daha fazla parametre kestirimleri ve dağılımın şekline ilişkin varsayımları içeren istatistiksel süreçler tartışıldı. Örneğin *t* testinde, evren varyansının ($\sigma^2$) kestiriminde örneklem varyansı ($s^2$) kullanılır ve örneklemin seçildiği evrenin normal olduğu varsayımı vardır. Bu tür testlere **parametrik testler** denir.

**parametrik testler**

**parametrik olmayan testler**

**dağılımdan bağımsız testler**

Parametre kestirimine veya dağılıma ilişkin varsayımlara dayanmayan testler de vardır. Bu tür testler de genellikle **parametrik olmayan testler** veya **dağılımdan bağımsız testler** olarak adlandırılır. Genellikle, bir test, parametrik değilse aynı zamanda dağılımdan bağımsızdır ve gerçekten testin dağılımdan bağımsız doğası bizim için değerli olandır. Bu iki ifade sıklıkla birbiri yerine kullanılıyor olsa da bir şekilde daha genel bir kullanımı olduğundan, burada parametrik olmayan testler ifadesi kullanılacaktır.

**yeniden örnekleme testleri**

Parametrik olmayan testlerin çoğu **yeniden örnekleme testleri** kategorisi altında yer alır. Yeniden örnekleme testleri, açıkça veya üstü kapalı olarak, yokluk hipotezinin doğru olduğu varsayımı altında, bir evrenden alınan çok sayıdaki örneklemden belirlenen sonuçların yorumlanmasına ve elde edilen sonucun yeniden örnekleme sonuçları ile karşılaştırılmasına dayanır. (Bu bölümde daha sonra tartışılacak olan parametrik olmayan sıra (rank) testleri yeniden örnekleme kuramına dayanır; fakat bu testler nadir olarak bu yolla tanımlanır ve pek çok kimse sıra testlerini bu şekilde düşünmez.)

**özyükleme**

**tesadüfîleştirme**

Sıra sayıları yerine, ham puanlarla ilgili yeniden örnekleme süreçleri daha sonra **özyükleme** ve **tesadüfîleştirme** testleri olarak ayrıştırılabilir. Kitap boyunca bu tür süreçlerin her ikisinden de örnekler verilecektir, fakat bunların önemli avantajlarını açıklamak için burada daha ayrıntılı olarak tartışalacaktır. Bu teknikler, *t* testi gibi, parametrik bir testin, gerekliliği olan varsayımlarla ilgili endişe duyduğumuzda veya ortanca üzerine bir güven aralığının biçimlendirilmesi gibi, istediğimiz doğru parametrik süreçlere sahip olduğu halde dağılımın normal olduğuna ilişkin şüphemiz olduğu durumlarda kullanılır.

Parametrik olmayan testlerin başlıca avantajı –ki aynı zamanda en açık olanı- bu testlerin, örneklemin alındığı evren(ler)in biçimiyle ilgili oldukça ciddi sınırlayıcı varsayımlara dayanıyor olmamasıdır. Bu demek değildir ki parametrik olmayan testlerde herhangi bir dağılım varsayımı yoktur; sadece parametrik testler için olanlardan çok daha genel varsayımları gerektirirler. Test edilecek yokluk hipotezi, örneğin, iki evrenin simetrik veya aynı şekle sahip olup olmadıklarına dayalı olabilir. Bununla birlikte bu testlerden hiçbiri dağılımın belirli bir biçimine ilişkin olası bir varsayım içermez; bir başka deyişle, testin geçerliği evrendeki değişkenlerin normal dağılıp dağılmamasından etkilenmez. Diğer taraftan, bir parametrik test genelde normallik varsayımının bir biçimini içerir ve eğer bu varsayım karşılanmazsa, testten elde edilen sonuçlar doğru olmayabilir. Buna ek olarak, parametrik test varsayımlarının şüphe altında olması, belirli bir veri kümesinde, karşılık gelen parametrik olmayan testle karşılaştırıldığında parametrik testin gücünün daha düşük olmasına sebep olabilir. Belki de parametrik olmayan testlerin en açık sözcüsü Bradley'dir[1] (1968). Bradley, parametrik olmayan testlerin altında yatan varsayımların tanımlarını ve bunların rollerini en açık şekilde ifade etmiştir.

Parametrik olmayan testlerin bir diğer özelliği de –ki genellikle çoğunda bir avantaj olarak ortaya çıkabilir - özellikle ortancaya, ortalamaya göre daha duyarlı olmalarıdır. Böylece, eğer ilgilendiğiniz verinin doğası öncelikle ortancaya dayalı ise burada sunulan testler size özellikle faydalı olacaktır.

Hemen hemen her durumda parametrik testlerin kullanılmasının iyi olduğunu savunanlar, parametrik olmayan testlerin varsayımlarında daha özgür olduklarını reddetmemektedirler. Bununla birlikte, parametrik testlerin gerekliliği olarak söz edilen varsayımların pratikte fazlaca sınırlayıcı olduğunu ve parametrik testlerin, dağılım varsayımlarının karşılanmamasından önemli ölçüde etkilenmediklerini kabul etmektedirler. Varsayımların kar-

---

[1] Bradley'nin çalışması, ham puanları yeniden örnekleme süreçlerinin yaygın şekilde kullanılmasından önce geldiğinden, buraya alınmamıştır.

şılanmadığında, parametrik testlerin üstünlüğüne dair bir örnek için Rasmussen'e (1987) bakınız.

Parametrik olmayan testlerin genel olarak en önemli dezavantajı, karşılık gelen parametrik testlere göre daha düşük güce sahip olmalarıdır. Genelde, parametrik testlerin varsayımları sağlandığında karşılık gelen parametrik olmayan testlerin gücü ile aynı güce sahip olması için parametrik olmayan teste karşı gelen parametrik teste göre daha fazla sayıda gözlem (veri) gerekir. Böylece verilen bir veri kümesi için parametrik olmayan teste karşılık gelen parametrik testte, doğru olmayan yokluk hipotezinin reddedilmesi olasılığı daha fazladır. Ayrıca, dağılım varsayımlarının orta derecede karşılanmaması durumunda bile parametrik testlerin avantajını koruduğu düşünülür. Bununla birlikte, çok sayıda çalışma mükemmel uygunluktaki veri kümesi için parametrik olmayan testlerin, karşılık gelen parametrik testlerden daha büyük güce sahip olduğunu göstermektedir. Sorun olan bizim genellikle parametrik olmayan testlerin ne zaman daha güçlü olacağını bilmememizdir.

Bazı parametrik olmayan testlerin fazladan avantajları vardır. Bu testlerin çoğunda ham puanlar sıralandığından ve bu sıra sayıları üzerinde çalışıldığından, bir veya daha fazla uç puandan etkilenmeyen, merkezî eğilimlerdeki farklılığın test edilmesine dayanırlar. Bir veri kümesindeki uç değerler, parametrik testlerin gücünü gerçekten düşürebilir, çünkü bu uç değerler, ortalamanın yanlı bir şekilde uç değerlere doğru kaymasına sebep olduğu gibi, varyansı da olduğundan daha fazla büyütür, ki bu hata teriminin büyümesine sebep olur (ardından ortalama farklar artabilir veya azalabilir).

Sonraki kesim özyüklemeye giriş olacaktır. Bu süreçler, örneklem dağılımının ve standart hatanın analitik olarak üretilemediği, ortanca gibi istatistiklerle ilgilenildiği durumlarda özellikle önemlidir.[2] **Özyükleme süreçleri** yokluk hipotezlerini test etmeye imkân sağlamasına rağmen, geliştirilmelerinin asıl nedeni, özellikle değişkenliğin kestirilmesinde olmak üzere, parametrelerin kestirilmesinde sık kullanılmasıdır. Özyükleme, bir evrenden *çekilen yerine konularak* (with replacement) oluşturulan yeniden örnekleme verisi olmasıyla ayrılır. (Diğer taraftan, **permütasyon testleri**nde örneklem, çekilen yerine konulmadan oluşturulur ve çoğu kez hipotezlerin test edilmesinde kullanılır.) Özyüklemeye ait bu tanıtımın amacı, özyükleme hakkında çok yorucu bir tartışmaya girilmesi değil, özyüklemenin ne olduğunun genel olarak tanıtılmasıdır. Yanlılığı azaltmak için uygun önemli düzeltmeler vardır, ancak özyüklemenin genel doğasını anlamada bizim bunları tartışmamıza gerek yoktur.

**özyükleme süreçleri**

**permütasyon testleri**

Özyüklemenin kullanılmasına ait, biri ortanca diğeri korelasyon katsayılarıyla ilgili iki örnek ele alacağız. Daha sonra, verinin tüm permütasyonlarını aramaya -veya olası tüm permütasyonları elde etmemiz mümkün olabilseydi çekileni geri koymadan örneklemeye dayalı- pertmütasyon testlerine geçeceğiz. Özyükleme, çekileni geri koyarak örneklemeyi gerektirirken, permütasyon testleri **çekileni geri koymadan örneklemeyi** gerektirir.

**çekileni geri koymadan örnekleme**

## 18.1 Genel Bir Yaklaşım Olarak Özyükleme

Önceki bölümlerde, evren dağılımı normal olmadığında veya dağılımın şekli bilinmediğinde bir istatistiğin örneklem dağılımını açıklamada özyükleme tekniklerinin kullanımı görülmüştü. Gerçekten özyükleme, çoğunlukla hipotezlerin test edilmesi yerine evren parametrelerinin kestirilmesinde kullanılır. Bununla birlikte, aşağıda görüleceği gibi, hipotezlerin test edilmesinde de önemli bir rol oynar. Bir an için bir evren ortalamasının güven aralığı sınırlarını belirlemek istediğimizi düşünelim. Eğer evrenin dağılımının normal olduğunu biliyorsak, evren ortalaması ve standart sapmasının kestiricisi olarak örneklem ortalamasını ve standart sapmasını kullanabiliriz. Daha sonra standart sapmayı $n$'ye (örneklem büyüklüğü) bölerek ortalamanın standart hatasını hesaplarız ve sonra da $z$ ya da

2 Eğer evren normal dağılıma sahip değilse, ortancanın standart hatası yaklaşık olarak ortalamanın standart hatasının 1,25 katıdır. Bununla birlikte, eğer dağılım çarpık ise, ortancanın standart hatası kolayca hesaplanamaz.

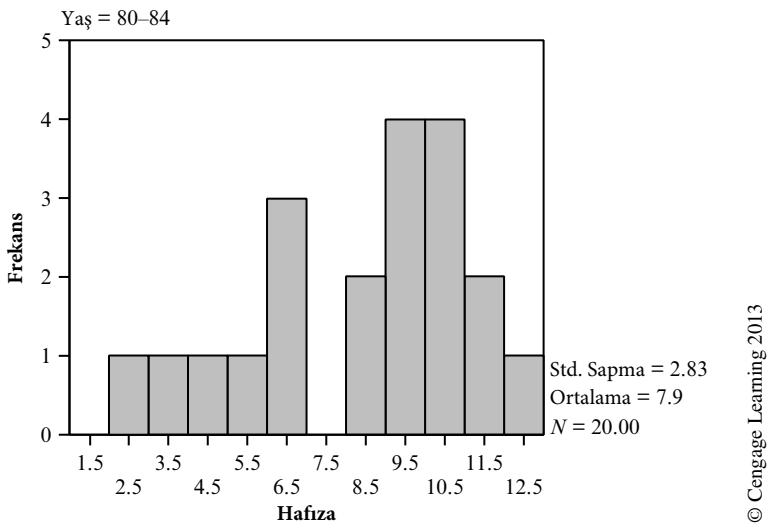
$t_{(,025,\ sd)}$'yi kullanarak güven aralığı sınırlarını buluruz. Ancak bu yöntem sadece örneklemenizin sonsuz büyüklükteki evrenlerden çekildiği durumlarda işe yarar. Örneklemin normal dağılması, evren dağılımının ciddi düzeyde normal olmadığı ve örneklemin küçük olduğu durumlarda çok rastlanan bir durum değildir. Fakat, evren dikkate değer ölçüde normal değil ve örneklem küçük ise veya ortalama yerine farklı bir parametre için güven aralığı belirlenmek istenirse ne olacak? İşte bu durumda, özyükleme devreye girer.

Eğer ortalama için güven aralığının hesaplanması sorulsa ve verinin elde edildiği evrenin normal dağıldığı söylense, bu problem çözülebilir. Bilhassa, ortalamanın standart hatasının, evrenin standart sapmasının (örneklem standart sapması ile kestirilmiş olabilir) *n*'nin kareköküne bölümüne eşit olduğunu biliriz. Normal (veya *t*) dağılımı kullanarak ortalamadan itibaren uygun sayıda standart hataların dağılımlarını hesaplayabilir ve cevabı bulabiliriz. Ancak, ortalama yerine ortancanın güven aralığı sınırları sorulmuş olsaydı, o zaman tıkanacaktık, çünkü ortancanın standart hatasını hesaplamak için kolay bir formülümüz yoktur. Şimdi ne yapacağız? Özyüklemeyi işe koşacağız.

Macauley (1999, bireysel iletişim), yaşlı yetişkinlerden zihinsel durum bilgisi toplamıştır. Bağımlı değişkenlerden biri, 80-84 yaş aralığında yer alan 20 katılımcının Sinirsel Davranışlar Bilişsel Durum Sınavı (Neurobehavioral Cognitive Status Examination) puanlarıdır. Beklenildiği gibi bu dağılım, negatif kayışlıdır, çünkü -kesinlikle tümü değil- ancak katılımcıların bir kısmı zihinsel fonksiyonlarının bir miktarını yitirmiştir. Orijinal veri, Şekil 18.1'de gösterilmiştir. Macauley, bu yaş grubu için evren ortancasının güven aralığını belirlemek istemiştir. Bu durumda, yukarıda altı çizilen problemlerin her ikisiyle de karşı karşıya kalmıştır. Evrenin normal dağıldığı varsayımındaki güven aralığını temel almak mantıklı görünmez ve ortancaya dair bir güven aralığı istenmektedir, fakat ortancanın standart hatası için uygun bir formülümüz yoktur.

Yapılacak şey, evren dağılımının *kesinlikle örneklem ile aynı dağılıma sahip olduğunu* varsaymaktır. Diğer bir deyişle, katılımcı evreninin dağılım şeklinin *kesinlikle* Şekil 18.1'deki ile aynı olduğunu varsaymaktır.

Şekil 18.1'de görüldüğü gibi sonsuz büyüklükteki evreni oluşturmak zorlu bir süreç gibi gelse de oldukça kolaydır ve önceki bölümlerde bu süreç birkaç kez gerçekleştirilmiştir. Tüm yapmak zorunda olduğumuz, Şekil 18.1 ile resmedilen örneklemi temel almak ve bu örneklemden, çekileni geri koyarak istenilen çoklukta örneklemler oluşturmaktır. Çekileni geri koyarak örneklem oluşturulduğundan, sonsuz büyüklükteki evrenden örneklem oluşturuluyor gibidir. Tüm özyükleme programları bu şekilde çalışır. Bir başka deyişle,




**Şekil 18.1** 80-84 yaş arası katılımcıların hafıza puanları örneklem dağılımı

Şekil 18.1'de resmedilen sonsuz büyüklükteki bir evrendeki 20 bireysel gözlem, örneklem dağılımından çekilen geri konularak oluşturulan 20 bireysel gözleminin kesinlikle aynısıdır. İleride, örneklem dağılımının şeklinin aynısına dayalı bir evren oluşturulduğunda, bunu sözde evren olarak tanımlayacağız.

## 18.2 Bir Örneklem ile Özyükleme

Macauley, geç yetişkin katılımcıların hafıza puanlarının ortancasına ait %95 güven aralığını belirlemek istemişti. Yukarıda belirttiğim gibi, evren puanlarının normal dağıldığına ilişkin şüphesi vardı ve ortancanın standart hatasını tanımlayan genel bir formül de yoktu. Her iki durum da istenilen güven aralığının tanımlanmasına yardımcı olmaz. Burada tüm yapılması gereken; evren dağılımının, örneklem dağılımına tümüyle yansımış olduğunu varsaymak ve bu evrenden çok sayıda yeni örneklem (her biri için n = 20) oluşturmaktır. Her bir örneklem için ortancayı hesapladı ve elde ettiği ortancaların dağılımını inceledi. Daha sonra, örneklem ortancalarının %95 etrafındaki değerlerini deneysel olarak belirledi.

Macauley'nin problemi basit bir bilgisayar programı ile kolayca çözülür. Bunlardan biri olan R programı, kitabın web sitesinden bulunabilir. Adı "Giftware"[3] (Armağan) olan bu programa, www.statistics101.net adresinden erişilebilir. (Bu programın denenmesi, yazar tarafından tavsiye edilmektedir.) İndirilebilir bir diğer programı da kısaca göreceğiz. Bu program önce veri kümesini okur ve daha sonra örneklem ortancasını hesaplar. Ardından 10.000 yeni örneklem oluşturur ve ortanca değerlerini belirler. Program daha sonra ham veri dosyasından çekileni geri koyarak 20 gözlem seçer ve örneklem için ortancayı hesaplar. Bu süreci 9.999 kez tekrarlar. Süreç tamamlandığında tekrar tekrar oluşturulan örneklemlerin ortancalarının histogramını çizer ve %95 güven aralığını elde eder, sonuçların en düşük ve en yüksek %2,5'luk kesme değerlerini bulur. Bu sürecin sonuçları Şekil 18.2'de gösterilmiştir. Şekil boyunca çizilen yatay çizgi, 6 ≤ ortanca ≤ 10 olacak şekilde, %95'lik güven aralığını temsil eder. (Örneklem ortancalarının minimum ve maksimum değerleri 5 ve 11'dir.)

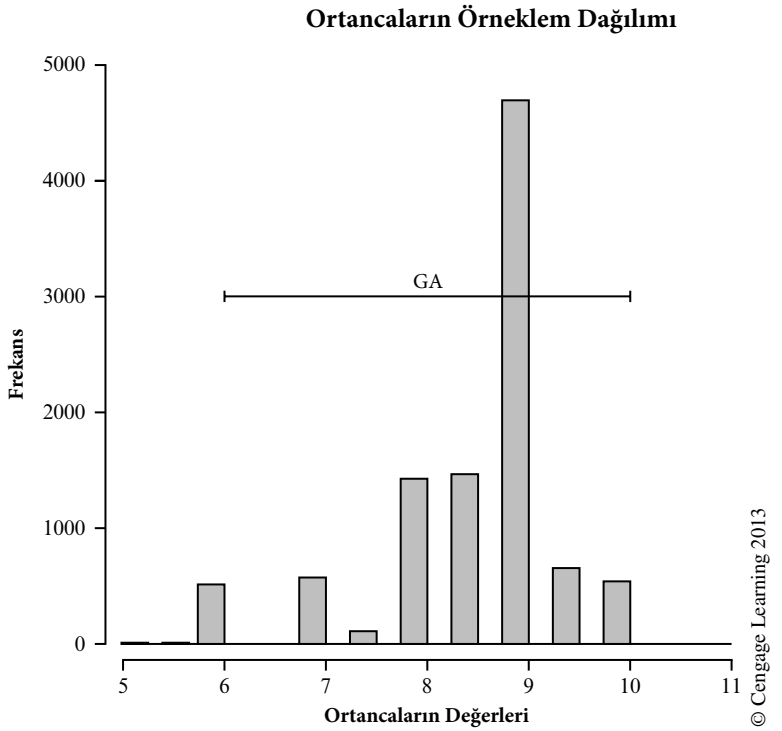
Şekil 18.2'deki dağılım oldukça ayrıktır çünkü ortanca, sınırlı sayıda elemanı olan bir kümenin orta değeri olacaktır. Örneğin, ne kadar çok sayıda örneklem seçerseniz seçiniz ortancanın değeri 9,63 olarak bulunamaz. Bu belirli evren için ortancalar 2 ile 12 arasında yer alan bir tam sayı (veya sıralı veride art arda gelen iki tamsayının ortalaması) olmak zorundadır. Başka bir olasılık yoktur.

En iyi şekilde %95 olasılıklı güven aralığını hesaplamak için, dağılımın uçlarındaki %2,5'luk kesme değerler bulunmalıdır.[4] Ortancalar ile oluşan oldukça ayrık dağılımda, dağılımın en düşük %2,5'luk kesme noktasında hiçbir değer yoktur. En uçta, 5 olan ortanca ve altında 4/10.000 = %,04 ve 6 olan ortanca ve altında (496 + 4) / 10.000 = %5. Dağılımın bir diğer ucunda, 9 ve altında 4006/10.000 = %40,6 ve 10 ve altında 9997/10.000 = %99,97. Yukarıdaki sonuçları elde etmek için ayrılan serilerdeki; 0,025 x 10.000 = 250'nci ve 0,975 x 10.000 = 9750'nci gözlemler alınır. Örneklem dağılımına ilişkin bir örnek bize, verilen yaş aralığındaki bireylerin performansını daha iyi anlamamızı sağlar. Gerçekten genellikle "normal davranış" için alt sınır olarak alınan 10'un altında puanı olan birkaç bireyin olması o yaş grubundaki insanlar için bizi farklı bir güven aralığı araştırmaya yöneltir. Bu veri için özyükleme kullanmak oldukça mantıklı olacakken biz burada, bu sorunun peşine düşmeyeceğiz.

---

3 Giftware ile belirtilen bilgisayar programını ücretsiz olarak indirebilirsiniz. Eğer beğenirseniz yazara bir şey gönderebilirsiniz - en azından hoş bir not.

4 Güven aralığı sınırlarının oluşturulmasında kullanılan en basit yaklaşımdır ve ortancanın örneklem dağılımının 2,5 ile 97,5 yüzdeliklerine dayanır. Çok daha ayrıntılı kestirimler de vardır ancak burada verilen, yaklaşımı en iyi ortaya koyandır.




**Şekil 18.2** 10.000 özyükleme örneklem ortancasının sonuçları

Bu örnek, özyükleme için en ilgi uyandırıcı bir örnek gibi görünmeyebilir, çünkü özyüklemeyi kesin olmayan bir hale getiriyor. Her şeye rağmen, iyi bir örnektir çünkü gerçek bir verinin zaman zaman ne kadar garip olduğunu yansıtmaktadır. Bununla birlikte, göreceğimiz gibi, tüm veri kümeleri bu şekildeki gibi ayrık dağılımlara neden olmaz. Buna ek olarak, sonucun ayrıklığı süreçte değil verinin doğasında vardır. Eğer bu evrenden 10.000 örneklem seçseydik ve *t* değerlerini hesaplasaydık sonuçta elde edilen dağılım nerdeyse ayrık olurdu. Buradaki problem, sürekli (örneğin, normal) dağılımlardan değil, sınırlı sayıdaki farklı değerlerden oluşan dağılımdan örneklemler seçiyor olmamız. Normalliğin varsayılması mantıklı değilse, normal dağılımlardakine benzer iyi grafikler elde edileceğini varsaymak da mantıklı değildir.

## 18.3 Bir Korelasyon Katsayısında Özyükleme Güvenlik Sınırları

Korelasyon problemlerinde standart yaklaşım, korelasyon katsayısını hesaplamak ve daha sonra korelasyonun 0,00'dan manidar olarak farklı olduğu üzerine kurulu alternatif hipotezi test etmektir. Bununla birlikte, 0,00'dan manidar farklılık gösterse dahi önemli olmayacak derecede düşük birçok manidar korelasyon vardır. Bu nedenle son zamanlarda verilen öneme paralel olarak güven sınırlarında da etki büyüklüğü ölçülerinin kullanılması artmıştır.

Bölüm 9'da Fisher'ın arcsin transformasyonunda gördüğümüz gibi,

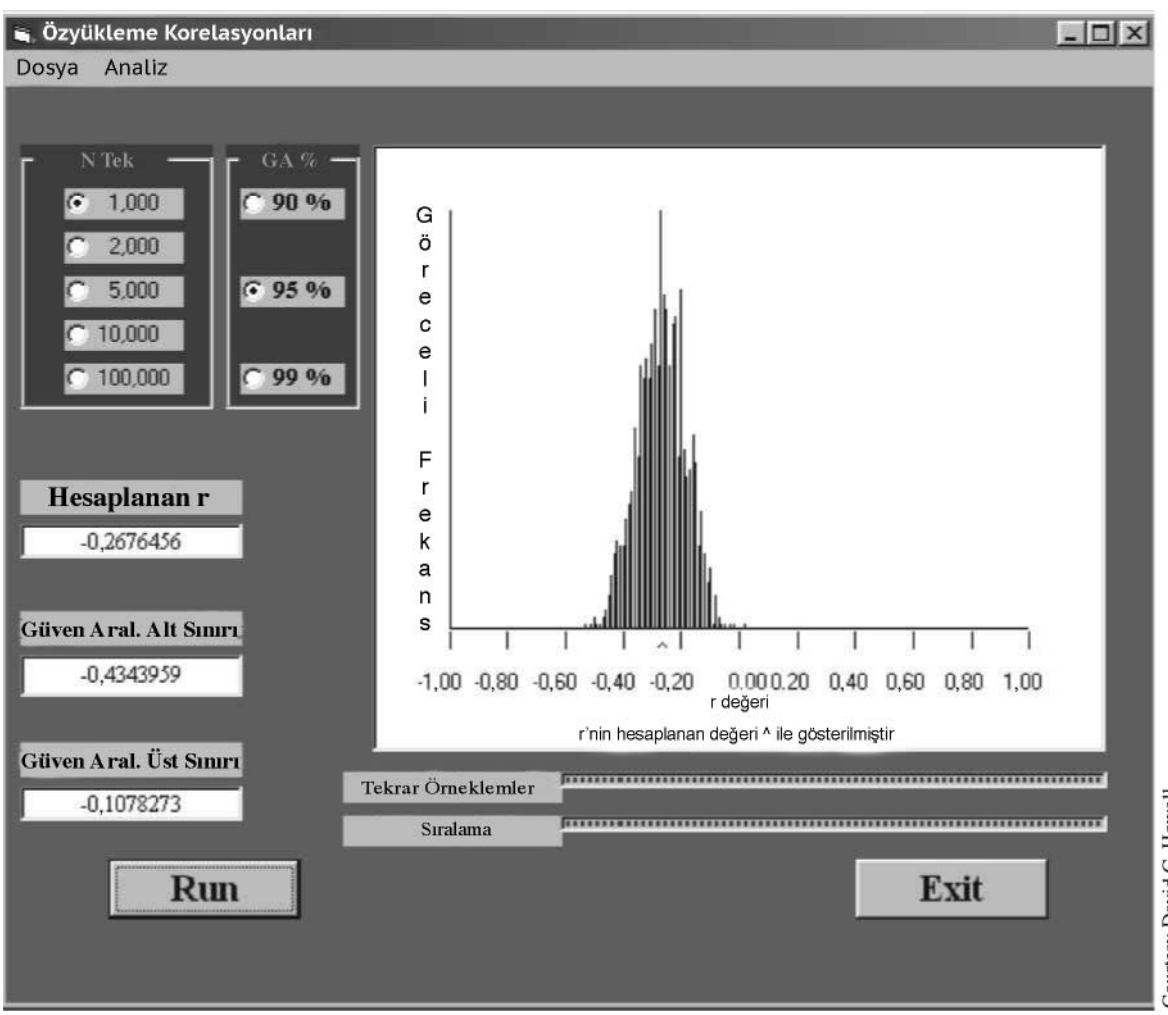
$$r' = (0.5)\log_e\left|\frac{1+r}{1-r}\right|$$

$p \neq 0$ olduğunda $r$'nin çarpık örneklem dağılımının düzeltilmesi için bir yol sağlar. İlgi çekici bir alternatif de evrendeki ilişkiyi yansıtan iki kutuplu veri varsayımı altında öz-

yüklenen örneklemler seçmek ve sonra dağılımın her iki ucundan $\alpha/2$ yüzdelik için kesme değerleri alarak güven sınırlarını basitçe elde etmektir.

Bir örnek olarak, yaşlı yetişkinlerin zihinsel durum puanlarından oluşan Macauley'nin veri kümesini alabiliriz. Macauley'nin veri kümesi yaşları 60 ile 97 arasında olan 123 bireyi içerir ve bireylerin yaşları ile hafıza performansları arasındaki ilişkiye bakabiliriz. İki değişken arasındaki korelasyonun negatif olması olasılığını bekleriz ancak korelasyonun manidarlığı, güçlü bir ilişkinin gerçekte nasıl olduğuna dair daha iyi fikir veren korelasyona ait güven sınırları kadar kullanışlı değildir. Güven sınırlarının elde edilmesinde özyükleme yaklaşımı, örneklemdeki *XY* çiftlerinden, çekileni yerine koyarak, her birinde 123 veri bulunan örneklemler içerir, değişkenler arasındaki korelasyon hesaplanır ve bu süreç çok fazla sayıda tekrarlanır. Daha sonra örneklem dağılımının yüzde 2,5 ve 97,5 değerleri bulunur ve bu değerler bize %95 güven sınırlarını verir.

Bu sürecin gerçekleştirilebileceği bir bilgisayar programı yazdım, bu programa, http://www.uvm.edu/~dhowell/StatPages/Resampling/Resapmlingpackage.zip adresinden ulaşılabilir. (Diğer pek çok özyükleme süreçleri de hesaplanabilir.) Ayrıca kitabın Web sitesinde R programı da vardır. Puan çiftlerinin sözde-evreninden çekileni geri koyarak tekrar tekrar 1000 örneklem çekilerek elde edilen sonuçlar Şekil 18.3'te gösterilmiştir.



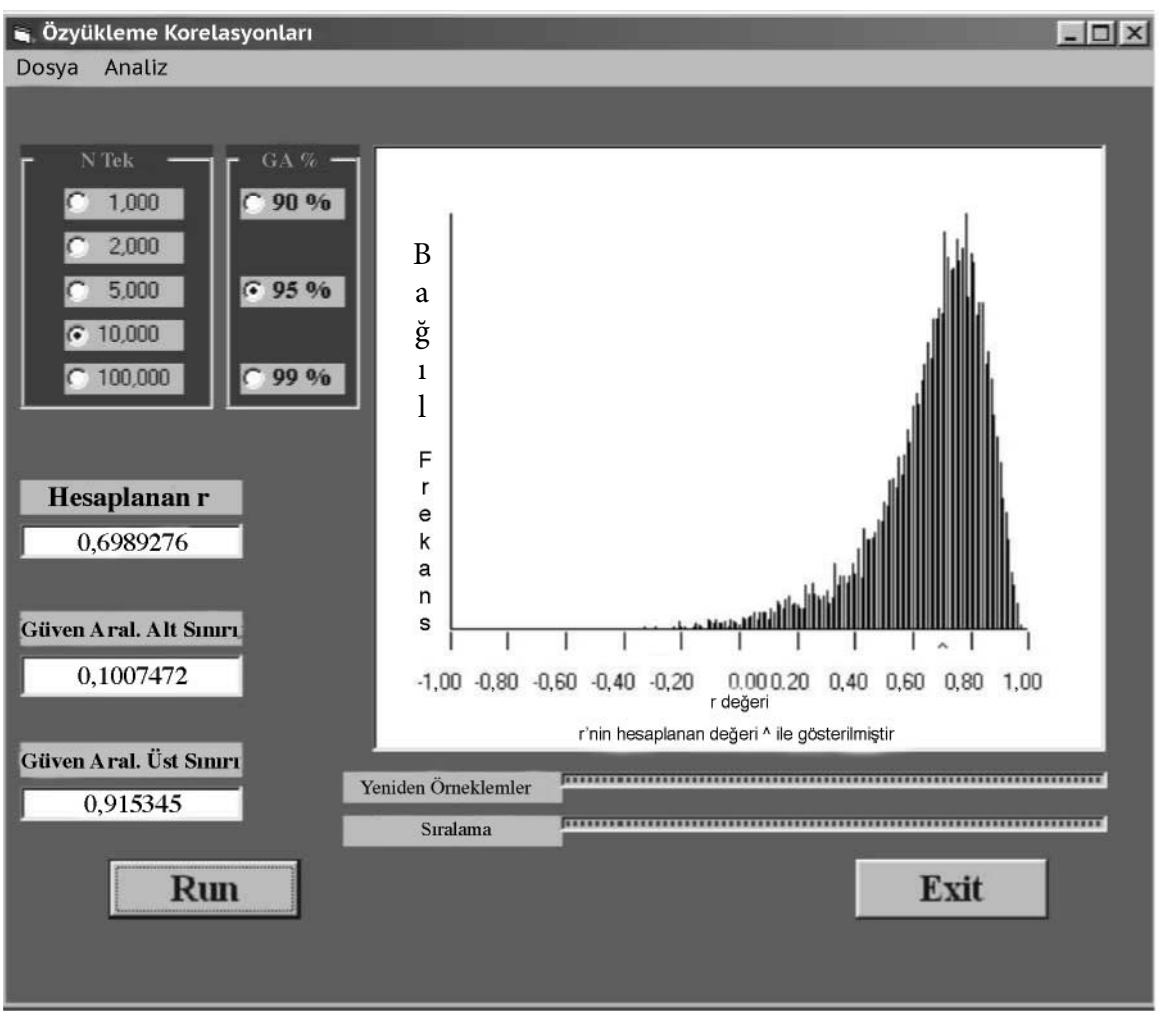

Courtesy David C. Howell

**Şekil 18.3** 10.00 kez üretilen örneklem ortancası özyükleme sonuçları

Şekil 18.3'ün merkezinde *r*'nin örneklem dağılımını görebilirsiniz. Solda elde edilen korelasyon (–,268) ile güven aralığının üst ve alt sınırları yer alır. Bunlar –,43 ve –,11'dir. Bu iki sınır değeri de 0,00'ın aynı tarafında olduğundan, korelasyon katsayısının manidar olduğunu biliriz. Güven aralığının genişliği sizi şaşırtabilir, korelasyon katsayılarının güven aralıkları genellikle böyle geniştir. Aralığın üst sınırında yer alan korelasyon katsayısı manidar gelebilir fakat alt sınırındaki değerden şüphe ederim.

Macauley'den alınan örnekte, korelasyon katsayısı (-,268) oldukça düşük olup 0,00 etrafında yaklaşık simetrik dağılım gösterir. Hoaglin, Mosteller ve Tukey (1983), beta-endorphin'in stres üzerindeki rolünü araştırmıştır. Beta-endorphin düzeyinin stres durumunu arttırıp arttırmadığını test etmekle ilgilenmişlerdir. 19 hastanın ameliyattan önceki 12 saatlik süreçte ve ameliyattan hemen önceki 10 dakika içinde yine aynı hastaların beta-endorphin düzeylerini kaydetmişlerdir. Beta-endorphin verisi üzerinde aynı analiz uygulandığında, çok sayıdaki korelasyon için örneklem dağılımının çarpık olduğu kolayca görülebilir. Bu sonuç, Şekil 18.4'te gösterilmiştir.

Şekil 18.4'te, iki ilginç sonuç görülmektedir. İlki, dikkat edilirse, korelasyon oldukça yüksek (*r* =,699) olduğundan, örneklem dağılımı da oldukça negatif kayışlıdır. Buna ek olarak, güven aralığı sınırları da dikkat çekecek kadar asimetriktir. Üst sınır, *r*'den nerdeyse 20 birim daha yüksek bir değer olan 0,91'dir. Bununla birlikte, alt sınır da *r*'den yaklaşık




Courtesy David C. Howell

**Şekil 18.4** 10.000 örneklemde beta-endorphin veri kümesi için r'nin örneklem dağılımı

60 birim daha düşük bir değer olan 0,11'dir. Her ne zaman çok yüksek korelasyon olursa, örneklem dağılımı çarpık ve güven aralığı sınırları asimetrik olacaktır.

Güven aralığı sınırlarının özyükleme kestirimlerinin mükemmel tartışması, Mooney ve Duval'da (1993) bulunabilir. Yazarlar, uygulanması oldukça kolay olan yanlılık için düzeltme üzerine tartışmışlardır. Hem özyükleme hem de tesadüfîleştirme (randomization) testleri için mükemmel bilgiler Edgington (1995), Manly (1997) ve Efron ve Tibshirani'de (1993) bulunabilir. Efron, özyükleme yaklaşımının tartışmasız en etkin geliştiricisidir ve Tibshirani ile olan kitabı önemli bir kaynaktır. Good'un (2000) çalışmasında permütasyon testlerinin gösterimi yer almaktadır ve Lunnenborg'un (2000) çalışmasında ise yeniden örnekleme yöntemleri ayrıntılı ve oldukça okunabilir bir düzeyde ele alınmıştır.

Özyükleme ve yeniden örnekleme ile ilgili daha fazla bilgi için http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html web adresine başvurulabilir. Bu özel sayfalar, özyüklemeyi de içeren yeniden örnekleme süreçlerinin altında yatan tüm felsefeyi ve parametrik süreçlerde farklılık gösteren yolları kapsar. Bu alan oldukça hızlı genişlemekte ve yeni sonuçlar düzenli bir şekilde yayımlanmaya devam edilmektedir.

Kendi programlarım olan, www.r-project.org adresinden ücretsiz olarak indirilebilen R programını ve bu programın ticari bir uygulaması olan S-Plus'ı kişisel nedenlerle en çok sevdiğimden, esnekliğinden ve tekrarlı örneklemlere uygulanabilirliğinden dolayı yeniden örnekleme süreçlerinde kullanılan çok iyi programlar olarak görmekteyim. Bu bölümdeki örnekler için R'de yazılan örnek programlara kitabın web sitesinden ulaşılabilir.

## 18.4 Eşleştirilmiş İki Örneklem ile Yeniden Örnekleme Testleri

Şimdi özyüklemeden, sözde-evrenden çekileni yerine koyarak çektiğimiz çok sayıda örneklemi içeren, tesadüfîleştirme (randomization) veya permütasyon testlerine geçeceğiz. Sözde evrenden çekilen örneklemler, tüm gözlemler kümesini içerir ve onları tesadüfî olarak karıştırarak durumlara/şartlara yine tesadüfî olarak atarlar. Kısa zaman sonra özyüklemeye döneceğiz.

Hoaglin, Mostller ve Tukey'nin (1983) beta-endorphin'in stres üzerindeki rolü ile ilgili çalışma verisi Tablo 18.1'de fmol/ml[5] birimi ile verilmiştir.

Buradaki veri ikili puanlardan oluştuğundan, özellikle bu ikili puanlar arasındaki farklarla ilgileneceğiz. Endorphin düzeyleri ortalamalarının değişmediğini ifade eden, ortalama fark puanının 0 olduğu yokluk hipotezinin test edilmesini isteyeceğiz. Fark puanları Tablo'nun en alt satırında gösterilmiştir ve açıkça görüleceği gibi farkların çoğu pozitiftir. Bu örnekteki farkların grafiğini çizerseniz oldukça pozitif kayışlı bir dağılım elde edersiniz ki bu durum standart parametrik *t* test kullanmaktan vazgeçilmesine neden olur. Ayrıca, özellikle farkların ortancasıyla ilgilenirseniz *t* testinin kullanılması uygun olmaz. Bu problem, yeniden örnekleme istatistikeri ile çözülür.

**Tablo 18.1** Operasyondan 12 saat ve 10 dakika önce endorphin düzeyleri

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 saat | 10,0 | 6,5 | 8,0 | 12,0 | 5,0 | 11,5 | 5,0 | 3,5 | 7,5 | 5,8 | 4,7 |
| 10 dakika | 20,0 | 14,0 | 13,5 | 18,0 | 14,5 | 9,0 | 18,0 | 6,5 | 7,4 | 6,0 | 25,0 |
| Fark | 10,0 | 7,5 | 5,5 | 6,0 | 9,5 | -2,5 | 13,0 | 3,0 | -0,1 | 0,2 | 20,3 |
| | | | | | | | | | | | |
| 12 saat | 8,0 | 7,0 | 17,0 | 8,8 | 17,0 | 15,0 | 4,4 | 2,0 | | | |
| 10 dakika | 12,0 | 15,0 | 42,0 | 16,0 | 52,0 | 11,5 | 2,5 | 2,1 | | | |
| Fark | 4,0 | 8,0 | 25,0 | 7,2 | 3,50 | -3,5 | -1,9 | 0,1 | | | |

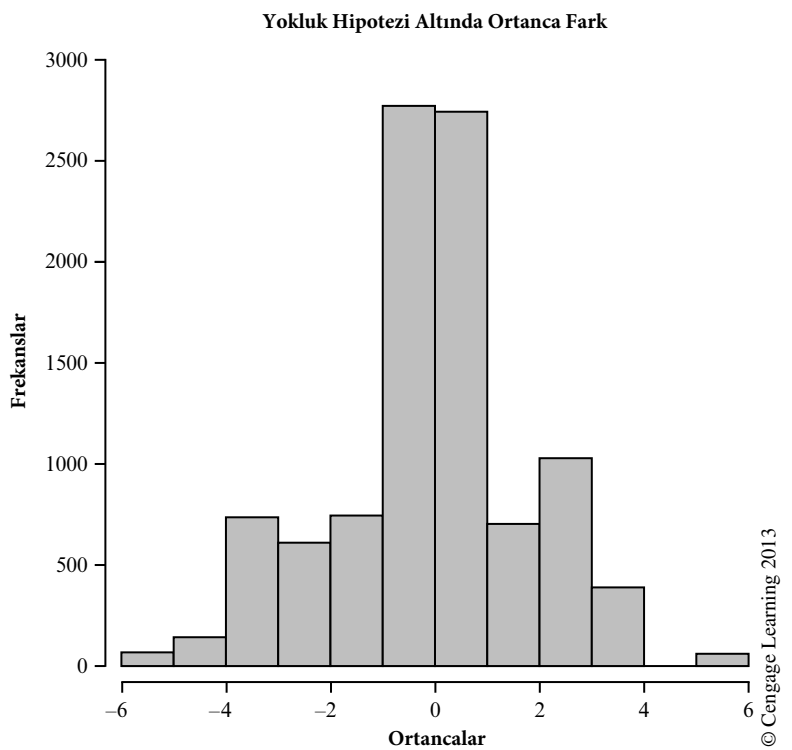
© Cengage Learning 2013

5 Açıklamanın daha kolay olması için, fark puanlarının 0 çıkmasını engellemek amacıyla çok önemsiz iki değişiklik yaptım. Farkların 0,00 olduğu durumlarda normal olarak bu ölçümleri veri kümesinden çıkarırız.

Genellikle permütasyon testi olarak ifade edilen yeniden örnekleme süreci, eğer yokluk hipotezi doğruysa ve stres beta-endorphin düzeyini etkilemiyorsa hastaların 10-dakika değerinin daha yüksek olması 12-saat değerinin daha düşük olması kadar olasıdır düşüncesine dayanır. Eğer bir hastanın değeri 8,0 ile 13,5 ve yokluk hipotezi doğru ise 13,5 değerinin 12-saat ölçümünden elde edilmesi, 10-dakika ölçümünden elde edilmesi kadar olasıdır. 8,0 ve 13,5 değerleri değişik sıralarda elde edilebileceğinden bu teste "permütasyon" testi adı verilir. $H_0$ altında, her bir farklılık, pozitif veya negatif olsun, eşit şansa sahiptir. Bu bize, $H_0$ altında verinin nasıl modellenebileceğini söyler. Farkların, pozitif ve negatif işaretlerinin tesadüfî olduğu ve 19 farklı değerden oluşan çok sayıda örneklem seçilir. Her bir örneklem için farkların ortancası hesaplanır ve bu farkların örneklem dağılımının grafiği çizilir. Hatırlarsak bu grafik, $H_0$ doğru olduğunda farkların örneklem dağılımıdır. Elde edilen ortancalar farkını, $H_0$ testindek bu dağılıma dayalı olarak karşılaştırabiliriz.

Bu sürecin sonuçları, Şekil 18.5'te gösterilmiştir.

Şekil 18.5 oldukça güven vericidir çünkü şekil, yokluk hipotezi doğru olduğunda, beklendiği gibi, yeniden örneklenen ortancalar 0 etrafında simetrik dağılır. Bu şekilden, $H_0$ altında, elde edilen en uç ortanca değerin 6 olduğunu görebiliriz. Burada görülmemesine rağmen, çift yönlü olasılık 18/10000 = ,0018 için sadece 10 ortanca, 6 veya daha yüksek ve 8 ortanca -6 veya daha düşük hesaplanmıştır. Bu sonuçlar bize, "eğer $H_0$'ın doğru olduğu bir modelden örneklem alsaydık, bulduğumuz kadar uç bir ortanca elde etme olasılığımızın çok düşük olacağını" söylerdi. Bu nedenle, yokluk hipotezini reddederiz ve operasyon zamanı yaklaştıkça beta-endorphin düzeyinin arttığı sonucuna varırız. Bu oldukça güzel bir bulgudur çünkü endorphin, vücutta ağrıya karşı bir ilâç gibi etkide bulunmaktadır.




**Şekil 18.5** Endorphin çalışması için yeniden örneklenen ortancaların histogramı

## 18.5 Bağımsız İki Örneklem ile Yeniden Örnekleme Testleri

Şimdi de bağımsız iki örneklem için *t* testinin yeniden örnekleme karşılığını ele alalım. Kullanacağımız örnek, Epping-Jordan, Compas ve Howell'ın (1994), kadınların göğüs kanserinde iyileşmeden kaçınma etkileri üzerinde yaptıkları çalışmanın veri kümesini dikkate alacaktır. Epping-Jordan, kanser oldukları düşüncesinden kaçınan, sürekli olarak bunu düşünmeyi reddeden kadınların bu tür bir kaçınma davranışı sergilemeyen kadınlara göre bir yıllık periyotta daha zayıf bir tedavi süreci geçirip geçirmediklerini araştırmakla ilgilenmişlerdir. Göğüs kanseri tanısı konulduktan hemen sonra kaçınma davranışı gösterme sıklığı veri kümesini toplamışlar, bir yıl sonunda hastalar, hastalıklarında iyileşme olanlar (49 hasta) ve iyileşme olmayanlar veya hayatını kaybedenler (28 hasta) olmak üzere sınıflandırılmıştır. Bu gruplar sırasıyla, "Başarılı" ve "Başarısız" olarak isimlendirilmiştir. Veri kümesi Tablo 18.2'de gösterilmiştir. Epping-Jordan, daha sonra, bu iki grubun daha önce kaydedilen kaçınma davranışı sıklıklarını karşılaştırmışlardır.

Bu örnek için, iki grubun ortalamalarını kolayca karşılaştırabiliyor olmamıza rağmen grupların ortancalarını karşılaştıracağız.

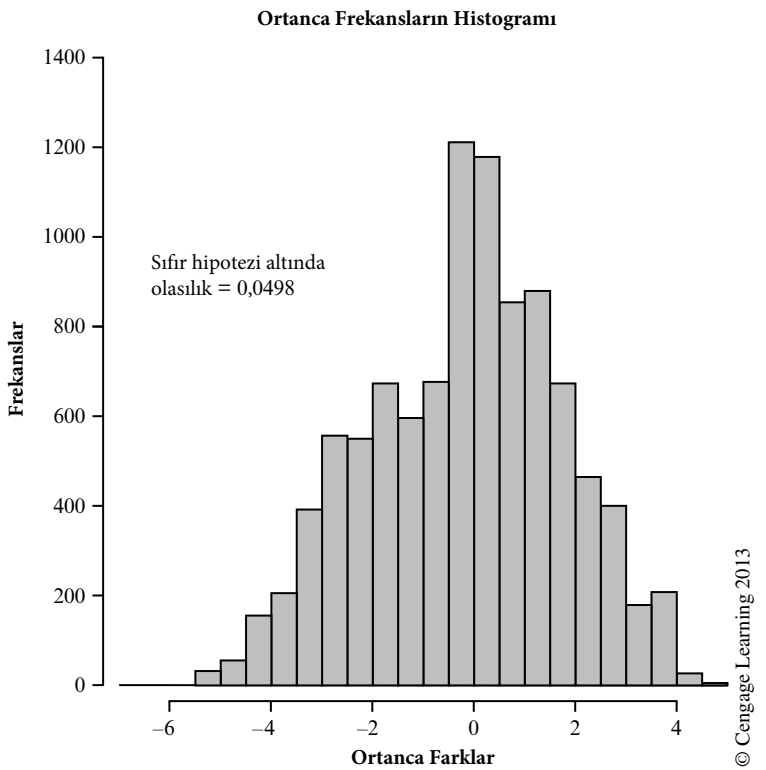
Epping-Jordan'ın çalışmasında, yokluk hipotezi doğruysa, iki grubun (Başarılı ve Başarısız) aynı evrenden geldiklerini düşünebiliriz. Herhangi bir "kaçınma puanı" Başarılı gruptan 49 puan, kalan 18 puanı da Başarısız gruptan tesadüfî örneklem seçerek modellenebilir. (Dikkat edilirse, buradaki örneklem çekilen geri konulmadan oluşturulmuştur. Tüm puanlar, gruplardan birine atanmıştır.) Bu iki grubun ortancaları arasındaki fark, $H_0$ altında elde edilebilecek ortanca farkıyla karşılaştırılır. Bu süreç pek çok kez tekrarlanır ve elde edilen ortanca farklarına bakılır (başarılı gruptan 49 puan, Başarısız gruptan 18 puan tesadüfî olarak alınarak). Son olarak, elde edilen fark ile yokluk hipotezinin modellenmesinden elde edilen fark karşılaştırılır.

Yukarıdaki sürecin uygulanması oldukça kolaydır. 67 puanın bir şapka içinde karıştırılarak, bunlardan 49'unun tesadüfî olarak çekildiği ve ilk gruba atandığı, kalan 18 puanın da ikinci gruba konulduğu düşünülebilir. Daha sonra ortancaları ve ortancaların farklılığı hesaplanır, bu işlem 10.000 kez tekrarlanır. Bu sürecin sonuçları Şekil 18.6'da gösterilmiştir. Yeniden örneklemenin uygulandığı bu sürecin programına, kitabın Web sitesinden ulaşılabilir.

**Tablo 18.2** Epping-Jordan ark.'dan kaçınma veri kümesi (1994)

| Başarılı | | | | Başarısız | | | |
|---|---|---|---|---|---|---|---|
| 19 | 14 | 17 | 10 | 18 | 17 | 17 | 21 |
| 23 | 12 | 10 | 14 | 17 | 15 | 8 | 12 |
| 20 | 21 | 8 | 12 | 16 | 11 | 27 | 18 |
| 8 | 11 | 13 | 23 | 13 | 22 | 18 | 18 |
| 11 | 9 | 8 | 20 | 22 | 16 | | |
| 13 | 15 | 18 | 15 | | | | |
| 13 | 8 | 16 | 15 | | | | |
| 16 | 14 | 11 | 19 | | | | |
| 10 | 12 | 12 | 15 | | | | |
| 8 | 12 | 12 | 17 | | | | |
| 20 | 18 | 25 | 12 | | | | |
| 9 | 23 | 11 | 21 | | | | |
| 13 | | | | | | | |
| Ortanca | 13 | | | 17 | | | |
| *n* | 49 | | | 18 | | | |

© Cengage Learning 2013




**Şekil 18.6** Epping-Jordan ve arkadaşlarının veri kümesine ait yeniden örnekleme sonuçlarının özeti

Tablo 18.2'den, kaçınma puanları ortancasının Başarılı grup için 13, Başarısız grup için ise 17 olduğunu görebiliriz. Başarısız grubun, tedavinin (işlem) başlarındaki kaçınma davranışlarının daha fazla olduğu sergilenmiştir. Kaçınma puanları ortancalar farkı -4'tür. Grafikten anlaşılabildiği gibi, ±4 ortanca farkının görülmesinin çok olası olmadığı ve bunun, yokluk hipotezi altındaki olasılığının 0,0498 olduğu görülmektedir. Başarısız gruptaki hastaların tedavinin başındaki kaçınma davranışının, Başarılı olarak sınıflanan gruptakilere göre manidar bir şekilde daha fazla olduğu sonucu çıkarılabilir.[6]

## 18.6 Wilcoxon Sıra-Toplamları Testi

**Wilcoxon sıra-toplamları testi**

Şimdi, özyükleme ve tesadüfileştirmeden sonra, daha geleneksel olan parametrik olmayan testlere geçeceğiz. Genelde bu şekilde düşünmesek de geleneksel testler gerçekten permütasyon testleridir, aslında değiştirilenler sıralardır. Bu noktayı bir miktar tartışacağız. Bu testlerden en genel olanı ve en iyi bilineni, bağımsız iki örneklem için **Wilcoxon sıra-toplamları testi**'dir. Bir miktar farklılıkların bulunmasına rağmen bu test sıklıkla, bağımsız iki örneklem $t$ testinin parametrik olmayan karşılığı olarak düşünülürse de ondan biraz farklı ve biraz daha geniştir. Wilcoxon testinin yokluk hipotezi, iki örneklemin eş evrenlerden (sadece aynı ortalamaya sahip evrenler değil) tesadüfi seçildiği hipotezidir; ancak özellikle merkezî eğilimdeki evren farklılıklarına karşı duyarlıdır. Böylece, $H_0$'ın reddedil-

6 Bu verinin ortalamaları için $t$ testi uygulasaydık bu olasılık ,0397 olacaktı.

mesi genellikle, iki dağılımın farklı merkezî eğilimlere sahip olduğu anlamını taşır; fakat bunu reddetmenin gerkçekte evrenler arasındaki daha başka farklılıklardan ortaya çıkmış olması olasıdır. Dikkat edilirse, bir kazanç sağlandığında (varsayımların özgürlüğü) diğer taraftan bedeli ödenmektedir (belirliliğin kaybı).

Wilcoxon sıra-toplamları testinin mantıksal temelinin anlaşılması oldukça kolaydır. Varsayalım ki $n_1$ gözlem içeren grup 1 ve $n_2$ gözlem içeren grup 2 olmak üzere iki grubumuz olsun. Bir sonraki varsayımımız da yokluk hipotezinin önemli ölçüde yanlış olması ve grup 1'in örneklemlendiği evrendeki değerler, grup 2'nin örneklemlendiği evrendeki değerlerden genellikle daha düşük olsun. Böylece, eğer $n_1 + n_2 = N$ verinin tümünü, hangi gruba ait olduğunu önemsemeden, en düşükten en yükseğe doğru sıralarsak, düşük sıra numaralarının özellikle grup 1'e yüksek sıra numaralarının grup 2'ye düşeceğini bekleriz. Bir adım daha ilerlediğimizde, her bir grup için sıraların toplamlarını belirlersek, grup 1 için sıraların toplamı grup 2 için sıraların toplamından kayda değer ölçüde düşük olacaktır.

Şimdi de tam tersi durumu ele alalım; yokluk hipotezinin doğru olduğunu ve iki örneklem verisinin eş evrenlerden örneklendiğini düşünelim. Bu durumda, N verinin tümünü hangi gruba ait olduklarını önemsemeden sıralarsak, düşük sıraların da yüksek sıraların da her iki grupta ve grup 1'in sıraları toplamı ile grup 2'nin sıraları toplamının yaklaşık eşit olacağını bekleriz. Bu durumlar Tablo 18.3'te gösterilmiştir.

Wilcoxon, testini yukarıda açıklanan mantığa dayandırır; bir gruptaki sıraların toplamını test istatistiği olarak kullanır. Eğer bu toplam diğer toplama göre çok küçükse, yokluk hipotezi reddedilir. Daha özel olarak küçük gruba verilen sıraların toplamını test istatistiği olarak alırız, eğer $n_1 = n_2$ ise iki toplam da küçük olan grup yerine geçer.[7] Böyle bir durumda yokluk hipotezini test etmek için Wilcoxon istatistiği ($W_S$) tabloları kullanılır.

Özel bir örnek olarak, aşağıda yer alan, yerel bir hastanedeki Kalp Hastalarının bir grup ve aynı hastanedeki Ortopedi Hastalarının kontrol grubu olduğu ve bu grupların son günlerde yaşadıkları stresli olayların sayısının oluşturduğu hipotetik veriyi ele alalım. Çok iyi bilindiği gibi, stresli hayat olayları (evlilik, yeni bir iş, eşin vefatı vb.) hastalıklarla ilişkilidir ve ortalama olarak, çoğu kalp hastasının ortopedik hastalara (bir binayı yıkarken bileğini ya da futbol oynarken bileğini kıranlar) daha fazla stresli olaylar yaşadığını beklemek mantıklıdır.Veriden bu beklentinin doğrulanacağı açıktır. Bazı sebeplerden dolayı stres puanlarının evren dağılımının simetrik olmadığından (özellikle kalp hastaları için, eğer araştırma hipotezimiz doğruysa) şüphe duyarız, parametrik olmayan test kullanmayı tercih ederiz. Bu durumda, bağımsız iki grup olduğundan Wilcoxon sıra-toplamları testini kullanırız.

**Tablo 18.3** $H_0$'ın doğru ve yanlış olduğu durumlardaki tipik sonuçların gösterimi

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $H_0$ Yanlış | | | | | | | | | | | |
| Ham Veri | 10 | 12 | 17 | 13 | 19 | 20 | 30 | 26 | 25 | 33 | 18 | 27 |
| Sıra ($R_i$) | 1 | 2 | 4 | 3 | 6 | 7 | 11 | 9 | 8 | 12 | 5 | 10 |
| $\Sigma(R_i)$ | | | | 23 | | | | | | 55 | | |
| | $H_0$ Doğru | | | | | | | | | | | |
| Ham Veri | 22 | 28 | 32 | 19 | 24 | 33 | 18 | 25 | 29 | 20 | 23 | 34 |
| Sıra ($R_i$) | 4 | 8 | 10 | 2 | 6 | 11 | 1 | 7 | 9 | 3 | 5 | 12 |
| $\Sigma(R_i)$ | | | | 41 | | | | | | 37 | | |

© Cengage Learning 2013

7 Daha küçük gruptaki sıralar toplamı artı daha büyük gruptaki sıralar toplamı bir sabite eşit olduğu için bu toplamlardan sadece birini kullanmaya ihtiyaç duyarız.

| | Kardiyak Hastaları | | | | | | Ortopedik Hastalar | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ham Veri | 32 | 8 | 7 | 29 | 5 | 0 | 1 | 2 | 2 | 3 | 6 |
| Sıra | 11 | 9 | 8 | 10 | 6 | 1 | 2 | 3.5 | 3.5 | 5 | 7 |

Wilcoxon testini uygulamak için önce 11 puanın tümü en düşükten en yükseğe doğru sıralanır, eşit puanlara aynı sıra numarası verilir (Bölüm 10'daki sıralamanın açıklamasına bakınız). Ortopetik grup, iki gruptan daha düşük puanlara sahiptir, eğer bu gruptaki hastalar genellikle daha az stresli yaşam olaylarıyla karşılaşmışlarsa, bu gruba dair belirlenen sıraların toplamı da daha düşük olur. Daha düşük değere sahip grubun (ortopetik grup) sıralar toplamını $W_S$ ile gösterirsek;

$$W_S = 2 + 3{,}5 + 3{,}5 + 5 + 7 = 21$$

buluruz.

Elde edilen WS değerini Wilcoxon tablosunu (Ek WS) kullanarak değerlendirebiliriz. Bu değer, yokluk hipotezi doğru olduğunda şansla elde edilmesi beklenen WS'nin en düşük değeridir. Ek WS'den n1 = 5 denek için daha küçük Grup ve n2 = 6 denek daha büyük grup (n1 her zaman, Grup büyüklükleri eşit değilse küçük gruptaki denek sayısıdır) olmak üzere, a = 5,025 (tek kuyruklu) için giriş 18'dir. Bunun anlamı ,025 düzeyinde, tek yönlü ya da ,05 düzeyinde çift-yönlü test için gruplar arası farklılığın manidar olması için WS'nin 18'e eşit veya daha küçük olması gerekir. WS değerini 21 bulduğumuz için H0'ı reddedemeyiz. Yokluk hipotezi altındaki gerçek olasılık ,1198'dir. (Kıyaslamak amacıyla, bu veriye t testi uygulasaydık, bir örneklemdeki varyansın diğer örneklemdeki varyansın nerdeyse 50 katı daha fazla olduğunu görmezden gelerek ve kalp hastalığı puanlarının dağılımının doğru olduğu kestiriminde bulunursak, 9 serbestlik derecesinde t değeri 1,92 olup p = ,110 ile sonuç manidar olmayacaktı. Ham verinin ortalamalarına yeniden örnekleme programını uygularsak en uç sonucun olasılığı ,059 olacaktır. Benzer test ortancalar için uygulanırsa uç değerin olasılığı da ,059 olacaktır. Sıralara uygulanan yeniden örnekleme testi p = ,11'i üretir).

Ek$W_S$'deki değerler tek-yönlü test içindir ve daha düşük gruptaki sıralar toplamının manidar şekilde *küçük* olması durumunda yokluk hipotezinin reddedilmesine sebep olur. Bununla birlikte, $H_0$ yanlış ise bu durumda daha yüksek sıra numaralarının daha düşük grupta toplanmış olması, sıralar toplamının şans beklentisinden daha düşük olması yerine daha yüksek olması mümkündür. Bu sorunun üstesinden gelmek için bir yol olarak, tüm veri yeniden sıralanır ancak bu kez sıralama yüksekten düşüğe doğru gerçekleştirilir. Eğer bu sıralamayı yaptıysak, daha düşük sıra numaraları şimdi düşük grupta görünecektir ve önceki süreci burada sürdürebiliriz. Bununla birlikte, yeniden sıralanmış veriye bu süreci uygulamak zorunda değiliz. $W_S'$ istatistiği hesaplanarak ve sıra toplamlarının dağılımının simetri özelliği kullanılarak benzer şeyi gerçekleştirebiliriz. $W_S'$ istatistiği, eğer sıralamayı tersine, yüksekten düşüğe yapmışsak bulabileceğimiz düşük grup için sıralar toplamıdır: $W_s'$ istatistiği için

$$W_S' = 2\overline{W} - W_S$$

eşitliğini kullanırız. Bu eşitlikte, $2\overline{W} = n_1(n_1 + n_2 + 1)$'dir ve Ek $W_S$'deki tabloda gösterilmiştir. Daha sonra, $W_S'$ değerini, tek-yönlü dağılımın üst sınır değerini, tablo değeri ile karşılaştırırız. $H_0$'ın çift-yönlü testi için (normalde istenilen) $W_S$ ve $W_S'$ değerlerini hesaplarız, hangisi *küçük* ise o değer ile tabloya gireriz ve listelenen $\alpha$ değerinin iki katını alırız.

$W_S$ ve $W_S'$ değerlerinin gösterimi için Tablo 18.4'te yer alan iki veri kümesini ele alalım. Dikkat edilirse, bu iki kümede de aynı derecede uç değerler vardır, öyle ki ilk veri kümesi için en düşük sıra numaralı beş değerin dördü ve ikinci veri kümesi için en yüksek sıra numaralı beş değerin dördü grup 1'dedir. Ayrıca, küme 1 için $W_S$ değeri küme 2 için $W_S'$ değerine eşittir, tersi de doğrudur. Böylece, eğer daha düşük grup için hem $W_s$ hem de $W_S'$ değerlerini hesaplayacağımız bir kural oluşturursak ve tablolardan $W_S$ ve $W_S'$'nin daha düşük değerine başvurursak, her iki veri kümesi için de aynı sonuca ulaşırız.

**Tablo 18.4** Wilcoxon sıra-toplamları testi için veri örneği

| Küme 1 | Grup 1 | | | | Grup 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | 2 | 15 | 16 | 19 | 18 | 23 | 25 | 37 | 82 |
| Sıralar | 1 | 2 | 3 | 5 | 4 | 6 | 7 | 8 | 9 |
| $W_S = 11$ | | | | | | | | | |
| $W'_S = 29$ | | | | | | | | | |

| Küme 2 | Grup 1 | | | | Grup 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | 60 | 40 | 24 | 21 | 23 | 18 | 15 | 14 | 4 |
| Sıralar | 9 | 8 | 7 | 5 | 6 | 4 | 3 | 2 | 1 |
| $W_S = 29$ | | | | | | | | | |
| $W'_S = 11$ | | | | | | | | | |

© Cengage Learning 2013

## Normal Dağılıma Yaklaşma

Ek $W_S$, $n_1$ ve $n_2$'nin 25'e eşit veya daha düşük değerleri için uygundur. $n_1$ veya $n_2$'nin daha büyük değerleri için normal dağılıma yaklaşan $W_S$ dağılımı kullanılabilir. Bu dağılım, aşağıdaki ortalama ve standart hata değerlerine sahiptir:

$$\text{Ortalama} = \frac{n_1(n_1 + n_2 + 1)}{2}$$

ve

$$\text{Standart hata} = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Dağılım normal olduğu için dağılımın ortalaması ve standart sapması (standart hatası) bilindiğinden $z$ değerini hesaplayabiliriz:

$$z = \frac{\text{İstatistik} - \text{Ortalama}}{\text{Standart Sapma}} = \frac{W_S - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$

Daha sonra en az elde edilen değer kadar düşük $W_S$'nin gerçek olasılık değerinin yaklaşımını normal dağılım tablolarından buluruz. (Eşit $z$ değerleri üreteceklerinden, sadece işaretleri farklılaşabilir, bu durumda $W_S$ veya $W'_S$ değerlerinin hangisinin kullanılacağının bir önemi yoktur.)

Daha yüksek sıra numaralarının düşük gruplarda bulunduğu durumdaki hesaplamaların ve normal yaklaşımın kullanımının gösterimi için (bu şekildeki küçük örneklemler için yaklaşımın kullanılmasına gerçekte ihtiyaç olmamasına rağmen), Tablo 18.5'teki veri kümesini ele alalım. Bu veri, ilk üç ay boyunca hiçbir doğum öncesi bakım almamış ve ilk aydan itibaren bakım almış annelerin bebeklerinin doğum ağırlıklarından (gram olarak) oluşan hipotetik (ancak mantıksız değildir) bir veridir.

Tablo 18.5'teki veri kümesi için daha düşük gruba ait sıralar toplamı 100'e eşittir. Ek $W_S$'den $2\overline{W} = 152$ bulunur ve böylece $W'_S = 2\overline{W} - W_S = 52$'dir. 52, 100'den küçük olduğundan Ek $W_S$'teki tabloya $W'_S = 52$, $n_1 = 8$ ve $n_2 = 10$ değerleriyle girilir. ($n_1$ küçük örneklemin büyüklüğü olarak tanımlanır.) Çift-yönlü test istediğimizden, $\alpha$'nın tablo değerinin iki katı alınır. $\alpha$ = ,05'teki çift-yönlü test için kritik değer 53'tür. Bunun anlamı, $H_0$ doğru ise $W_S$ veya $W'_S$ değerlerinin sadece %5 oranında 53'e eşit veya 53'ten küçük olmasını bekleriz. $W_S$'nin, reddetme bölgesine düşen, elde ettiğimiz değeri 52'dir ve $H_0$'ı reddederiz. Böylece, ilk üç ay boyunca hiçbir doğum öncesi bakım almamış annelerin bebeklerinin daha düşük ağırlıkta doğdukları sonucuna varırız. Bu sonuç, ilk üç aya kadar doğum öncesi

**Tablo 18.5** Farklı düzeylerde doğum öncesi bakım almış olan annelerin doğan bebeklerinin ağırlıklarına ilişkin hipotetik veri

| Bakımın Başlangıcı | | | |
|---|---|---|---|
| Üçüncü Ay | | İlk Ay | |
| Doğum Ağırlığı | Sıralar | Doğum Ağırlığı | Sıralar |
| 1680 | 2 | 2940 | 10 |
| 3830 | 17 | 3380 | 16 |
| 3110 | 14 | 4900 | 18 |
| 2760 | 5 | 2810 | 9 |
| 1700 | 3 | 2800 | 8 |
| 2790 | 7 | 3210 | 15 |
| 3050 | 12 | 3080 | 13 |
| 2660 | 4 | 2950 | 11 |
| 1400 | 1 | | |
| 2775 | 6 | | |

$$W_S = \sum(\text{Grup 2'deki Sıralar}) = 100$$
$$W_S' = 2\overline{W} - W_S = 152 - 100 = 52$$
$$= \frac{W_S - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}}$$
$$= \frac{100 - \frac{8(8 + 10 + 1)}{2}}{\sqrt{\frac{8(10)(8 + 10 + 1)}{12}}}$$
$$= \frac{100 - 76}{\sqrt{126{,}6667}} = 2{,}13$$

© Cengage Learning 2013

bakım almamış olmak, bebeklerin daha küçük doğmasına *neden olmak*tadır anlamına gelmeyecektir, sadece bakımı ertelemeyle (örneğin; genç anneler, az beslenme veya yoksulluk) ilişkili olan bu değerlerin daha düşük ağırlıkla da ilişkili olduğu anlamına gelir.

$W_S$'nin değerlendirilmesi için normal yaklaşımın kullanılması Tablo 18.5'in alt kısmında görülmektedir. Buradan, bulduğumuz değer $z$ = 2,13'tür. Ek z'den, 100 kadar büyük veya 52 kadar küçük $W_S$'nin olasılığı ($z$'nin uç değerleri ±2,13), 2(,0166) = ,033'tür. (R fonksiyonu kullanılarak hesaplanan kesin olasılık değeri ,03428'dir ve ,033'e oldukça yakın bir değerdir.) Bu değer, $\alpha$ = ,05'in geleneksel kesim noktasından daha düşük olduğundan $H_0$'ı reddedeceğiz ve doğum öncesi erken bakım almamış olmak ile daha düşük ağırlıkta bebek dünyaya getirmek arasında ilişki olduğunu söyleyecek yeterli kanıtımız olduğu yorumunu yaparız. Dikkat edilirse, hem kesin çözümle hem de normal yaklaşımla, $H_0$'a göre aynı sonuca ulaşılmıştır. Bununla birlikte, tesadüfîleştirme kullanılarak ortalamalarda yapılan yeniden örnekleme testi $p$ = ,059 (çift-yönlü) değerini verir. (Aynı veri kümesi için $t$'nin hesaplanmasında size yol gösterici olacaktır.)

## Eşit Sayıların Olması Hali

Veri kümesi birbirine eşit puanlar içerdiğinde, sıralara bağlı olan bir test bir şekilde bozulmuş olur. Puanların eşitliği pek çok şekilde ele alınabilir. Eşit puanlara eşit sıralar ata-

nabilir (yapmış olduğumuz gibi), madenî bir para atılabilir ve eşit puanlara ardışık sıralar verilebilir veya $H_0$ hipotezini reddetmenin zor olacağı şekilde farklı sıralar atanabilir. Gerçek uygulamada, çoğu kişi basitçe eşit sıralar atar. İstatistiksel olarak ilerlenebilecek en iyi yol olmasa da bu yöntem açıkça en genel kullanılan ve bizim burada kullanacağımız yöntemdir.

### Yokluk Hipotezi

Wilcoxon sıra-toplamları testi, örneklem alınan iki puan kümesinin eşdeğer evrenlerden geldiği yokluk hipotezini değerlendirir. Bu test, karşılık gelen *t* testinde (ortalamaların ele alındığı), test edilen yokluk hipotezinden daha kapsamlıdır. Eğer iki evren dağılımının aynı şekle ve dağılıma sahip olduğu varsayılıyorsa, sıra-toplamları testi ile test edilen yokluk hipotezi gerçekten iki evrenin merkeze eğilimi (bu durumda ortancalar) ile ilgili olacaktır ve eğer iki evren dağılımı da simetrik ise bu durumda test edilen ortalamalardır. Her durumda, sıra-toplamları testi özellikle merkeze eğilimdeki farklılıklara duyarlıdır.

### Wilcoxon Testi ve Yeniden Örnekleme Yöntemleri

**sıra-tesadüfîleştirme testleri**

Wilcoxon testinin ilginç olan özelliği gerçekten daha önce görmediğiniz bir durumdur. Wilcoxon, testini, sıralanmış veri kümesinde permütasyon testinden üretmiştir ve bu şekildeki testler sıklıkla **sıra-tesadüfîleştirme** testleri olarak ifade edilirler. Diğer bir deyişle, daha önceki verimizi ele alır, bu veriyi sıralarsak ve standart bir permütasyon testi uygularsak (her permütasyonu bir ve yalnız bir kere çekerek uygularsak gerçekten bir tesadüfîleştirme testidir), Wilcoxon testi ile ortaya çıkan sonucun aynısını elde ederiz. Wilcoxon, testini, bilgisayarlardan uzun yıllar önce hesaplamayı başarmıştır, bu nedenle ve sıra sayıları kullanılarak uygulanan testler için için tablolar oluşturmuştur. Herhangi bir hesaplama yapmadan sıralar hakkında pek çok şey bilmekteyiz, örneğin sıraların toplamları ve ortalaması gibi. Eğer elimizde beş sayı varsa bu sayıları 1'den 5'e kadar sıralayacağımızı biliriz ve her bir ham verinin ne olduğu önemli olmaksızın bu sıraların toplamı 15 eder. Bu Wilcoxon'a, sonuçlarının örneklem dağılımını bir kez ve yalnızca bir kez üretmesine izin verir ve böylece tablosunu oluşturur.

### Mann-Whitney *U* İstatistiği

**Mann-Whitney *U* testi**

Wilcoxon sıra-toplamları testinin en genel rakibi **Mann-Whitney *U* testi**dir. Mann-Whitney U testini çok fazla tartışmaya ihtiyaç duymuyoruz; ancak, iki test de birbirine eş testler olduğundan $W_S$ ile $U$ arasında mükemmel bir doğrusal ilişki vardır. Bu testi burada ele almamızın tek nedeni $U$'nun referans alındığı bir durum ile karşılaşabilirsiniz ve bu durumda bu testin ne olduğunu bilmek isteyeceksinizdir. $n$, iki örneklemden küçük olanın büyüklüğünü göstermek üzere,

$$U = \frac{n_1(n_1 + 2n_2 + 1)}{2} - W_S$$

ile ifade edilir. Bu eşitlikten, verilen her örneklem büyüklüğü için $U$ ve $W_S$ arasında sadece bir sabit kadar fark olduğunu (kritik değerlerinde de olduğu gibi) görebiliriz. Bu iki istatistik arasındaki ilişkiden dolayı daima $U$ değerlerini $W_s$ değerlerine dönüştürebiliriz ve Ek $W_S$'yi kullanarak $W_S$ değerlerini değerlendirebiliriz.

## 18.7 Wilcoxon Eşleştirilmiş-Çiftler İşaretli-Sıralar Testi

Wilcoxon saygınlığını sadece bağımsız gruplar için parametrik olmayan en popüler testi geliştirtidiği için değil, aynı zamanda eşleştirilmiş gruplar (veya eşleştirilmiş puanlar) için

en popüler testi geliştirdiğinden kazanmıştır. Bu test, ilişkili örneklemler için *t* testinin parametrik olmayan paralelidir ve ilişkili (eşleştirilmiş) iki örneklemin aynı evrenlerden veya aynı ortalamaya sahip simetrik evrenlerden geldiği yokluk hipotezini test eder. Daha açık olarak, fark puanları (evrende) dağılımının sıfır civarında simetrik olduğu yokluk hipotezini test eder. Bu da normallik varsayımı sağlandığında karşılık gelen *t* testi ile test edilen hipotezin paralelidir.

**Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi**

**Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi**nin altında yatan mantığın geliştirilmesi, sıralar-toplamı testi kadar anlaşılırdır ve basit bir örnek ile gösterilebilir. Varsayalım ki uzun-dönem koşunun kan basıncını düşürdüğünü test etmek istiyoruz. Bu hipotezi test etmek için çok sayıda katılımcının kan basıncını ölçeriz ve katılımcılardan altı ay boyunca sistematik olarak bir koşu programına katılmalarını sağlar ve bu periyot sonunda tekrar kan basınçlarını ölçeriz. Bağımlı değişkenimiz altı aylık periyotta katılımıcıların kan basınçlarındaki değişim olacaktır. Eğer koşu süreci kan basıncını düşürüyorsa katılımcıların çoğunun ikinci kez ölçülen kan basınçlarının daha düşük ve böylece ön-son ölçüm farkının pozitif olmasını; ayrıca, kan basıncı yükselen (ön-son ölçüm farkın negatif olması) katılımcılarda da bu yükselmenin oldukça az olmasını bekleriz. Diğer taraftan, koşmak, kan basıncını kontrol etmede etkili olmayan bir yöntem ise puanlar farkının yaklaşık yarısı pozitif ve yarısı negatif olur ve pozitif farkların büyüklüğü de yaklaşık negatif farkların büyüklüğü kadardır. Bir başka deyişle, $H_0$ gerçekten doğruysa, tahminimiz doğrultusunda çok değişiklik beklemeyiz, tahminimiz dışında küçük değişiklikler bekleriz. Dikkat edilirse, burada iki beklenti vardır: (1) Değişiklilerin çoğu aynı yönlü olacaktır; (2) farklı yönde olanlar da az sayıda olacaktır. İkinci beklenti işaret testine götürür, fakat güç kaybına uğrarız.

Aşağıda verilen sayısal bir örnek ile gösterilmek üzere, Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi uygulanırken önce her bir ikili ölçümün fark puanlarını hesaplarız. Daha sonra, bu fark puanlarının işaretine bakmaksızın sıralarız, farklardaki işaretleri karşılık gelen sıralara atarız ve son olarak pozitif ve negatif sıraları ayrı ayrı toplarız. İki toplamdan mutlak değerce (işaretine bakılmaksızın) küçük olanı *T* test istatistiği (*T*) olarak alır ve Ek *T*'deki tablo değeri ile karşılaştırırız. (*T* hesaplanırken, sıralara cebirsel işaretlerin atanmasının sadece kolaylık için olduğuna dikkat etmek önemlidir. Bu iki durumu birbirinden ayırmak için farklı olanlar yuvarlak içine alınabilir veya altı çizilir.)

Varsayalım ki yukarıda sözü edilen çalışma sonucu, 6-aylık koşu süreci öncesi ve sonrası elde edilen sistolik (kasımlı) kan basıncı veri kümesi aşağıdaki gibi olsun:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Önce: | 130 | 170 | 125 | 170 | 130 | 130 | 125 | 160 |
| Sonra: | 120 | 163 | 120 | 135 | 143 | 136 | 124 | 120 |
| Fark ($B - A$): | 10 | 7 | 5 | 35 | −13 | −6 | 1 | 40 |
| Farkın Sırası: | 5 | 4 | 2 | 7 | 6 | 3 | 1 | 8 |
| İşaretli Sıra: | 5 | 4 | 2 | 7 | −6 | −3 | 1 | 8 |

$$T_+ = \sum(\text{pozitif sıralar}) = 27$$

$$T_- = \sum(\text{negatif sıralar}) = -9$$

© Cengage Learning 2013

Tablonun ilk iki satırı, katılımcıların 6-aylık koşu programı öncesi ve sonrası ölçülen kan basınçlarıdır. Üçüncü satır, "önceki" ölçümlerin "sonraki" ölçümlerden çıkarılmasıyla elde edilen farkları gösterir. Dikkat edilirse, sadece iki katılımcıda negatif değişim gözlenmektedir – artan kan basıncı. Buradaki fark puanları, normale yakın dağılıma sahip bir evreni yansıtıyor gibi görünmediklerinden parametrik olmayan bir test kullanmayı tercih ederiz. Tablonun dördüncü satırında, tüm fark puanları, değişimin yönü göz önüne alınmaksızın sıralanmıştır; beşinci satırda, katılımcıların kan basıncı artanlar ile azalanları ayırt etmek için sıralara uygun olan işaretler eklenmiştir. Tablonun sonunda pozitif ve ne-

gatif işaretlerin toplamı ($T_+$ ve $T_-$) görülür. *T;* $T_+$ ve $T_-$'nin mutlak değerce küçük olanı ifadesiyle tanımlandığından $T = 9$'dur.

Elde edilen bu T değeri, bir kısmı Tablo 18.6'da gösterilen Ek *T*'ye başvurularak değerlendirilir. Bu tablonun formatı daha önce gördüklerimizden oldukça farklıdır. Tablodan nasıl yararlanılacağını kolayca açıklayabilmek için bir benzetmeden yararlanacağız. Varsayalım ki bir madenî paranın hilesiz olup olmadığını test etmek için parayı sekiz kez atalım ve eğer çok az tura gelirse yokluk hipotezini $\alpha = ,05$ düzeyinde (tek-yönlü) reddedelim. Madenî paranın sekiz kez atılması dışında, $H_0$ altında *tam* olarak ,05 olasılığına sahip herhangi bir çıktı kümesi yoktur. Bir veya daha az tura gelme olasılığı ,0352'dir ve iki veya daha az tura gelme olasılığı da ,1445'tir. Böylece, $\alpha = ,05$'te çalışmak istersek, bir veya daha az sayıdaki tura için reddedebiliriz, bu durumda Tip I hata olasılığı ,0352'dir (,05'in altında) ya da iki veya daha az sayıdaki tura için reddedebiliriz, bu durumda Tip I hata olasılığı ,1445'tir (,05'ten çok büyük). Aynı problem Y durumunda da ortaya çıkar çünkü Y de yazı ve tura gelme olasılıklarını veren binom dağılımı gibi ayrık bir dağılımdır.[8] Ek *T*'de $n = 8$ ile $\alpha = ,025$ düzeyinde tek-yönlü (veya $\alpha = ,05$ düzeyinde çift-yönlü) test için 3 (0,0195) ve 4 (0,0546) girdilerini buluruz. Bu girdiler bize, $\alpha = ,025$'te (tek-yönlü) çalışmak istiyorsak, ki bu $\alpha = ,05$'te çift-yönlü test ile aynıdır, ya $T \leq 3$ için $H_0$'ı reddedebiliriz (gerçekte $\alpha$'nın ,0195'e eşit olduğu durumda) ya da $T \leq 4$ için $H_0$'ı reddedebiliriz (gerçekte $\alpha$'nın ,0273'e eşit olduğu durumda). Çift-yönlü test istediğimizden, olasılıkların iki katı alınacaktır: 3 (0,0390) ve 4 (0,0546). *T* değerini 9 elde ettiğimiz için hangi kesme noktasını seçersek seçelim $H_0$'ı reddedemeyiz. Böylece, kısa (6-aylık) periyotta yapılan koşunun kan basıncında bir etki göstermediği şüphesini duymamız için hiçbir neden olmadığı sonucuna

**Tablo 18.6** *T*'nin alt kuyruk kritik değerleri ve bunlara ilişkin olasılıklar (Ek *T*'den özetlenmiştir).

| Nominal α (Tek Yönlü) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ,05 | | | ,025 | | ,01 | | ,005 | |
| *N* | *T* | α | *T* | α | *T* | α | *T* | α |
| 5 | 0 | ,0313 | | | | | | |
| | 1 | ,0625 | | | | | | |
| 6 | 2 | ,0469 | 0 | ,0156 | | | | |
| | 3 | ,0781 | 1 | ,0313 | | | | |
| 7 | 3 | ,0391 | 2 | ,0234 | 0 | ,0078 | | |
| | 4 | ,0547 | 3 | ,0391 | 1 | ,0156 | | |
| 8 | 5 | ,0391 | 3 | ,0195 | 1 | ,0078 | 0 | ,0039 |
| | 6 | ,0547 | 4 | ,0273 | 2 | ,0117 | 1 | ,0078 |
| 9 | 8 | ,0488 | 5 | ,0195 | 3 | ,0098 | 1 | ,0039 |
| | 9 | ,0645 | 6 | ,0273 | 4 | ,0137 | 2 | ,0059 |
| 10 | 10 | ,0420 | 8 | ,0244 | 5 | ,0098 | 3 | ,0049 |
| | 11 | ,0527 | 9 | ,0322 | 6 | ,0137 | 4 | ,0068 |
| 11 | 13 | ,0415 | 10 | ,0210 | 7 | ,0093 | 5 | ,0049 |
| | 14 | ,0508 | 11 | ,0269 | 8 | ,0122 | 6 | ,0068 |
| … | … | … | … | … | … | … | … | … |

© Cengage Learning 2013

8 Benzer durum, Wilcoxon sıra-toplamlar testinde de ortaya çıkar, ancak bu test için olan standart tablolar sadece tutucu bir kesme değeri verirler.

varırız. 6 aydan daha fazla sürecek koşu programı, yaşam için harcanmış bir alışkanlık olacaktır.

## Eşit Veri Olması Hali

Bir veri kümesinde, eşit veri iki şekilde ortaya çıkabilir. Birinde, aynı katılımcılardan önce ve sonra elde edilen puanlar arasındaki fark 0 olabilir (0'ın bir işareti yoktur). Bu durumda, veride yanlılığa sebep olunmasına rağmen, normalde bu katılımcıyı kümeden çıkarırız ve örneklem büyüklüğünü düşürmüş oluruz.

Buna ek olarak, eşit fark puanları, eşit sıralama yapmamıza neden olur. Eşit puanların her ikisi de aynı işarete sahipse bu eşitleri istediğimiz herhangi bir şekilde ayırabiliriz (bu durumun sonuca herhangi bir etkisi olmaz). Eğer puanlar farklı işaretlere sahipse normal olarak eşit puanları atar ve sürece her zamanki gibi devam ederiz.

## Normal Dağılıma Yaklaşım

Örneklem büyüklüğü, Ek *T*'nin üst sınır değeri olan 50'den büyük olduğunda, *T*'nin değerlendirilmesi için normal yaklaşım uygundur. Daha büyük örneklemler için *T*'nin örneklem dağılımının, aşağıda gösterilen ortalama ve standart hata değerleri ile yaklaşık olarak normal dağıldığını biliyoruz:

$$\text{Ortalama} = \frac{n(n+1)}{4} \quad \text{ve} \quad \text{Standart hata} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Böylece, *z* değerini aşağıdaki eşitlik yardımıyla hesaplayabiliriz:

$$z = \frac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

Elde edilen *z* değerini Ek *z*'deki değer ile karşılaştırırız. Buradaki süreç, sıra-toplamları testinde kullanılan sürece doğrudan benzer olduğundan tekrar edilmeyecektir.

Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testinin kullanılmasına ait bir diğer ilginç örnek Manning, Hall ve Gold'un (1990) çalışmasında bulunur. Bu araştırmacılar, glikozun hatırlama üzerindeki etkisini araştırmışlar, özellikle yaşlıların ezberlediklerini hatırlamalarındaki performansları üzerindeki etkileriyle ilgilenmişlerdir. Literatürde, çok az glikoz tüketen katılımcıların hafızalarının zayıf olduğunu ve diğer nöro-psikolojik testlerde düşük performans gösterdiklerine dair dikkate değer bulgular yer alır.

Manning ve diğerleri, 17 yaşlı gönüllüden, sabahın erken saatlerinde yaklaşık 250 gram glikoz veya sakarin ile tatlandırılmış 8-ons limonlu içecek içtikten sonra birtakım testleri uygulamalarını istemişlerdir. Sakarin, şeker tadı verir, ancak kandaki glikoz seviyesini yükseltmez. Her iki durum altındaki katılımcılar, testleri uygulamışlar ve daha sonra bu katılımcılara ait ölçümler eşleştirilmiştir. Sadece 16 katılımcının ölçümlerinin bulunduğu testlerdeki performanslardan birinde, katılımcılardan bir pasajı okumaları ve beş dakika sonra bu pasajı hatırlamaları istenmiştir. Buradaki bağımlı değişken açıkça tanımlanmamıştır ancak tahminimiz, katılımcıların pasajdaki hatırladıkları belirli ifadelerin sayısı olduğudur.

Tablo 18.7'de görülen veri kümesi, Manning ve arkadaşlarının çalışmasında yer alan test sonuçları, standart sapma ve ortalamaya yakın değerler elde edilecek şekilde üretilmiştir. Ek *T*'den $N = 16$ ve $\alpha = ,05$ düzeyinde çift-yönlü test ile bulunan kritik *T* değeri 35 veya 36'dır. $T_{\text{hesaplanan}} = 14,5$, her iki değerden de düşüktür ve böylece elde edilen sonuç manidardır. Glikoz durumunun hatırlamayı arttırdığı sonucu ile Manning ve arkadaşlarınınki ile aynı bulguya ulaşılmıştır.

**Tablo 18.7** Glikoz veya sakarin çözeltisi içtikten sonra yaşlı katılımcıların hatırlama puanları

| Katılımcı | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glikoz | 0 | 10 | 9 | 4 | 8 | 6 | 9 | 3 | 12 | 10 | 15 | 9 | 5 | 6 | 10 | 6 |
| Sakarin | 1 | 9 | 6 | 2 | 5 | 5 | 7 | 2 | 8 | 8 | 11 | 3 | 6 | 8 | 8 | 4 |
| Fark | −1 | 1 | 3 | 2 | 3 | 1 | 2 | 1 | 4 | 2 | 4 | 6 | −1 | −2 | 2 | 2 |
| Pozitif sıralar | | 3 | 12.5 | 8.5 | 12.5 | 3 | 8.5 | 3 | 14.5 | 8.5 | 14.5 | 16 | | | 8.5 | 8.5 |
| Negatif sıralar | −3 | | | | | | | | | | | | −3 | −8.5 | | |

$T_+ = \Sigma$ (positif sıralar) = 121.5
$T_- = \Sigma$ (negatif sıralar) = 14.5

© Cengage Learning 2013

Normal dağılıma yaklaşımın kullanılmasına dair bir örnek olarak, $N = 16$ için $T$'nin 14,5 olduğu değer ile ilişkili normal sapmayı ($z$ puanı) hesaplayabiliriz:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{14{,}5 - \frac{(16)(17)}{4}}{\sqrt{\frac{(16)(17)(33)}{24}}} = -2{,}77$$

Bu z değerinin $H_0$ altındaki çift-yönlü olasılığı ,0056'dır. Ortlamalara uygulanan yeniden örnekleme süreci $p$ = ,002 (çift-yönlü) değerini üretecektir.

## 18.8 İşaret Testi

**işaret testi**

Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi, eşleştirilmiş çiftler farkları için dağılımdan bağımsız mükemmel bir testtir. Bununla birlikte, Student's $t$ testinden farklı olarak, ham veri farklarının yerini sıralar alır, böylece bazı veri noktaları arasındaki ince farklılıklar kaybolarak verinin maksimum kullanımı azalır. Student's $t$ testinin varsayımları sağlandığında, bu test de bir şekilde daha az güce sahiptir. Bununla birlikte, varsayımların sağlanmadığı durumda daha fazla güce sahip olabilir. Varsayımlar açısından daha fazla özgürlük sunarken bunun bedelinin gücün daha da azalmasıyla ödendiği bir test de **işaret testi**dir. İşaret testinde, değerlerin tümü göz ardı edilerek sadece farkların işaretine bakılır ve daha da fazla bilgi kaybı yaşanır. Sonuç olarak daha da fazla güç kaybedilir. Bu testi Bölüm 6'da tartışmıştık ancak tamamlayıcı olması amacıyla burada ikinci bir örnek vereceğim.

Önceki bölümde yer alan Manning ve ark. (1990) örneğini kullanabiliriz. Bu örneğin işaret testi için iyi bir seçim olup olmadığı tartışılabilir; çünkü Wilcoxon testi, çok sayıda eşit sıralara dayanır. Bu tartışma tümüyle ikna edici değildir; çünkü eşit sıralar nasıl ayrıştırılırsa ayrıştırılsın sonuçlar aynı olacaktır, fakat Manning ve arkadaşlarının sonucu yeterince sağlam olup; işaret testinin de istatistiksel olarak manidar çıkması açıklayıcı olacaktır.

Manning ve arkadaşlarından alınan veri Tablo 18.8'de tekrarlanmıştır. Bu veriden, 16 katılımcıdan 3'ünün sakarin alma durumunda hatırlama düzeylerinin yüksek olmasına karşın 13'ünün glikoz alımı durumunda daha yüksek olduğu gözlenmektedir. İşaret testi ba-

**Tablo 18.8** Manninh ve arkadaşlarının (1990) veri kümesinden

| Katılımcı | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glikoz | 0 | 10 | 9 | 4 | 8 | 6 | 9 | 3 | 12 | 10 | 15 | 9 | 5 | 6 | 10 | 6 |
| Sakarin | 1 | 9 | 6 | 2 | 5 | 5 | 7 | 2 | 8 | 8 | 11 | 3 | 6 | 8 | 8 | 4 |
| Fark | −1 | 1 | 3 | 2 | 3 | 1 | 2 | 1 | 4 | 2 | 4 | 6 | −1 | −2 | 2 | 2 |
| İşaret | − | + | + | + | + | + | + | + | + | + | + | + | − | − | + | + |

© Cengage Learning 2013

sitçe, iki farklı durum altında hatırlamanın eşit düzeyde ise "3'e 13" olarak ayırımın ortaya çıkmasının olası olup olmadığı sorusunu sorar.

Bu test, birçok şekilde düzenlenebilir. $p$ = ,50 verildiğine göre, 16 olaydan 13 veya daha fazlasının doğru olması olasılığı binom olasılık dağılımı yardımıyla bulunabilir. Standart tablolardan veya binom formülünden aşağıdaki değerler elde edilir:

| | |
|---|---|
| $p(13)$ | = ,0085 |
| $p(14)$ | = ,0018 |
| $p(15)$ | = ,0002 |
| $p(16)$ | = ,0000 |
| Toplam | ,0105 |

Binom dağılımı, $p$ = ,50 için simetrik olduğundan bu olasılığın iki katını alarak $p$ = ,021 çift-yönlü olasılığını elde ederiz. Bu olasılık, ,05'ten küçük olduğundan yokluk hipotezini reddederiz ve glikoz alınması durumunda hatırlamanın daha fazla olduğu sonucuna varırız.

Bu problemi, Bölüm 5'te verilen normal dağılıma yaklaşımla da çözebiliriz. Sadece yaklaşımın kesinliğindeki farklılık dışında kesinlikle yine aynı sonuca ulaşırız.

Bir diğer üçüncü olasılık da diğerleriyle mantıksal olarak eşit olan $\chi^2$ uyum iyiliği testinin kullanılmasıdır. Bu durumda, her bir hücrede beklenen frekansları 8 olarak alırız; çünkü eğer iki durum eşit hatırlamaya sebep olursa, 16 katılımcının daha iyi hatırlama düzeyinin her bir durum için yarı yarıya olmasını bekleriz. Buradan, aşağıdaki tabloyu düzenleriz:

| | Glikoz | Sakarin |
|---|---|---|
| Gözlenen | 13 | 3 |
| Beklenen | 8 | 8 |

$$x^2 = \sum \frac{(0 - E)^2}{E} = \frac{(13 - 8)^2}{8} + \frac{(3 - 8)^2}{8} = 6,25$$

Serbestlik derecesi 1 ($sd$=1) için $\chi^2$'nin kritik değeri 3,84 olup $H_0$'ı reddederiz ve yine farkın manidar olduğu sonucuna varırız. ($\chi^2 \geq 6,25$'in olasılığı ,0124'tür; küçük örneklemler için binom olasılığı ile yeterli derecede tutarlıdır.) Bu testlerin üçü de az veya çok birbirine denktir ve hangisi daha uygun ise onu kullanabiliriz.

## 18.9 Kruskal-Wallis Tek-Yönlü Varyans Analizi

**Kruskal-Wallis tek-yönlü varyans analizi**

**Kruskal-Wallis tek-yönlü varyans analizi**, üç veya daha çok bağımsız gruplar için Wilcoxon sıra-toplamları testinin doğrudan genelleştirilmiş halidir. Bölüm 11'de tartışılan tek-yönlü varyans analizinin parametrik olmayan karşılığıdır. Tüm örneklemlerin aynı evrenden çekildiği hipotezini test eder ve özellikle merkeze eğilmedeki farklılıklara duyarlıdır.

Kruskal-Wallis testini uygulamak için verinin tümü, hangi gruba ait olduğuna bakılmaksızın sıralanır ve her bir grup için sıralar toplamı hesaplanır. Toplamlar, $R_i$ ile gösterilir. Eğer yokluk hipotezi doğru ise $R_i$'lerin az veya çok birbirine eşit olması beklenir (örneklem büyüklüklerinden kaynaklanan fark dışında). $R_i$'lerin birbirinden farklılık derecesinin ölçüsü aşağıdaki eşitlik ile elde edilir:

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N + 1)$$

**Tablo 18.9** Problem çözme veri kümesine uygulanan Kruskal-Wallis testi

| Depresan | | Uyarıcı | | Etkisiz İlaç | |
|---|---|---|---|---|---|
| Puan | Sıra | Puan | Sıra | Puan | Sıra |
| 55 | 9 | 73 | 15 | 61 | 11 |
| 0 | 1.5 | 85 | 18 | 54 | 8 |
| 1 | 3 | 51 | 7 | 80 | 16 |
| 0 | 1.5 | 63 | 12 | 47 | 5 |
| 50 | 6 | 85 | 18 | | |
| 60 | 10 | 85 | 18 | | |
| 44 | 4 | 66 | 13 | | |
| | | 69 | 14 | | |
| $R_i$ | 35 | | 115 | | 40 |

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1)$$

$$= \frac{12}{19(20)}\left(\frac{35^2}{7} + \frac{115^2}{8} + \frac{40^2}{4}\right) - 3(19+1)$$

$$= \frac{12}{380}(2228{,}15) - 60$$

$$= 70{,}36 - 60$$

$$= 10{,}36$$

$$\chi^2_{.05}(2) = 5{,}99$$

© Cengage Learning 2013

Burada;
$k$ = grup sayısı
$n_i$ = gruptaki gözlem sayısı
$R_i$ = gruptaki sıralar toplamı
$N = \sum n_i$ = toplam örneklem büyüklüğü

anlamındadır. Daha sonra $H$, $k–1$ serbestlik derecesinde $\chi^2$ dağılımı ile karşılaştırılır.

Bir örnek olarak, Tablo 18.9'da bulunan verinin; depresan, uyarıcı ve etkisiz ilâç verilen katılımcıların bir saat içinde 85 problemden çözdükleri (doğru veya yanlış) problem sayısı olduğunu varsayalım. Dikkat edilirse, depresan gruptaki üç katılımcı hiçbir şey yapamayacak kadar depresiftir ve uyarıcı grubundaki üç katılımcı ise 85 problemin tümünü çözmek için girişimde bulunmuştur. Bu veri, şüphesiz normal olmayan bir dağılıma sahiptir ve Kruskal-Wallis testi kullanırız. Hesaplamalar, tablonun alt kısmında verilmiştir. $3 - 1 = 2$ serbestlik derecesinde $\chi^2$ ile test edeceğimiz $H$'nin elde edilen değeri 10,36'dır. $\chi^2_{,05}(2)$'nin kritik değeri Ek $\chi^2$'den 5,99 bulunur. $10{,}36 > 5{,}99$ olduğundan $H_0$ reddedebiliriz ve "üç farklı ilâç farklı performans düzeylerine sebep olmaktadır" sonucuna ulaşırız.

## 18.10 *k* Sayıda İlişkili Örneklem İçin Friedman Sıra Testi

***k* sayıda ilişkili örneklem için Friedman sıra testi**

Bu bölümde tartışacağımız son test, tek-yönlü terkrarlı-ölçümler varyans analizinin parametrik-olmayan karşılığı, ***k* sayıda ilişkili örneklem için Friedman sıra testi**dir. Çok tanınan ekonomist Milton Friedman tarafından geliştirilmiştir. Bu test, ham puanlar yerine sıralara uygulanan tekrarlı-ölçümler varyans analizi ile yakından ilişkilidir. Her bir gruptaki puanların her işlem için puanların özdeş evrenlerden geldiği yokluk hipotezini test eder ve merkezî eğilimdeki evren farklılıklarına oldukça duyarlıdır.

Varsayalım ki "bir konuşma için puanlayıcının, konuşmanın kalitesi hakkındaki kararı (verdiği puan), konuşmacının kullandığı araç sayısıyla ilgilidir" hipotezini test etmek istiyoruz.

Araştırmacı, yerel iş gruplarına çeşitli konularda konferaslar veren 17 konuşmacıya ulaşmış olsun. Her konferansçı aynı konfreansı birbirine denk fakat farklı üç gruba veriyor ve izleyiciler konferansların birinde hiçbir görsel araç kullanmıyorlar, birinde ana noktalar için bazı saydamlar kullanıyorlar ve bir diğerinde de dikkate aldığı her önemli nokta için saydamlar ve çevirici-çizelgeler (flip-charts) kullanıyorlar. Her konfreansın sonunda izleyicilerden, konfreansı veren kişiyi 75-puanlık bir dereceleme ölçeğine göre puanlamaları isteniyor ve izleyicilerin tümünden elde edilen puanların ortalaması bağımlı değişken olarak alınıyor. Tüm konfreansçılar üç durumda da konfreans verdiklerinden verinin ilişkili olmasını bekleriz. İyi konfreansçı olmayan bir kişi ne kadar görsel araç kullanıyor olursa olsun iyi konfreans veremiyordur. Hipotetik veri, Tablo 18.10'da gösterilmiştir ve burada yüksek puan daha yüksek beğeniyi belirtir. *Her bir dinleyicinin* ham puanlarına ait sıra sayıları parantez içinde belirtilmiştir.

Eğer yokluk hipotezi doğruysa her bir konuşmacı için sıralamalar tesadüfi olarak dağılacaktır. Böylece, konuşmacıların biri hiçbir görsel araç kulanmadan en iyi olabilirken bir diğeri pek çok görsel araç kullanarak en iyi olabilir vb. Eğer durum bu ise, her durumda-

**Tablo 18.10** Ders verenlerin puanlanmasına ilişkin hipotetik veri kümesi

| | Görsel Araçların Sayısı | | |
|---|---|---|---|
| Dersi Veren | Hiç | AZ | Çok |
| 1 | 50 (1) | 58 (3) | 54 (2) |
| 2 | 32 (2) | 37 (3) | 25 (1) |
| 3 | 60 (1) | 70 (3) | 63 (2) |
| 4 | 58 (2) | 60 (3) | 55 (1) |
| 5 | 41 (1) | 66 (3) | 59 (2) |
| 6 | 36 (2) | 40 (3) | 28 (1) |
| 7 | 26 (3) | 25 (2) | 20 (1) |
| 8 | 49 (1) | 60 (3) | 50 (2) |
| 9 | 72 91) | 73 92) | 75 (3) |
| 10 | 49 (2) | 54 (3) | 42 (1) |
| 11 | 52 (2) | 57 (3) | 47 (1) |
| 12 | 36 (2) | 42 (3) | 29 (1) |
| 13 | 37 (3) | 34 (2) | 31 (1) |
| 14 | 58 (3) | 50 (1) | 56 (2) |
| 15 | 39 (1) | 48 (3) | 44 (2) |
| 16 | 25 (2) | 29 (3) | 18 (1) |
| 17 | 51 (1) | 63 (2) | 68 (3) |
| | 30 | 45 | 27 |

$$\chi_F^2 = \frac{12}{Nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3N(K+1)$$
$$= \frac{12}{17(3)(4)}(30^2 + 45^2 + 27^2) - 3(17)(4)$$
$$= \frac{12}{204}(3654) - 204$$
$$= 10{,}94$$

© Cengage Learning 2013

ki (sütundaki) sıralar toplamı yaklaşık olarak eşit olur. Diğer taraftan, birkaç görsel araç kullanmak, konferansı en popüler kılıyorsa konferanscıların çoğu bu durumda en yüksek puanlara sahip olur ve üç durumdaki sıralar toplamı da şüphesiz eşit olmaz.

Friedman testini uygulamak için her bir konferansçının ham puanlarını ayrı ayrı sıralar ve her bir durum için sıraların toplamını alırız. Daha sonra, aşağıdaki eşitlik yardımıyla yaptığımız hesaplama ile toplamların değişkenliğini değerlendiririz:

$$\chi_F^2 = \frac{12}{Nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3N(k+1)$$

Bu eşitlikte;

$R_i$ = i. durumu için sıralar toplamı
$N$ = konuşmacıların sayısı
$k$ = durum sayısı

Hesaplanan Ğ2F değeri, k – 1 serbestlik derecesinde standart Ğ2 dağılımı ile değerlendirilir.

Tablo 18.9'daki veri kümesi 2 serbestlik derecesinde $\chi^2_F = 10{,}94$'tür. $\chi^2_{,05}(2) = 5{,}99$ olduğundan $H_0$'ı reddederiz ve konferansların kalitesinin değerlendirilmesi, konferansta kullanılan görsel araçların kullanılmasına bağlı olarak değişir. Elde edilen veri, derslerde bazı görsel araçların kullanılmasını tavsiye ederken çok fazla kullanımın derslerde anlatılanlara gölge düşürebileceğini ifade eder. (*Dikkat*: Test ettiğimiz yokluk hipotezi katılımcılar (konferans verenler) arasında hiçbir şekilde farklılık olmadığını söylemekteydi ve gerçekten katılımcı farkları sıralama sürecinde tamamıyla saf dışı bırakılmıştı.)

## Anahtar Terimler

Parametrik testler (Giriş)
Parametrik olmayan testler (Giriş)
Dağılımdan-bağımsız testler (Giriş)
Yeniden örnekleme testleri (Giriş)
Çekileni geri koyarak örnekleme (Giriş)
Özyükleme süreçleri (Giriş)
Permütasyon testleri (Giriş)
Tesadüfileştirme testleri (Giriş)
Çekileni geri koymadan örnekleme (Giriş)
Wilcoxon sıra-toplamları testi (18.6)
Sıra-tesadüfileştirme testleri (18.6)
Mann-Whitney U testi (18.6)
Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi (18.7)
İşaret testi (18.8)
Kruskal-Wallis tek-yönlü varyans analizi (18.9)
*k* sayıda ilişkili örneklem için Friedman sıra testi (18.10)

## Alıştırmalar

18.1 McConaughy (1980), genç çocukların hikâyeleri organize ederken basit betimleyici ("ve sonra...") modeller kullanırken, yetişkin çocukların sıradan ifadeler ve sosyal çıkarımları birleştirerek kullandıklarını savunmuştur. Varsayalım ki farklı yaşlarda iki çocuk grubundan, henüz okumuş oldukları hikâyeleri anlatmalarını istiyoruz. Daha sonra, özetlerinde kullandıkları çıkarımları sayıyoruz. Elde ettiğimiz bu veri aşağıdaki gibidir:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Genç Çocuklar: | 0 | 1 | 0 | 3 | 2 | 5 | 2 |
| Yetişkin Çocuklar: | 4 | 7 | 6 | 4 | 8 | 7 | |

© Cengage Learning 2013

a. Çift-yönlü sıra-toplamlar testi kullanarak bu veriyi analiz ediniz.
b. Nasıl bir sonuca varırsınız?
c. Yeniden örnekleme yapabileceğiniz programa ulaşabilseniz bu veriyi nasıl analiz ederdiniz?

18.2 Kapp, Frysinger, Gallagher ve Hazelton (1979), beynin bir bölgesindeki (amygdala) doku bozukluğunun genellikle korku ile ilgili belirli tepkilerde (örneğin, kalp atışının yavaşlaması) azalmaya sebep olduğunu kanıtlamışlardır. Eğer korku azalırsa, doku bozukluğu olan hayvanların trenden kaçmaları da daha zor olur çünkü farkındalığı sağlayan uyarıcı azalacaktır. İki grup tavşan olduğunu varsayalım; bir grubun beyin bölgesinde doku bozukluğu olsun, diğeri herhangi bir işlem görmemiş kontrol grubu olsun. Aşağıdaki veri kümesi, her tavşanın kaçınma tepkisini kaç denemede öğrendiğini gösterir:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Doku bozukluğu olan grup | 15 | 14 | 15 | 8 | 7 | 22 | 36 | 19 | 14 | 18 | 17 |
| Kontrol grubu | 9 | 4 | 9 | 10 | 6 | 6 | 4 | 5 | 9 | | |

© Cengage Learning 2013

a. Wilcoxon-sıra toplamları testini (çift-yönlü) kullanarak veriyi analiz ediniz.

b. Nasıl bir sonuca varırsınız?

18.3 Alıştırma 18.2'deki veriyi normal dağılım yaklaşımını kullanarak tekrar analiz ediniz.

18.4 Alıştırma 18.2'deki veriyi uygun bir tek-yönlü test kullanarak tekrar analiz ediniz.

18.5 Nurcombe, Fitzhenry-Coor (1979), diagnostik tekniklerdeki eğitimlerin, klinisyenlerin bir durum ile ilgili karar verirken daha fazla hipotez üretmelerini sağladığını tartışmışlardır. Klinikte çalışmaya yeni başlanmış olan 10 psikiyatristin olduğunu ve bu psikiyatristlerden bir görüşme kaydının olduğu videoları kaydetmelerini istediğimizi varsayalım. Daha sonra yazmış oldukları ifadelerdeki hipotezleri sayalım. Deneyin sonunda benzer videolar ile aynı çalışmayı tekrar edelim.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Bireyler: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Önce: | 8 | 4 | 2 | 2 | 4 | 8 | 3 | 1 | 3 | 9 |
| Sonra: | 7 | 9 | 3 | 6 | 3 | 10 | 6 | 7 | 8 | 7 |

© Cengage Learning 2013

a. Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testini kullanarak veriyi analiz ediniz.

b. Nasıl bir sonuca varırsınız?

18.6 Alıştırma 18.5'i dikkate alarak:

a. Normal yaklaşımı kullanarak veriyi tekrar analiz ediniz.

b. İki analizden elde ettiğiniz cevaplar (18.5a ve 18.6a) biriyle ne kadar tutarlıdır? Niçin tamamıyla aynı sonuçlar elde edilmemiştir?

18.7 Alıştırma 18.6'daki önce ve sonraki veri kümeleri arasındaki farkı test etmek için yeniden örnekleme sürecini nasıl uygulayabilirsiniz?

18.8 İlk-doğan çocukların daha sonra doğan çocuklara göre daha bağımsız olduklarına ilişkin bir görüş vardır. 25-puanlık bir bağımsızlık ölçeği geliştirdiğimizi varsayalım ve 20 ilk-doğan çocuk ile bunların ikinci doğan kardeşlerini bu ölçeği kullanarak puanlayalım. Aynı ölçeğin uygulamasını bu kardeşler yetişkin olduklarında da uygulayalım ve böylece yaş etkisini ortadan kaldıralım. Bağımsız puanları aşağıdaki gibi elde edelim (puanı yüksek olan kişi daha bağımsız anlamına gelir).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Kardeş Çiftleri:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| **Birinci Doğan:** | 12 | 18 | 13 | 17 | 8 | 15 | 16 | 5 | 8 | 12 |
| | 13 | 5 | 14 | 20 | 19 | 17 | 2 | 5 | 15 | 18 |
| **İkinci Doğan:** | 10 | 12 | 15 | 13 | 9 | 12 | 13 | 8 | 10 | 8 |
| | 8 | 9 | 8 | 10 | 14 | 11 | 7 | 7 | 13 | 12 |

© Cengage Learning 2013

a. Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testini kullanarak veriyi analiz ediniz.

b. Nasıl bir sonuca varırsınız?

18.9 Normal yaklaşımı kullanarak Alıştırma 18.8'deki veriyi tekrar analiz ediniz.

18.10 Alıştırma 18.8'deki veri için standart yeniden örnekleme testini nasıl uygularız?

18.11 Alıştırma 18.8'in sonuçlarının beklediğimiz kadar belirgin bir kesim-noktası yoktur. İlk-doğanların puanlarının bir fonksiyonu olarak farkların grafiğini çiziniz. Bu grafik bize ne tavsiye eder?

18.12 Wilcoxon sıra-toplamları testi ile buna karşılık gelen *t* testinin yokluk hipotezlerinin test edilmesi arasındaki fark nedir?

18.13 Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi ile buna karşılık gelen *t* testinin yokluk hipotezlerinin test edilmesi arasındaki fark nedir?

18.14 Parametrik-olmayan testlerin lehine olan argümanlardan biri sıralanmış ölçek düzeyinde olan veri kümeleri için daha uygun olduklarıdır. Bu durum, kitabın daha önceki farklı bir bölümünde yer almıştır. Bu argümanın niçin iyi bir argüman olmadığına dair bir sebep söyleyiniz.

18.15 *t* testi kullanarak yokluk hipotezini reddetmek, uygun olan parametrik-olmayan bir testi kullanarak yokluk hipotezini reddetmeye göre neden daha belirli bir açıklamadır?

18.16 İngilizce I dersini veren üç rakip profesör, en iyi öğrencileri yetiştirme gururu taşıdıklarını iddia etmektedir. Bunu belirlemek üzere, her sınıftan tesadüfi olarak sekiz öğrenci seçilerek tümüne aynı sınav uygulanır. Tarafsız bir profesör, öğrencilerin hangi sınıfta olduğunu bilmeksizin sınav kâğıtlarını puanlar. Elde edilen veri kümesi aşağıda verilmiştir:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Profesör A: | 82 | 71 | 56 | 58 | 63 | 64 | 62 | 53 |
| Profesör B: | 55 | 88 | 85 | 83 | 71 | 70 | 68 | 72 |
| Profesör C: | 65 | 54 | 66 | 68 | 72 | 78 | 65 | 73 |

© Cengage Learning 2013

Uygun olan testi uygulayınız ve sonuca varınız.

18.17 Bir psikolog, suç işlemiş ergenler için düzenlediği grup evinin suçluluğu azaltmada başarılı olduğunu göstermek ister. Evde problemlerin olduğu polis tarafından testpit edillen, ailesiyle birlikte yaşayan dokuz ergeni, üvey bir ailede yaşayan benzer dokuz ergeni ve grup evinde yaşayan dokuz ergeni örneklem olarak alır. Gösterge değişkeni olarak da okul kayıtlarından elde ettiği; son eğitim döneminde ergenlerin okuldan kaçtıkları gün sayısını alır. Aşağıdaki veriyi ele alarak, uygun sonuca varınız:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Doğal Aile: | 15 | 18 | 19 | 14 | 5 | 8 | 12 | 13 | 7 |
| Üvey Aile: | 16 | 14 | 20 | 22 | 19 | 5 | 17 | 18 | 12 |
| Grup Evi: | 10 | 13 | 14 | 11 | 7 | 3 | 4 | 18 | 2 |

© Cengage Learning 2013

18.18 Grup evi çalışmasını değerlendirmek için bir diğer yöntem olarak, suç işlediği bildirilmiş 12 ergeni aldığımızı varsayalım. Bu ergenlerin (1) grup evinde kalmadan önceki bir ay boyunca, (2) grup evinde kaldıkları bir ay boyunca ve (3) grup evinden ayrıldıktan sonraki bir ay boyunca okuldan kaçtıkları gün sayısını alalım. Elde ettiğimiz veri kümesi aşağıda verilmiş olsun:

| Ergenler: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Önce: | 10 | 12 | 12 | 19 | 5 | 13 | 20 | 8 | 12 | 10 | 8 | 18 |
| Süreçte: | 5 | 8 | 13 | 10 | 10 | 8 | 16 | 4 | 14 | 3 | 3 | 16 |
| Sonra: | 8 | 7 | 10 | 12 | 8 | 7 | 12 | 5 | 9 | 5 | 3 | 2 |

© Cengage Learning 2013

Friedman testini uygulayınız. Nasıl bir sonuca vardınız?

18.19 Kitabın bu bölümünde, tek-yönlü varyans analizinin uygulandığı veri için bir tesadüfileştirme (randomizasyon) testi uygulaması verilmemiştir (Alıştırma 18.17'deki gibi), ancak hesaplamaların gerçekleştirilebileceği bir bilgisayar programı olsaydı uygun bir analiz önerilebilirdi. Bu testin ana hatları nasıl olacaktır?

18.20 Alıştırma 18.19'da sözü edilen test, kitabın Web sitesinde Windows programında mümkündür. Bu programı Alıştırma 18.18'deki veri için çalıştırıp sonuçları not ediniz. (Burada bir "read-me" dosyası olacak, burada size yeniden örnekleme programının nasıl çalıştırılacağı gösterilmiştir.)

18.21 Alıştırma 18.18'de tanımlanan çalışmanın Alıştırma 18.19'da tanımlanan çalışmaya göre avantajı nedir?

18.22 Alıştırma 18.15'teki veriye Friedman testi uygulamak mümkün olsa, kaybımız ne olurdu?

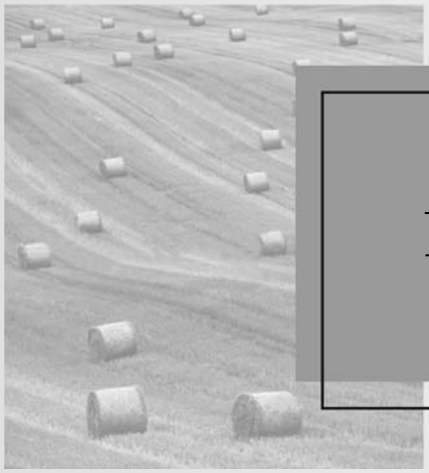
18.23 Alıştırma 18.15'teki veri için ikinci durumda; 10 bireyden 3'ünün daha az, 7'sinin daha fazla hipotez kullandığını söyleyebiliriz. Bu durumu $\chi^2$ ile test edebiliriz. Bu veriye bu testin uygulanmasının Friedman testi uygulamadan farkı nedir?

18.24 İstatistiksel testlerin tarihi gerçekten ilk olarak çay-tadımı deneyi (Fisher, 1935) ile başlar ve bu örnek ile de bu kitabı sonlandırmak uygun olacaktır. Küçük bir çay salonu sahibi, müşterilerinin bir çay poşetinin ilk sunumu ile ikinci ve üçüncü sunumları arasındaki farkı söyleyebileceklerini düşünmez (belki de bunun için küçük bir çay salonudur). Çay salonu sahibi, sekiz farklı marka çay poşeti seçer ve her biriyle üçer bardak çay hazırlar. Daha sonra bu çayların, bir grup müşterisi tarafından 20-puanlık bir ölçek ile puanlanmasını ister (hangi bardakta hangi çayın olduğunu bilmeden). Daha yüksek puanların, çayın daha iyi olduğunu gösterdiği veri aşağıda verilmiştir:

| Çay Markası | Birinci Bardak | İkinci Bardak | Üçüncü Bardak |
|---|---|---|---|
| 1 | 8 | 3 | 2 |
| 2 | 15 | 14 | 4 |
| 3 | 16 | 17 | 12 |
| 4 | 7 | 5 | 4 |
| 5 | 9 | 3 | 6 |
| 6 | 8 | 9 | 4 |
| 7 | 10 | 3 | 4 |
| 8 | 12 | 10 | 2 |

© Cengage Learning 2013

Friedman testi kullanarak, uygun sonuca varınız.



# Ekler

| | |
|---|---|
| **Ek: Veri** | Alıştırmalar İçin Veri |
| **Ek: Bilgisayar Veri Kümeleri** | Bilgisayar Alıştırmalarında Kullanılan Veri Kümeleri |
| **Appendix $\chi^2$** | $\chi^2$ Dağılımının Üst Yüzdelik Noktaları |
| **Appendix F** | *F* Dağılımının Kritik Değerleri |
| **Appendix ncF** | Merkezî Olmayan *F* Dağılımının Kritik Değerleri |
| **Appendix Polinomlar** | Dik Polinom Katsayıları |
| **Appendix Power** | $\delta$'nın Bir Fonksiyonu Olarak Güç ve Manidarlık Düzeyi ($a$) |
| **Appendix *q*** | Student Range İstatistiğinin ($q$) Kritik Değerleri |
| **Appendix *r′*** | *r*'den *r′*'ne Fisher Dönüşümleri |
| **Appendix *t*** | *t* Dağılımının Üst Yüzdelik Noktaları |
| **Appendix *T*** | Wilcoxon Eşleştirilmiş-Çiftler İşaretli-Sıralar Testi İçin *T*'nin Kritik Alt-Kuyruk Değerleri (ve İlgili Olasılıkları) |
| **Appendix *t′*** | Bonferroni Çoklu Karşılaştırmalar Testi İçin Kritik Değerler |
| **Appendix $t_d$** | Dunnett *t* İstatistiğinin ($t_d$) Kritik Değerleri |
| **Appendix $W_S$** | Bağımsız İki Örnekleme Ait ($N_1 \leq N_2$) Sıra-Toplamları Testi İçin $W_S$'in Alt-Kuyruk Değerleri |
| **Appendix *z*** | Normal Dağılım ($z$) |

## Ek: Veri Kümesi

Howel ve Huessy (1985), çocukluk yıllarında dikkat eksikliği semptomu (attention deficite disorder (ADD) -daha önceleri hiperkinezi veya en düşük beyin bozukluğu (inimal brain disfunction) olarak bilinen- gösteren ve göstermeyen 386 çocuk üzerinde bir araştırma yaptıklarını rapor etmişlerdir. 1965 yılında, kuzeybatı Wermont'da bir grup okuldaki bütün ikinci-sınıf öğretmenlerinden öğrencilerinin her biri hakkında genel olarak ADD ile ilgili bir anketi cevaplamaları istendi. Aynı öğrencilere ait anketler, öğrenciler dördüncü ve beşinci sınıflarda iken, sadece bu veri kümesiyle ilgili olarak, tekrar dolduruldu, bu üçer puanın ortalamaları hesaplandı ve ADDSC puanı olarak adlandırıldı. Yüksek puanlar daha çok ADD göstergesi olarak alındı. Bu öğrencilerin dokuzuncu ve onuncu sınıfların sonundaki performansları okul kayıtlarından elde edildi. Bu veri, gelecekteki davranışların öncekilerden kestirilip kestirilemeyeceği, igili değişkenleri ve bunlar arasındaki ilişkilerin akademik olarak incelenip incelenemeyeği sorularının çalışılmasına fırsat sunar. Her bölümün sonundaki pek çok alıştırmada adı geçen bu veriye ait değişkenlerin tanımları şöyledir:

| | |
|---|---|
| **ADDSC** | İlkokulda ADD-benzeri üç davranış puanının ortalaması |
| **CİNSİYET** | 1 = erkek; 2 = kız |
| **REPEAT** | 1 = en az bir sınıfı tekrar etti; 2 = herhangi bir sınıfı tekra etmedi |
| **IQ** | Grupça-uygulanan IQ testinden elde edilen |
| **ENGL** | Dokuzuncu sınıftaki İngilizce düzeyi: 1 = üniversite hazurlık; 2 = genel; 3 = yardım |
| **ENGG** | Dokuzuncu sınıftaki İngilizce notu: 1 = A; 3 = B ve böyle devam ediyor |
| **GPA** | Dokuzuncu sınıftaki genel başarı ortalaması |
| **SOCPROB** | Dokuzuncu sınıftaki sosyal problemleri |
| **DROPOUT** | 1 = liseyi bitirmeden terk; 0 = terk etmedi |

## Ek: Bilgisayar Veri Kümeleri

(www.uvm.edu/~dhowell/methods8/) Web sitesinde çok sayıda veri kümesi vardır. Veri kümeleri gerçek çalışmalardan elde edilen veri kümeleri veya bunların kombinasyonlarıdır; veri, örneklerde ve bölümlerin sonlarındaki alıştırmalarda yer alan veri kümeleri gerçek çalışmalardan ve bunların benzetimlerinden oluşturulmuştur. Aynı zamanda, belirli noktaları açıklamak amacıyla üretilen iki tesdadüfî sayılar kümesi ihtiva eder.

Bu veri kümelerinin hepsi standart ASCII dosyalarıdır, hemen bütün bilgisayar programlarınca okunabilir ve gerektiğinde herhangi bir bilgisayar sisteminde düzenlenebilir (örneğin, Microsoft Worldpad). Ek olarak, ASCII dosyası (bazen text [metin] dosyası veya DOS dosyası olarak anılır) üreten herhangi bir kelime işlemcisi (word processor) yardımıyla düzenlenebilir. Her dosya ilk satırda belirtilen değişkenlere sahiptir.

Aşağıda açıklanan, birçok bölümde ödev olarak verilen alıştırmalarda odaklanılan veri kümeleri alışılmadık şekilde karmaşıktır. Aşağıdaki tanımlarla verinin çekildiği küme ve veri kümesinin nasıl düzenlendiğinin açıklanması amaçlanmıştır. Veri kümeleri tablolardan ve alıştırmalardan basitçe doğrudan çekilebilir ve yapıları metinden çıkarılabilir.

Ayrıca, bu Web sitesinde, kitapta yer alan örnek ve alıştırmaların pek çoğuna ait veri kümelerinin kopyaları da vardır.

### Add.dat

Howell ve Huessy'den (1985) gelir. Veri yukarıda tanımlanmıştır.

| Değişken Adı | Sütunlar | Tanım |
|---|---|---|
| ID | 1-3 | Katılımcı tanımlama numarası |
| ADDSC | 5-6 | 3 yıllık ortalama ADD puanı |
| GENDER (CİNSİYET) | 8 | 1 = erkek; 2 = kadın |
| REPEAT (TEKRAR) | 10 | 1 = sınıf tekrarı var: 2 = sınıf tekrarı yok |
| IQ | 12-14 | Grup-uygulamalı IQ testinden elde edilmiştir |
| ENGL | 16 | İngilizce düzeyi: 1 = üniversite hazırlama; 2 = genel; 3 = yardım |
| ENGG | 18 | İngilizce notu: 4 = A; 3 = B ve böyle devam ediyor |
| GPA | 20-23 | Dokuzuncu sınıftaki genel başarı ortalaması |
| SOCPROB | 25 | Dokuzuncu sınıftaki sosyal problemleri: 0 = hayır; 1= evet |
| DROPOUT | 27 | 1 = liseyi bitirmeden terk; 0 = terk etmedi |

Veri Compas'tan (1990) kişisel iletişimle alınmıştır.

Veri kümesinin ilk dört satırı aşağıda gösterilmiştir:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 45 | 1 | 0 | 111 | 2 | 3 | 2,60 | 0 | 0 |
| 2 | 50 | 1 | 0 | 102 | 2 | 3 | 2,75 | 0 | 0 |
| 3 | 49 | 1 | 0 | 108 | 2 | 4 | 4,00 | 0 | 0 |
| 4 | 55 | 1 | 0 | 109 | 2 | 2 | 2,25 | 0 | 0 |

## Badcancer.dat

Bu veri ve veri kümelerinin arkasındaki çalışmanın tanımı için bundan sonraki bölümde yer alan Cancer.dat adlı dosyayı görünüz. Bu dosyadaki veri, Cancer.dat adlı dosyadaki veriye sadece kasıtlı olarak hata eklenmesiyle elde edilmiştir.

Bu veri kasıtlı olarak bir alıştırma olması amacıyla değiştirilmiştir. Hata eklenmiş ve en az bir değişken bozulmuştur. Verinin doğrusu Cancer.dat dosyasında olandır, bu veri gelecekteki analizlerde kullanılmalıdır. Gerçekte her programın hatalar bulunup düzeltilinceye kadar çalışmaması muhtemeldir ve çalıştırıldığında imkânsız değerler kalacaktır. Birçok hatanın bulunmasının en kolay yolu dosyayının çıktısını almak ve sütunları taramaktır.

## Cancer.dat

Bu veri dosyası Compas'ın (1990, kişisel iletişim) kanser hastalarında ve onların ailelerinde stresin etkileri üzerine yaptığı çalışmadan gelir. Burada, toplanan verinin sadece küçük bir kısmı gösterilmiştir; öncelikle, veri, çocukların davranış problemleri, hastalarda ve onların eşlerindeki psikolojik bozukluklarla ilgilidir. Veri 89 aileden toplanmıştır, kümedeki pek çok veri noktası boştur, çünkü çalışmanın yapıldığı zaman bu veri yoktu. Buna rağmen, bu ön veri kümesi, öndemli psilolojik değişkenleri görme şansı vermektedir.

Cancer.dat dosyasındaki veri için kod kitabına (değişkenler listesi, tanımlar, yerleri ve makul değerleri) ait örnek veri aşağıda gösterilmiştir.

Kayıp veri noktaları "," ile belitilmiştir. Verinin ilk dört satırı aşağıda görülmektedir:

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 101 | 2 | 62 | 50 | 52 | 39 | 52 | 1 | 42 | 44 | 41 | 40 | 42 | . | . | . | . | . | . | . |
| 104 | 1 | 56 | 65 | 55 | 40 | 57 | 2 | 53 | 73 | 68 | 67 | 71 | 1 | 11 | 12 | 28 | 58 | 57 | 60 |
| 105 | 1 | 56 | 57 | 67 | 65 | 61 | 2 | 41 | 67 | 63 | 66 | 65 | 2 | 7 | 7 | 15 | 47 | 48 | 45 |
| 106 | 2 | 41 | 61 | 64 | 53 | 57 | 1 | 60 | 60 | 59 | 67 | 62 | 1 | 6 | 10 | 15 | 49 | 52 | 48 |

| Değişken | Tanımlar | Sütunlar | Normal Değerler |
|---|---|---|---|
| FamNum | Family ID number | 1–3 | 100–400 |
| | GSI Değişkenleri | | |
| | Patient (Hasta) Değişkenleri | | |
| SexP | Hastanın Cinsiyeti | 5 | 1 = erkek; 2 = kadın |
| SomTP | Somatizm *T* puanı | 8–9 | 41–80 |
| DepTP | Depresyon *T* puanı | 12–13 | 42–80 |
| AnxTP | Kaygı *T* puanı | 16–17 | 38–80 |
| HosTP | Düşmanlık *T* puanı | 20–21 | 39–80 |
| GSITP | Küresel Semptom İndeksi *T* puanı | 24–25 | 33–80 |
| | Spouse (Eş) Değişkenleri | | |
| SexS | Eşin cinsiyeti | 27 | 1 = erkek; 2 = kadın |
| SomTS | Somatizm *T* puanı | 30–31 | 41–80 |
| DepTS | Depresyon *T* puanı | 34–35 | 42–80 |
| AnxTS | Kaygı *T* puanı | 38–39 | 38–80 |
| HosTS | Düşmanlık *T* puanı | 42–43 | 39–80 |
| GSITS | GSI *T* puanı | 46–47 | 33–80 |
| | Child Behavior Checklist (Çocuk Davranışları Kontrol Listesi) Değişkenleri | | |
| SexChild | Çocuğun cinsiyeti | 49 | 1 = erkek; 2 = kadın |
| Intern | İçselleştirme alt ölçeği | 51–52 | 0–98 |
| Extern | Dışsallaştırma alt ölçeği | 54–55 | 0–102 |
| TotBP | Toplam davranış problemleri | 57–58 | 0–240 |
| InternT | İçselleştirme *T* puanı | 60–61 | 33–100 |
| ExternT | Dışsallaştırma *T* puanı | 63–64 | 30–100 |
| TotBPT | Toplam davranış problemi *T* puanı | 66–67 | 30–100 |

Veri, Compas'tan (1990) kişisel iletişim yoluyla alınmıştır.

## Epineq.dat, Epinuneq.dat

Introini-Collison ve Mcgauch (1986) "vücutta üretilen hormonlar hafızada rol oynarlar" hipotezini incelemişlerdir. Özel olarak, eğitim sonrası epinephrine enjeksiyonlarının önceden öğrenilmiş ayırımcılığın kalıcılığı üzerindeki etkisine bakmışlardır. İlk eğitilen farelere Y-labirentinin sol kolunu seçerek hafif şoktan kaçmayı öğrettiler. Eğitimin hemen arkasından araştırmacılar, farelere 0,0; 0,3 veya 1,0 mg/kg epinephrine enjekte ettiler, burada yüksek dozlar engellenmiştir.

1 gün, 1 hafta veya 1 ay süren bir eğitimden sonra her fare tekrar Y-labirentine konuldu. Fakat bu defa, şoktan sonra sağ koluna doğru labirentten kaçmasına izin verildi. Muhtemelen, orijinal eğitimde güçlü hafıza, yeni görevi öğrenmede daha çok engelleyici olabilir ve denekler daha fazla hata yapabilir.

Bu deneye ait Epineq.dat ve Epinuneq.dat adlarıyla iki veri kümesi vardır. Orijinal çalışmada 18 hayvan üç dozaj grubu halinde kullanıldı, 1 gün sonra test edildi ve her gruptan 12 hayvan bir hafta ve bir aylık dönemlerle test edildi. Hatırlama aralığında 5 katılımcı 1,0 mg/kg grubundan çıkarılmış olmasına rağmen, Epinuneq.dat'tan elde edilen sonuçları içerecek şekilde orijinal sonuçlarla uyumlu hipotetik veri üretildi. İkinci veri kümesi 9 hüc-

renin her birinde 12 gözlem olacak şekilde oluşturuldu ve Epineq.dat olarak adlandırıldı. Her iki durumda da, tam sayı olan gerçek veriye göre hafif muhafazakâr sonuçlar verecek, fakat $H_0$'a göre kararların aynı olacağı veri oluşturmaya ihtiyaç vardır.

Her iki veri kümesinde, üç basamaklı ID vardır; dozaj 5. sütunda 1, 2 veya 3 olarak kodlanmıştır; tutma (retention) aralığı 7 sütunda 1, 2, veya 3 ile kodlanmıştır; öğrenmedeki ayırıma ait hata sayıları 9'uncu sütında belirtilmiştir. Veri kümesine ait ilk dört satır aşağıdaki gibidir:

```
001  1    1    0
002  1    1    3
003  1    1    4
004  1    1    2
```

## Mirault.dat

Mirault (1990), 381 üniversite öğrencisından veri toplamıştır, bunlardan bazılarının ailesi öğrencilerin çocukluk yıllarında ölmüştür. Üç grup öğrenci vardır. Grup 1'dekiler çocukken ebeveynlerini kaybedenlerden oluşturulmuştur. Grup 2'dekilerin ebeveynleri halen hayattadırlar ve birbirleriyle evlidirler. Grup 3'teki öğrencilerin ebeveynleri boşanmışlardır.

Mirault, ebeveyn kaybının, kişinin mevcut sempton belirtileri düzeyine (Brief Symptom Inventory, Dragos, 1983 ile ölçülen) ve bireyin algıladığı geleceğe ait güvenlik eksikliği üzerindeki etkisinin gözlenmesiyle ilgiliydi. İlgi uzayında, veri kümesi sadece güvenlik eksikliği ölçümlerini ihtiva etmektedir, alt ölçek ölçümlerini bulundurmamaktadır. Aynı zamanda sosyal destek için bir ölçüm vardır. Bütün ölçülerde yüksek puanlar, ölçülen özelliğin yüksekliğinin ölçüsünü belirtir.

Değişkenler ve bunların dosyadaki yerleri aşağıdaki örnek veride listelenmiştir.

Kayıp veri noktaları "." ile belirtilmiştir. Verinin ilk üç satırı aşağıdaki örnekte gösterilmiştir.

```
002  2  1  1  4  2  .  .  42  53  59  57  49  57  47  51  46  51  112  24  66
007  1  2  1  2  .  1  18 65  80  64  71  72  73  63  67  67  72  100  23  73
008  2  2  1  1  4  .  .  52  67  60  62  65  78  60  65  58  65  118  28  64
```

| Değişken Adı | Sütunlar | Tanım |
|---|---|---|
| ID | 1–3 | Katılımcı kimlik numarası |
| Group | 5 | 1 = ölü; 2 = evli; 3 = boşanmış |
| Gender | 7 | 1 = erkek; 2 = kadın |
| YearColl | 9 | 1 = ilk yıl; 2 = ikinci sınıf ve devamı |
| College | 11 | 1 = güzel sanatlar ve bilimler; 2 = sağlık; 3 = mühendislik; 4 = işletme; 5 = tarım |
| GPA | 13 | 4 = A; 3 = B; 2 = C; 1 = D; 0 = F |
| LostPGen | 15 | Ölmüş olan ebeveynin cinsiyeti |
| AgeAtLos | 17–18 | Ebeveyn öldüğünde öğrencinin yaşı |
| SomT | 20–21 | Somatizasyon *T* puanı |
| ObsessT | 23–24 | Takıntı *T* puanı |
| SensitT | 26–27 | Kişiler arası duyarlılık *T* puanı |
| DepressT | 29–30 | Depresyon *T* puanı |
| AnxT | 32–33 | Kaygı *T* puanı |

| Değişken Adı | Sütunlar | Tanım |
|---|---|---|
| HostT | 35–36 | Düşmanlık T puanı |
| PhobT | 38–39 | Fobik kaygı T puanı |
| ParT | 41–42 | Paranoit düşünce T puanı |
| PsyT | 44–45 | Psikotizm T puanı |
| GSIT | 47–48 | Küresel semptom indeksi T puanı |
| PVTotal | 50–52 | Algılanan güvenlik açığı T puanı |
| PVLoss | 54–56 | Algılanan güvenlik açığı toplam puanı |
| SuppTotl | 58–60 | Sosyal destek puanı |

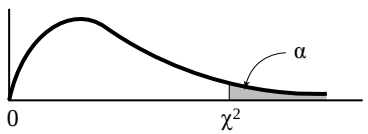
Mireault ve Bond'dan (1992) veri

### Stress.dat

Bu dosyadaki veri Compas ve arkadaşlarının kanser hastalarının stres ve onunla başa çıkma üzerine topladıkları verinin bir alt kümesidir. Dosya, aile numarası, cevaplayıcının cinsiyeti (1 = Erkek; 2 = Kadın) cevaplayıcının rolü (1 = Hasta; 2 = Eş) ve iki stres ölçümüdür (biri tanı konulduktan kısa süre sonra ve biri 3 ay sonra). Değişkenler aşağıdaki sıradadır: FamNum, Gender, Role, Time 1, Time 2. İlk altı durum aşağıdaki gibidir:

```
101   2   1   2   .
101   1   2   2   .
104   1   1   4   .
104   2   2   5   .
105   1   1   3   4
105   2   2   5   4
```

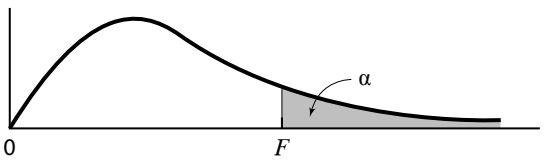
# Ek $\chi^2$: Dağılımın Üst Yüzdelik Noktaları




| df | 0,995 | 0,990 | 0,975 | 0,950 | 0,900 | 0,750 | 0,500 | 0,250 | 0,100 | 0,050 | 0,025 | 0,010 | 0,005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,00 | 0,00 | 0,00 | 0,00 | 0,02 | 0,10 | 0,45 | 1,32 | 2,71 | 3,84 | 5,02 | 6,63 | 7,88 |
| 2 | 0,01 | 0,02 | 0,05 | 0,10 | 0,21 | 0,58 | 1,39 | 2,77 | 4,61 | 5,99 | 7,38 | 9,21 | 10,60 |
| 3 | 0,07 | 0,11 | 0,22 | 0,35 | 0,58 | 1,21 | 2,37 | 4,11 | 6,25 | 7,82 | 9,35 | 11,35 | 12,84 |
| 4 | 0,21 | 0,30 | 0,48 | 0,71 | 1,06 | 1,92 | 3,36 | 5,39 | 7,78 | 9,49 | 11,14 | 13,28 | 14,86 |
| 5 | 0,41 | 0,55 | 0,83 | 1,15 | 1,61 | 2,67 | 4,35 | 6,63 | 9,24 | 11,07 | 12,83 | 15,09 | 16,75 |
| 6 | 0,68 | 0,87 | 1,24 | 1,64 | 2,20 | 3,45 | 5,35 | 7,84 | 10,64 | 12,59 | 14,45 | 16,81 | 18,55 |
| 7 | 0,99 | 1,24 | 1,69 | 2,17 | 2,83 | 4,25 | 6,35 | 9,04 | 12,02 | 14,07 | 16,01 | 18,48 | 20,28 |
| 8 | 1,34 | 1,65 | 2,18 | 2,73 | 3,49 | 5,07 | 7,34 | 10,22 | 13,36 | 15,51 | 17,54 | 20,09 | 21,96 |
| 9 | 1,73 | 2,09 | 2,70 | 3,33 | 4,17 | 5,90 | 8,34 | 11,39 | 14,68 | 16,92 | 19,02 | 21,66 | 23,59 |
| 10 | 2,15 | 2,56 | 3,25 | 3,94 | 4,87 | 6,74 | 9,34 | 12,55 | 15,99 | 18,31 | 20,48 | 23,21 | 25,19 |
| 11 | 2,60 | 3,05 | 3,82 | 4,57 | 5,58 | 7,58 | 10,34 | 13,70 | 17,28 | 19,68 | 21,92 | 24,72 | 26,75 |
| 12 | 3,07 | 3,57 | 4,40 | 5,23 | 6,30 | 8,44 | 11,34 | 14,85 | 18,55 | 21,03 | 23,34 | 26,21 | 28,30 |
| 13 | 3,56 | 4,11 | 5,01 | 5,89 | 7,04 | 9,30 | 12,34 | 15,98 | 19,81 | 22,36 | 24,74 | 27,69 | 29,82 |
| 14 | 4,07 | 4,66 | 5,63 | 6,57 | 7,79 | 10,17 | 13,34 | 17,12 | 21,06 | 23,69 | 26,12 | 29,14 | 31,31 |
| 15 | 4,60 | 5,23 | 6,26 | 7,26 | 8,55 | 11,04 | 14,34 | 18,25 | 22,31 | 25,00 | 27,49 | 30,58 | 32,80 |
| 16 | 5,14 | 5,81 | 6,91 | 7,96 | 9,31 | 11,91 | 15,34 | 19,37 | 23,54 | 26,30 | 28,85 | 32,00 | 34,27 |
| 17 | 5,70 | 6,41 | 7,56 | 8,67 | 10,09 | 12,79 | 16,34 | 20,49 | 24,77 | 27,59 | 30,19 | 33,41 | 35,72 |
| 18 | 6,26 | 7,01 | 8,23 | 9,39 | 10,86 | 13,68 | 17,34 | 21,60 | 25,99 | 28,87 | 31,53 | 34,81 | 37,15 |
| 19 | 6,84 | 7,63 | 8,91 | 10,12 | 11,65 | 14,56 | 18,34 | 22,72 | 27,20 | 30,14 | 32,85 | 36,19 | 38,58 |
| 20 | 7,43 | 8,26 | 9,59 | 10,85 | 12,44 | 15,45 | 19,34 | 23,83 | 28,41 | 31,41 | 34,17 | 37,56 | 40,00 |
| 21 | 8,03 | 8,90 | 10,28 | 11,59 | 13,24 | 16,34 | 20,34 | 24,93 | 29,62 | 32,67 | 35,48 | 38,93 | 41,40 |
| 22 | 8,64 | 9,54 | 10,98 | 12,34 | 14,04 | 17,24 | 21,34 | 26,04 | 30,81 | 33,93 | 36,78 | 40,29 | 42,80 |
| 23 | 9,26 | 10,19 | 11,69 | 13,09 | 14,85 | 18,14 | 22,34 | 27,14 | 32,01 | 35,17 | 38,08 | 41,64 | 44,18 |
| 24 | 9,88 | 10,86 | 12,40 | 13,85 | 15,66 | 19,04 | 23,34 | 28,24 | 33,20 | 36,42 | 39,37 | 42,98 | 45,56 |
| 25 | 10,52 | 11,52 | 13,12 | 14,61 | 16,47 | 19,94 | 24,34 | 29,34 | 34,38 | 37,65 | 40,65 | 44,32 | 46,93 |
| 26 | 11,16 | 12,20 | 13,84 | 15,38 | 17,29 | 20,84 | 25,34 | 30,43 | 35,56 | 38,89 | 41,92 | 45,64 | 48,29 |
| 27 | 11,80 | 12,88 | 14,57 | 16,15 | 18,11 | 21,75 | 26,34 | 31,53 | 36,74 | 40,11 | 43,20 | 46,96 | 49,64 |
| 28 | 12,46 | 13,56 | 15,31 | 16,93 | 18,94 | 22,66 | 27,34 | 32,62 | 37,92 | 41,34 | 44,46 | 48,28 | 50,99 |
| 29 | 13,12 | 14,26 | 16,05 | 17,71 | 19,77 | 23,57 | 28,34 | 33,71 | 39,09 | 42,56 | 45,72 | 49,59 | 52,34 |
| 30 | 13,78 | 14,95 | 16,79 | 18,49 | 20,60 | 24,48 | 29,34 | 34,80 | 40,26 | 43,77 | 46,98 | 50,89 | 53,67 |
| 40 | 20,67 | 22,14 | 24,42 | 26,51 | 29,06 | 33,67 | 39,34 | 45,61 | 51,80 | 55,75 | 59,34 | 63,71 | 66,80 |
| 50 | 27,96 | 29,68 | 32,35 | 34,76 | 37,69 | 42,95 | 49,34 | 56,33 | 63,16 | 67,50 | 71,42 | 76,17 | 79,52 |
| 60 | 35,50 | 37,46 | 40,47 | 43,19 | 46,46 | 52,30 | 59,34 | 66,98 | 74,39 | 79,08 | 83,30 | 88,40 | 91,98 |
| 70 | 43,25 | 45,42 | 48,75 | 51,74 | 55,33 | 61,70 | 69,34 | 77,57 | 85,52 | 90,53 | 95,03 | 100,44 | 104,24 |
| 80 | 51,14 | 53,52 | 57,15 | 60,39 | 64,28 | 71,15 | 79,34 | 88,13 | 96,57 | 101,88 | 106,63 | 112,34 | 116,35 |
| 90 | 59,17 | 61,74 | 65,64 | 69,13 | 73,29 | 80,63 | 89,33 | 98,65 | 107,56 | 113,14 | 118,14 | 124,13 | 128,32 |
| 100 | 67,30 | 70,05 | 74,22 | 77,93 | 82,36 | 90,14 | 99,33 | 109,14 | 118,49 | 124,34 | 129,56 | 135,82 | 140,19 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından hesaplanmıştır.

# Ek F: *F* Dağılımının Kritik Değerleri




## Tablo 1 $\alpha = 0{,}05$

| Payda İçin Serbestlik Derecesi | Pay İçin Serbestlik Derecesi | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | 25 | 30 | 40 | 50 |
| 1 | 161,4 | 199,5 | 215,8 | 224,8 | 230,0 | 233,8 | 236,5 | 238,6 | 240,1 | 242,1 | 245,2 | 248,4 | 248,9 | 250,5 | 250,8 | 252,6 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,35 | 19,37 | 19,38 | 19,40 | 19,43 | 19,44 | 19,46 | 19,47 | 19,48 | 19,48 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,89 | 8,85 | 8,81 | 8,79 | 8,70 | 8,66 | 8,63 | 8,62 | 8,59 | 8,58 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 | 5,86 | 5,80 | 5,77 | 5,75 | 5,72 | 5,70 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,77 | 4,74 | 4,62 | 4,56 | 4,52 | 4,50 | 4,46 | 4,44 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 | 3,94 | 3,87 | 3,83 | 3,81 | 3,77 | 3,75 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,64 | 3,51 | 3,44 | 3,40 | 3,38 | 3,34 | 3,32 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,35 | 3,22 | 3,15 | 3,11 | 3,08 | 3,04 | 3,02 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,14 | 3,01 | 2,94 | 2,89 | 2,86 | 2,83 | 2,80 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 | 2,85 | 2,77 | 2,73 | 2,70 | 2,66 | 2,64 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 | 2,72 | 2,65 | 2,60 | 2,57 | 2,53 | 2,51 |
| 12 | 4,75 | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 | 2,62 | 2,54 | 2,50 | 2,47 | 2,43 | 2,40 |
| 13 | 4,67 | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 | 2,53 | 2,46 | 2,41 | 2,38 | 2,34 | 2,31 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 | 2,46 | 2,39 | 2,34 | 2,31 | 2,27 | 2,24 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 | 2,40 | 2,33 | 2,28 | 2,25 | 2,20 | 2,18 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,35 | 2,28 | 2,23 | 2,19 | 2,15 | 2,12 |
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 | 2,31 | 2,23 | 2,18 | 2,15 | 2,10 | 2,08 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,27 | 2,19 | 2,14 | 2,11 | 2,06 | 2,04 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 | 2,23 | 2,16 | 2,11 | 2,07 | 2,03 | 2,00 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 | 2,20 | 2,12 | 2,07 | 2,04 | 1,99 | 1,97 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 | 2,15 | 2,07 | 2,02 | 1,98 | 1,94 | 1,91 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 | 2,11 | 2,03 | 1,97 | 1,94 | 1,89 | 1,86 |
| 26 | 4,23 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 | 2,07 | 1,99 | 1,94 | 1,90 | 1,85 | 1,82 |
| 28 | 4,20 | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 | 2,04 | 1,96 | 1,91 | 1,87 | 1,82 | 1,79 |
| 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 | 2,01 | 1,93 | 1,88 | 1,84 | 1,79 | 1,76 |
| 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 | 1,92 | 1,84 | 1,78 | 1,74 | 1,69 | 1,66 |
| 50 | 4,03 | 3,18 | 2,79 | 2,56 | 2,40 | 2,29 | 2,20 | 2,13 | 2,07 | 2,03 | 1,87 | 1,78 | 1,73 | 1,69 | 1,63 | 1,60 |
| 60 | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 | 1,84 | 1,75 | 1,69 | 1,65 | 1,59 | 1,56 |
| 120 | 3,92 | 3,07 | 2,68 | 2,45 | 2,29 | 2,18 | 2,09 | 2,02 | 1,96 | 1,91 | 1,75 | 1,66 | 1,60 | 1,55 | 1,50 | 1,46 |
| 200 | 3,89 | 3,04 | 2,65 | 2,42 | 2,26 | 2,14 | 2,06 | 1,98 | 1,93 | 1,88 | 1,72 | 1,62 | 1,56 | 1,52 | 1,46 | 1,41 |
| 500 | 3,86 | 3,01 | 2,62 | 2,39 | 2,23 | 2,12 | 2,03 | 1,96 | 1,90 | 1,85 | 1,69 | 1,59 | 1,53 | 1,48 | 1,42 | 1,38 |
| 1000 | 3,85 | 3,01 | 2,61 | 2,38 | 2,22 | 2,11 | 2,02 | 1,95 | 1,89 | 1,84 | 1,68 | 1,58 | 1,52 | 1,47 | 1,41 | 1,36 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

**Tablo 2** $\alpha = 0{,}025$

| Payda İçin Serbestlik Derecesi | Pay İçin Serbestlik Derecesi | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | 25 | 30 | 40 | 50 |
| 1 | 647,8 | 799,5 | 864,2 | 899,6 | 921,8 | 937,1 | 948,2 | 956,7 | 963,3 | 968,6 | 984,9 | 993,1 | 998,1 | 1001 | 1006 | 1008 |
| 2 | 38,51 | 39,00 | 39,17 | 39,25 | 39,30 | 39,33 | 39,36 | 39,37 | 39,39 | 39,40 | 39,43 | 39,45 | 39,46 | 39,46 | 39,47 | 39,48 |
| 3 | 17,44 | 16,04 | 15,44 | 15,10 | 14,89 | 14,73 | 14,62 | 14,54 | 14,47 | 14,42 | 14,25 | 14,17 | 14,12 | 14,08 | 14,04 | 14,01 |
| 4 | 12,22 | 10,65 | 9,98 | 9,60 | 9,36 | 9,20 | 9,07 | 8,98 | 8,90 | 8,84 | 8,66 | 8,56 | 8,50 | 8,46 | 8,41 | 8,38 |
| 5 | 10,01 | 8,43 | 7,76 | 7,39 | 7,15 | 6,98 | 6,85 | 6,76 | 6,68 | 6,62 | 6,43 | 6,33 | 6,27 | 6,23 | 6,18 | 6,14 |
| 6 | 8,81 | 7,26 | 6,60 | 6,23 | 5,99 | 5,82 | 5,70 | 5,60 | 5,52 | 5,46 | 5,27 | 5,17 | 5,11 | 5,07 | 5,01 | 4,98 |
| 7 | 8,07 | 6,54 | 5,89 | 5,52 | 5,29 | 5,12 | 4,99 | 4,90 | 4,82 | 4,76 | 4,57 | 4,47 | 4,40 | 4,36 | 4,31 | 4,28 |
| 8 | 7,57 | 6,06 | 5,42 | 5,05 | 4,82 | 4,65 | 4,53 | 4,43 | 4,36 | 4,30 | 4,10 | 4,00 | 3,94 | 3,89 | 3,84 | 3,81 |
| 9 | 7,21 | 5,71 | 5,08 | 4,72 | 4,48 | 4,32 | 4,20 | 4,10 | 4,03 | 3,96 | 3,77 | 3,67 | 3,60 | 3,56 | 3,51 | 3,47 |
| 10 | 6,94 | 5,46 | 4,83 | 4,47 | 4,24 | 4,07 | 3,95 | 3,85 | 3,78 | 3,72 | 3,52 | 3,42 | 3,35 | 3,31 | 3,26 | 3,22 |
| 11 | 6,72 | 5,26 | 4,63 | 4,28 | 4,04 | 3,88 | 3,76 | 3,66 | 3,59 | 3,53 | 3,33 | 3,23 | 3,16 | 3,12 | 3,06 | 3,03 |
| 12 | 6,55 | 5,10 | 4,47 | 4,12 | 3,89 | 3,73 | 3,61 | 3,51 | 3,44 | 3,37 | 3,18 | 3,07 | 3,01 | 2,96 | 2,91 | 2,87 |
| 13 | 6,41 | 4,97 | 4,35 | 4,00 | 3,77 | 3,60 | 3,48 | 3,39 | 3,31 | 3,25 | 3,05 | 2,95 | 2,88 | 2,84 | 2,78 | 2,74 |
| 14 | 6,30 | 4,86 | 4,24 | 3,89 | 3,66 | 3,50 | 3,38 | 3,29 | 3,21 | 3,15 | 2,95 | 2,84 | 2,78 | 2,73 | 2,67 | 2,64 |
| 15 | 6,20 | 4,77 | 4,15 | 3,80 | 3,58 | 3,41 | 3,29 | 3,20 | 3,12 | 3,06 | 2,86 | 2,76 | 2,69 | 2,64 | 2,59 | 2,55 |
| 16 | 6,12 | 4,69 | 4,08 | 3,73 | 3,50 | 3,34 | 3,22 | 3,12 | 3,05 | 2,99 | 2,79 | 2,68 | 2,61 | 2,57 | 2,51 | 2,47 |
| 17 | 6,04 | 4,62 | 4,01 | 3,66 | 3,44 | 3,28 | 3,16 | 3,06 | 2,98 | 2,92 | 2,72 | 2,62 | 2,55 | 2,50 | 2,44 | 2,41 |
| 18 | 5,98 | 4,56 | 3,95 | 3,61 | 3,38 | 3,22 | 3,10 | 3,01 | 2,93 | 2,87 | 2,67 | 2,56 | 2,49 | 2,44 | 2,38 | 2,35 |
| 19 | 5,92 | 4,51 | 3,90 | 3,56 | 3,33 | 3,17 | 3,05 | 2,96 | 2,88 | 2,82 | 2,62 | 2,51 | 2,44 | 2,39 | 2,33 | 2,30 |
| 20 | 5,87 | 4,46 | 3,86 | 3,51 | 3,29 | 3,13 | 3,01 | 2,91 | 2,84 | 2,77 | 2,57 | 2,46 | 2,40 | 2,35 | 2,29 | 2,25 |
| 22 | 5,79 | 4,38 | 3,78 | 3,44 | 3,22 | 3,05 | 2,93 | 2,84 | 2,76 | 2,70 | 2,50 | 2,39 | 2,32 | 2,27 | 2,21 | 2,17 |
| 24 | 5,72 | 4,32 | 3,72 | 3,38 | 3,15 | 2,99 | 2,87 | 2,78 | 2,70 | 2,64 | 2,44 | 2,33 | 2,26 | 2,21 | 2,15 | 2,11 |
| 26 | 5,66 | 4,27 | 3,67 | 3,33 | 3,10 | 2,94 | 2,82 | 2,73 | 2,65 | 2,59 | 2,39 | 2,28 | 2,21 | 2,16 | 2,09 | 2,05 |
| 28 | 5,61 | 4,22 | 3,63 | 3,29 | 3,06 | 2,90 | 2,78 | 2,69 | 2,61 | 2,55 | 2,34 | 2,23 | 2,16 | 2,11 | 2,05 | 2,01 |
| 30 | 5,57 | 4,18 | 3,59 | 3,25 | 3,03 | 2,87 | 2,75 | 2,65 | 2,57 | 2,51 | 2,31 | 2,20 | 2,12 | 2,07 | 2,01 | 1,97 |
| 40 | 5,42 | 4,05 | 3,46 | 3,13 | 2,90 | 2,74 | 2,62 | 2,53 | 2,45 | 2,39 | 2,18 | 2,07 | 1,99 | 1,94 | 1,88 | 1,83 |
| 50 | 5,34 | 3,97 | 3,39 | 3,05 | 2,83 | 2,67 | 2,55 | 2,46 | 2,38 | 2,32 | 2,11 | 1,99 | 1,92 | 1,87 | 1,80 | 1,75 |
| 60 | 5,29 | 3,93 | 3,34 | 3,01 | 2,79 | 2,63 | 2,51 | 2,41 | 2,33 | 2,27 | 2,06 | 1,94 | 1,87 | 1,82 | 1,74 | 1,70 |
| 120 | 5,15 | 3,80 | 3,23 | 2,89 | 2,67 | 2,52 | 2,39 | 2,30 | 2,22 | 2,16 | 1,94 | 1,82 | 1,75 | 1,69 | 1,61 | 1,56 |
| 200 | 5,10 | 3,76 | 3,18 | 2,85 | 2,63 | 2,47 | 2,35 | 2,26 | 2,18 | 2,11 | 1,90 | 1,78 | 1,70 | 1,64 | 1,56 | 1,51 |
| 500 | 5,05 | 3,72 | 3,14 | 2,81 | 2,59 | 2,43 | 2,31 | 2,22 | 2,14 | 2,07 | 1,86 | 1,74 | 1,65 | 1,60 | 1,52 | 1,46 |
| 1000 | 5,04 | 3,70 | 3,13 | 2,80 | 2,58 | 2,42 | 2,30 | 2,20 | 2,13 | 2,06 | 1,85 | 1,72 | 1,64 | 1,58 | 1,50 | 1,45 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Tablo 3 $\alpha = 0{,}01$

| Payda İçin Serbestlik Derecesi | Pay İçin Serbestlik Derecesi | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | 25 | 30 | 40 | 50 |
| 1 | 4048 | 4993 | 5377 | 5577 | 5668 | 5924 | 5992 | 6096 | 6132 | 6168 | 6079 | 6168 | 6214 | 6355 | 6168 | 6213 |
| 2 | 98,50 | 99,01 | 99,15 | 99,23 | 99,30 | 99,33 | 99,35 | 99,39 | 99,40 | 99,43 | 99,38 | 99,48 | 99,43 | 99,37 | 99,44 | 99,59 |
| 3 | 34,12 | 30,82 | 29,46 | 28,71 | 28,24 | 27,91 | 27,67 | 27,49 | 27,34 | 27,23 | 26,87 | 26,69 | 26,58 | 26,51 | 26,41 | 26,36 |
| 4 | 21,20 | 18,00 | 16,69 | 15,98 | 15,52 | 15,21 | 14,98 | 14,80 | 14,66 | 14,55 | 14,20 | 14,02 | 13,91 | 13,84 | 13,75 | 13,69 |
| 5 | 16,26 | 13,27 | 12,06 | 11,39 | 10,97 | 10,67 | 10,46 | 10,29 | 10,16 | 10,05 | 9,72 | 9,55 | 9,45 | 9,38 | 9,29 | 9,24 |
| 6 | 13,75 | 10,92 | 9,78 | 9,15 | 8,75 | 8,47 | 8,26 | 8,10 | 7,98 | 7,87 | 7,56 | 7,40 | 7,30 | 7,23 | 7,14 | 7,09 |
| 7 | 12,25 | 9,55 | 8,45 | 7,85 | 7,46 | 7,19 | 6,99 | 6,84 | 6,72 | 6,62 | 6,31 | 6,16 | 6,06 | 5,99 | 5,91 | 5,86 |
| 8 | 11,26 | 8,65 | 7,59 | 7,01 | 6,63 | 6,37 | 6,18 | 6,03 | 5,91 | 5,81 | 5,52 | 5,36 | 5,26 | 5,20 | 5,12 | 5,07 |
| 9 | 10,56 | 8,02 | 6,99 | 6,42 | 6,06 | 5,80 | 5,61 | 5,47 | 5,35 | 5,26 | 4,96 | 4,81 | 4,71 | 4,65 | 4,57 | 4,52 |
| 10 | 10,04 | 7,56 | 6,55 | 5,99 | 5,64 | 5,39 | 5,20 | 5,06 | 4,94 | 4,85 | 4,56 | 4,41 | 4,31 | 4,25 | 4,17 | 4,12 |
| 11 | 9,65 | 7,21 | 6,22 | 5,67 | 5,32 | 5,07 | 4,89 | 4,74 | 4,63 | 4,54 | 4,25 | 4,10 | 4,01 | 3,94 | 3,86 | 3,81 |
| 12 | 9,33 | 6,93 | 5,95 | 5,41 | 5,06 | 4,82 | 4,64 | 4,50 | 4,39 | 4,30 | 4,01 | 3,86 | 3,76 | 3,70 | 3,62 | 3,57 |
| 13 | 9,07 | 6,70 | 5,74 | 5,21 | 4,86 | 4,62 | 4,44 | 4,30 | 4,19 | 4,10 | 3,82 | 3,66 | 3,57 | 3,51 | 3,43 | 3,38 |
| 14 | 8,86 | 6,51 | 5,56 | 5,04 | 4,69 | 4,46 | 4,28 | 4,14 | 4,03 | 3,94 | 3,66 | 3,51 | 3,41 | 3,35 | 3,27 | 3,22 |
| 15 | 8,68 | 6,36 | 5,42 | 4,89 | 4,56 | 4,32 | 4,14 | 4,00 | 3,89 | 3,80 | 3,52 | 3,37 | 3,28 | 3,21 | 3,13 | 3,08 |
| 16 | 8,53 | 6,23 | 5,29 | 4,77 | 4,44 | 4,20 | 4,03 | 3,89 | 3,78 | 3,69 | 3,41 | 3,26 | 3,16 | 3,10 | 3,02 | 2,97 |
| 17 | 8,40 | 6,11 | 5,18 | 4,67 | 4,34 | 4,10 | 3,93 | 3,79 | 3,68 | 3,59 | 3,31 | 3,16 | 3,07 | 3,00 | 2,92 | 2,87 |
| 18 | 8,29 | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,84 | 3,71 | 3,60 | 3,51 | 3,23 | 3,08 | 2,98 | 2,92 | 2,84 | 2,78 |
| 19 | 8,18 | 5,93 | 5,01 | 4,50 | 4,17 | 3,94 | 3,77 | 3,63 | 3,52 | 3,43 | 3,15 | 3,00 | 2,91 | 2,84 | 2,76 | 2,71 |
| 20 | 8,10 | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,70 | 3,56 | 3,46 | 3,37 | 3,09 | 2,94 | 2,84 | 2,78 | 2,69 | 2,64 |
| 22 | 7,95 | 5,72 | 4,82 | 4,31 | 3,99 | 3,76 | 3,59 | 3,45 | 3,35 | 3,26 | 2,98 | 2,83 | 2,73 | 2,67 | 2,58 | 2,53 |
| 24 | 7,82 | 5,61 | 4,72 | 4,22 | 3,90 | 3,67 | 3,50 | 3,36 | 3,26 | 3,17 | 2,89 | 2,74 | 2,64 | 2,58 | 2,49 | 2,44 |
| 26 | 7,72 | 5,53 | 4,64 | 4,14 | 3,82 | 3,59 | 3,42 | 3,29 | 3,18 | 3,09 | 2,81 | 2,66 | 2,57 | 2,50 | 2,42 | 2,36 |
| 28 | 7,64 | 5,45 | 4,57 | 4,07 | 3,75 | 3,53 | 3,36 | 3,23 | 3,12 | 3,03 | 2,75 | 2,60 | 2,51 | 2,44 | 2,35 | 2,30 |
| 30 | 7,56 | 5,39 | 4,51 | 4,02 | 3,70 | 3,47 | 3,30 | 3,17 | 3,07 | 2,98 | 2,70 | 2,55 | 2,45 | 2,39 | 2,30 | 2,25 |
| 40 | 7,31 | 5,18 | 4,31 | 3,83 | 3,51 | 3,29 | 3,12 | 2,99 | 2,89 | 2,80 | 2,52 | 2,37 | 2,27 | 2,20 | 2,11 | 2,06 |
| 50 | 7,17 | 5,06 | 4,20 | 3,72 | 3,41 | 3,19 | 3,02 | 2,89 | 2,78 | 2,70 | 2,42 | 2,27 | 2,17 | 2,10 | 2,01 | 1,95 |
| 60 | 7,08 | 4,98 | 4,13 | 3,65 | 3,34 | 3,12 | 2,95 | 2,82 | 2,72 | 2,63 | 2,35 | 2,20 | 2,10 | 2,03 | 1,94 | 1,88 |
| 120 | 6,85 | 4,79 | 3,95 | 3,48 | 3,17 | 2,96 | 2,79 | 2,66 | 2,56 | 2,47 | 2,19 | 2,03 | 1,93 | 1,86 | 1,76 | 1,70 |
| 200 | 6,76 | 4,71 | 3,88 | 3,41 | 3,11 | 2,89 | 2,73 | 2,60 | 2,50 | 2,41 | 2,13 | 1,97 | 1,87 | 1,79 | 1,69 | 1,63 |
| 500 | 6,69 | 4,65 | 3,82 | 3,36 | 3,05 | 2,84 | 2,68 | 2,55 | 2,44 | 2,36 | 2,07 | 1,92 | 1,81 | 1,74 | 1,63 | 1,57 |
| 1000 | 6,67 | 4,63 | 3,80 | 3,34 | 3,04 | 2,82 | 2,66 | 2,53 | 2,43 | 2,34 | 2,06 | 1,90 | 1,79 | 1,72 | 1,61 | 1,54 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

# Ek *ncF*: Merkezî Olmayan *F* Dağılımının Kritik Değerleri

## Güç = 1 – (Tablo Girişi)

| | | $\psi$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0,50 | 1,0 | 1,2 | 1,4 | 1,6 | 1,8 | 2,0 | 2,2 | 2,6 | 3,0 |
| $sd_e$ | α | $sd_t = 1$ | | | | | | | | | |
| 2 | 0,05 | 0,93 | 0,86 | 0,83 | 0,78 | 0,74 | 0,69 | 0,64 | 0,59 | 0,49 | 0,40 |
| | 0,01 | 0,99 | 0,97 | 0,96 | 0,95 | 0,94 | 0,93 | 0,91 | 0,90 | 0,87 | 0,83 |
| 4 | 0,05 | 0,91 | 0,80 | 0,74 | 0,67 | 0,59 | 0,51 | 0,43 | 0,35 | 0,22 | 0,12 |
| | 0,01 | 0,98 | 0,95 | 0,93 | 0,90 | 0,87 | 0,83 | 0,78 | 0,73 | 0,62 | 0,50 |
| 6 | 0,05 | 0,91 | 0,78 | 0,70 | 0,62 | 0,52 | 0,43 | 0,34 | 0,26 | 0,14 | 0,06 |
| | 0,01 | 0,98 | 0,93 | 0,90 | 0,86 | 0,81 | 0,75 | 0,69 | 0,61 | 0,46 | 0,31 |
| 8 | 0,05 | 0,90 | 0,76 | 0,68 | 0,59 | 0,49 | 0,39 | 0,30 | 0,22 | 0,11 | 0,04 |
| | 0,01 | 0,98 | 0,92 | 0,89 | 0,84 | 0,78 | 0,70 | 0,62 | 0,54 | 0,37 | 0,22 |
| 10 | 0,05 | 0,90 | 0,75 | 0,66 | 0,57 | 0,47 | 0,37 | 0,28 | 0,20 | 0,09 | 0,03 |
| | 0,01 | 0,98 | 0,92 | 0,87 | 0,82 | 0,75 | 0,67 | 0,58 | 0,49 | 0,31 | 0,17 |
| 12 | 0,05 | 0,90 | 0,74 | 0,65 | 0,56 | 0,45 | 0,35 | 0,26 | 0,19 | 0,08 | 0,03 |
| | 0,01 | 0,97 | 0,91 | 0,87 | 0,81 | 0,73 | 0,65 | 0,55 | 0,46 | 0,28 | 0,14 |
| 16 | 0,05 | 0,90 | 0,74 | 0,64 | 0,54 | 0,43 | 0,33 | 0,24 | 0,17 | 0,07 | 0,02 |
| | 0,01 | 0,97 | 0,90 | 0,85 | 0,79 | 0,71 | 0,61 | 0,52 | 0,42 | 0,24 | 0,11 |
| 20 | 0,05 | 0,90 | 0,73 | 0,63 | 0,53 | 0,42 | 0,32 | 0,23 | 0,16 | 0,06 | 0,02 |
| | 0,01 | 0,97 | 0,90 | 0,85 | 0,78 | 0,69 | 0,59 | 0,49 | 0,39 | 0,21 | 0,10 |
| 30 | 0,05 | 0,89 | 0,72 | 0,62 | 0,52 | 0,40 | 0,31 | 0,22 | 0,15 | 0,06 | 0,02 |
| | 0,01 | 0,97 | 0,89 | 0,83 | 0,76 | 0,67 | 0,57 | 0,46 | 0,36 | 0,19 | 0,08 |
| ∞ | 0,05 | 0,89 | 0,71 | 0,60 | 0,49 | 0,38 | 0,28 | 0,19 | 0,12 | 0,04 | 0,01 |
| | 0,01 | 0,97 | 0,88 | 0,81 | 0,72 | 0,62 | 0,51 | 0,40 | 0,30 | 0,14 | 0,05 |
| $sd_e$ | α | $sd_t = 1$ | | | | | | | | | |
| 2 | 0,05 | 0,93 | 0,88 | 0,85 | 0,82 | 0,78 | 0,75 | 0,70 | 0,66 | 0,56 | 0,48 |
| | 0,01 | 0,99 | 0,98 | 0,97 | 0,96 | 0,95 | 0,94 | 0,93 | 0,92 | 0,89 | 0,86 |
| 4 | 0,05 | 0,92 | 0,82 | 0,77 | 0,70 | 0,62 | 0,54 | 0,46 | 0,38 | 0,24 | 0,14 |
| | 0,01 | 0,98 | 0,96 | 0,94 | 0,92 | 0,89 | 0,85 | 0,81 | 0,76 | 0,66 | 0,54 |
| 6 | 0,05 | 0,91 | 0,79 | 0,71 | 0,63 | 0,53 | 0,43 | 0,34 | 0,26 | 0,13 | 0,05 |
| | 0,01 | 0,98 | 0,94 | 0,91 | 0,87 | 0,82 | 0,76 | 0,70 | 0,62 | 0,46 | 0,31 |
| 8 | 0,05 | 0,91 | 0,77 | 0,68 | 0,58 | 0,48 | 0,37 | 0,28 | 0,20 | 0,08 | 0,03 |
| | 0,01 | 0,98 | 0,93 | 0,89 | 0,84 | 0,78 | 0,70 | 0,61 | 0,52 | 0,34 | 0,19 |
| 10 | 0,05 | 0,91 | 0,75 | 0,66 | 0,55 | 0,44 | 0,34 | 0,24 | 0,16 | 0,06 | 0,02 |
| | 0,01 | 0,98 | 0,92 | 0,88 | 0,82 | 0,74 | 0,65 | 0,55 | 0,45 | 0,26 | 0,13 |
| 12 | 0,05 | 0,90 | 0,74 | 0,64 | 0,53 | 0,42 | 0,31 | 0,22 | 0,14 | 0,05 | 0,01 |
| | 0,01 | 0,98 | 0,91 | 0,86 | 0,80 | 0,71 | 0,61 | 0,51 | 0,40 | 0,22 | 0,09 |
| 16 | 0,05 | 0,90 | 0,73 | 0,62 | 0,51 | 0,39 | 0,28 | 0,19 | 0,12 | 0,04 | 0,01 |
| | 0,01 | 0,97 | 0,90 | 0,84 | 0,77 | 0,67 | 0,57 | 0,45 | 0,34 | 0,16 | 0,06 |
| 20 | 0,05 | 0,90 | 0,72 | 0,61 | 0,49 | 0,36 | 0,26 | 0,17 | 0,11 | 0,03 | 0,01 |
| | 0,01 | 0,97 | 0,90 | 0,83 | 0,75 | 0,65 | 0,53 | 0,42 | 0,31 | 0,14 | 0,04 |
| 30 | 0,05 | 0,90 | 0,71 | 0,59 | 0,47 | 0,35 | 0,24 | 0,15 | 0,09 | 0,02 | 0,00 |
| | 0,01 | 0,97 | 0,88 | 0,82 | 0,72 | 0,61 | 0,49 | 0,37 | 0,26 | 0,10 | 0,03 |
| ∞ | 0,05 | 0,89 | 0,68 | 0,56 | 0,43 | 0,30 | 0,20 | 0,12 | 0,06 | 0,01 | 0,00 |
| | 0,01 | 0,97 | 0,86 | 0,77 | 0,66 | 0,53 | 0,40 | 0,28 | 0,18 | 0,05 | 0,01 |

*(devam ediyor)*

## Ek *ncF* *(devamı)*

| | | $\phi$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0,50 | 1,0 | 1,2 | 1,4 | 1,6 | 1,8 | 2,0 | 2,2 | 2,6 | 3,0 |
| $sd_e$ | α | $sd_t = 3$ | | | | | | | | | |
| 2 | 0,05 | 0,93 | 0,89 | 0,86 | 0,83 | 0,80 | 0,76 | 0,73 | 0,69 | 0,60 | 0,52 |
| | 0,01 | 0,99 | 0,98 | 0,97 | 0,96 | 0,96 | 0,95 | 0,94 | 0,93 | 0,90 | 0,88 |
| 4 | 0,05 | 0,92 | 0,83 | 0,77 | 0,71 | 0,63 | 0,55 | 0,47 | 0,39 | 0,25 | 0,14 |
| | 0,01 | 0,98 | 0,96 | 0,94 | 0,92 | 0,89 | 0,86 | 0,82 | 0,77 | 0,67 | 0,55 |
| 6 | 0,05 | 0,91 | 0,79 | 0,71 | 0,62 | 0,52 | 0,42 | 0,33 | 0,24 | 0,11 | 0,04 |
| | 0,01 | 0,98 | 0,94 | 0,91 | 0,87 | 0,82 | 0,76 | 0,69 | 0,61 | 0,44 | 0,29 |
| 8 | 0,05 | 0,91 | 0,76 | 0,67 | 0,57 | 0,46 | 0,35 | 0,25 | 0,17 | 0,06 | 0,02 |
| | 0,01 | 0,98 | 0,93 | 0,89 | 0,84 | 0,77 | 0,68 | 0,59 | 0,49 | 0,30 | 0,16 |
| 10 | 0,05 | 0,91 | 0,75 | 0,65 | 0,53 | 0,41 | 0,30 | 0,21 | 0,13 | 0,04 | 0,01 |
| | 0,01 | 0,98 | 0,92 | 0,87 | 0,80 | 0,72 | 0,62 | 0,52 | 0,41 | 0,22 | 0,09 |
| 12 | 0,05 | 0,90 | 0,73 | 0,62 | 0,50 | 0,38 | 0,27 | 0,18 | 0,11 | 0,03 | 0,01 |
| | 0,01 | 0,98 | 0,91 | 0,85 | 0,78 | 0,69 | 0,58 | 0,46 | 0,35 | 0,17 | 0,06 |
| 16 | 0,05 | 0,90 | 0,71 | 0,60 | 0,47 | 0,34 | 0,23 | 0,14 | 0,08 | 0,02 | 0,00 |
| | 0,01 | 0,97 | 0,90 | 0,83 | 0,74 | 0,64 | 0,51 | 0,39 | 0,28 | 0,11 | 0,03 |
| 20 | 0,05 | 0,90 | 0,70 | 0,58 | 0,45 | 0,32 | 0,21 | 0,13 | 0,07 | 0,01 | 0,00 |
| | 0,01 | 0,97 | 0,89 | 0,82 | 0,72 | 0,60 | 0,47 | 0,35 | 0,24 | 0,08 | 0,02 |
| 30 | 0,05 | 0,89 | 0,68 | 0,55 | 0,42 | 0,29 | 0,18 | 0,10 | 0,05 | 0,01 | 0,00 |
| | 0,01 | 0,97 | 0,87 | 0,79 | 0,68 | 0,55 | 0,42 | 0,29 | 0,18 | 0,05 | 0,01 |
| ∞ | 0,05 | 0,88 | 0,64 | 0,50 | 0,36 | 0,23 | 0,13 | 0,07 | 0,03 | 0,00 | 0,00 |
| | 0,01 | 0,97 | 0,84 | 0,73 | 0,59 | 0,44 | 0,30 | 0,18 | 0,10 | 0,02 | 0,00 |
| $sd_e$ | α | $sd_t = 4$ | | | | | | | | | |
| 2 | 0,05 | 0,94 | 0,89 | 0,87 | 0,84 | 0,81 | 0,77 | 0,74 | 0,70 | 0,62 | 0,54 |
| | 0,01 | 0,99 | 0,98 | 0,97 | 0,97 | 0,96 | 0,95 | 0,94 | 0,93 | 0,91 | 0,88 |
| 4 | 0,05 | 0,92 | 0,83 | 0,78 | 0,71 | 0,64 | 0,55 | 0,47 | 0,39 | 0,25 | 0,14 |
| | 0,01 | 0,98 | 0,96 | 0,94 | 0,92 | 0,89 | 0,86 | 0,82 | 0,78 | 0,67 | 0,56 |
| 6 | 0,05 | 0,92 | 0,79 | 0,71 | 0,62 | 0,52 | 0,41 | 0,31 | 0,23 | 0,10 | 0,04 |
| | 0,01 | 0,98 | 0,94 | 0,91 | 0,87 | 0,82 | 0,76 | 0,68 | 0,60 | 0,43 | 0,28 |
| 8 | 0,05 | 0,91 | 0,76 | 0,66 | 0,55 | 0,44 | 0,33 | 0,23 | 0,15 | 0,05 | 0,01 |
| | 0,01 | 0,98 | 0,93 | 0,89 | 0,83 | 0,76 | 0,67 | 0,57 | 0,47 | 0,28 | 0,14 |
| 10 | 0,05 | 0,91 | 0,74 | 0,63 | 0,51 | 0,39 | 0,27 | 0,18 | 0,11 | 0,03 | 0,01 |
| | 0,01 | 0,98 | 0,92 | 0,86 | 0,79 | 0,70 | 0,60 | 0,49 | 0,37 | 0,19 | 0,07 |
| 12 | 0,05 | 0,90 | 0,72 | 0,61 | 0,48 | 0,35 | 0,24 | 0,15 | 0,08 | 0,02 | 0,00 |
| | 0,01 | 0,98 | 0,91 | 0,85 | 0,76 | 0,66 | 0,55 | 0,42 | 0,31 | 0,13 | 0,04 |
| 16 | 0,05 | 0,90 | 0,70 | 0,57 | 0,44 | 0,31 | 0,19 | 0,11 | 0,06 | 0,01 | 0,00 |
| | 0,01 | 0,97 | 0,89 | 0,82 | 0,72 | 0,60 | 0,47 | 0,34 | 0,23 | 0,08 | 0,02 |
| 20 | 0,05 | 0,89 | 0,68 | 0,55 | 0,41 | 0,28 | 0,17 | 0,09 | 0,04 | 0,01 | 0,00 |
| | 0,01 | 0,97 | 0,88 | 0,80 | 0,69 | 0,56 | 0,42 | 0,29 | 0,18 | 0,05 | 0,01 |
| 30 | 0,05 | 0,89 | 0,66 | 0,52 | 0,37 | 0,24 | 0,14 | 0,07 | 0,03 | 0,00 | 0,00 |
| | 0,01 | 0,97 | 0,86 | 0,77 | 0,64 | 0,50 | 0,35 | 0,22 | 0,13 | 0,03 | 0,00 |
| ∞ | 0,05 | 0,88 | 0,60 | 0,45 | 0,29 | 0,17 | 0,08 | 0,04 | 0,01 | 0,00 | 0,00 |
| | 0,01 | 0,96 | 0,81 | 0,68 | 0,53 | 0,36 | 0,22 | 0,11 | 0,05 | 0,01 | 0,00 |

*Kaynak:* M. L. Tiku (1967), F testinin tabloları, *Journal of the American Statistical Association,* 62, 525-539'dan yazarın ve editörlerin izni ile kısaltılarak alınmıştır.

## Ek Polinom: Dik Polinom Katsayıları

| $k$ | Polynomial | $X = 1$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | $\Sigma a_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Linear | – 1 | 0 | 1 | | | | | | | | 2 |
| | Quadratic | 1 | – 2 | 1 | | | | | | | | 6 |
| 4 | Linear | – 3 | – 1 | 1 | 3 | | | | | | | 20 |
| | Quadratic | 1 | – 1 | – 1 | 1 | | | | | | | 4 |
| | Cubic | – 1 | 3 | – 3 | 1 | | | | | | | 20 |
| 5 | Linear | – 2 | – 1 | 0 | 1 | 2 | | | | | | 10 |
| | Quadratic | 2 | – 1 | – 2 | – 1 | 2 | | | | | | 14 |
| | Cubic | – 1 | 2 | 0 | – 2 | 1 | | | | | | 10 |
| | Quartic | 1 | – 4 | 6 | – 4 | 1 | | | | | | 70 |
| 6 | Linear | – 5 | – 3 | – 1 | 1 | 3 | 5 | | | | | 70 |
| | Quadratic | 5 | – 1 | – 4 | – 4 | – 1 | 5 | | | | | 84 |
| | Cubic | – 5 | 7 | 4 | – 4 | – 7 | 5 | | | | | 180 |
| | Quartic | 1 | – 3 | 2 | 2 | – 3 | 1 | | | | | 28 |
| 7 | Linear | – 3 | – 2 | – 1 | 0 | – 1 | 2 | 3 | | | | 28 |
| | Quadratic | 5 | 0 | – 3 | – 4 | – 3 | 0 | 5 | | | | 84 |
| | Cubic | – 1 | 1 | 1 | 0 | 1 | – 1 | 1 | | | | 6 |
| | Quartic | 3 | – 7 | 1 | 6 | 1 | – 7 | 3 | | | | 154 |
| 8 | Linear | – 7 | – 5 | – 3 | – 1 | 1 | 3 | 5 | 7 | | | 168 |
| | Quadratic | 7 | 1 | – 3 | – 5 | – 5 | – 3 | 1 | 7 | | | 168 |
| | Cubic | – 7 | 5 | 7 | 3 | – 3 | – 7 | – 5 | 7 | | | 264 |
| | Quartic | 7 | – 13 | – 3 | 9 | 9 | – 3 | – 13 | 7 | | | 616 |
| | Quintic | – 7 | 23 | – 17 | – 15 | 15 | 17 | – 23 | 7 | | | 2184 |
| 9 | Linear | – 4 | – 3 | – 2 | – 1 | 0 | 1 | 2 | 3 | 4 | | 60 |
| | Quadratic | 28 | 7 | – 8 | – 17 | – 20 | – 17 | – 8 | 7 | 28 | | 2772 |
| | Cubic | – 14 | 7 | 13 | 9 | 0 | – 9 | – 13 | – 7 | 14 | | 990 |
| | Quartic | 14 | – 21 | – 11 | 9 | 18 | 9 | – 11 | – 21 | 14 | | 2002 |
| | Quintic | – 4 | 11 | – 4 | – 9 | 0 | 9 | 4 | – 11 | 4 | | 468 |
| 10 | Linear | – 9 | – 7 | – 5 | – 3 | – 1 | 1 | 3 | 5 | 7 | 9 | 330 |
| | Quadratic | 6 | 2 | – 1 | – 3 | – 4 | – 4 | – 3 | – 1 | 2 | 6 | 132 |
| | Cubic | – 42 | 14 | 35 | 31 | 12 | – 12 | – 31 | – 35 | – 14 | 42 | 8580 |
| | Quartic | 18 | – 22 | – 17 | 3 | 18 | 18 | 3 | – 17 | – 22 | 18 | 2860 |
| | Quintic | – 6 | 14 | – 1 | – 11 | – 6 | 6 | 11 | 1 | – 14 | 6 | 780 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Ek Güç: $\delta$'nın Bir Fonksiyonu Olarak Güç ve Manidarlık Düzeyi ($\alpha$)

**İki Kuyruklu Test İçin**

| δ | *0,10* | *0,05* | *0,02* | *0,01* |
|---|---|---|---|---|
| 1,00 | 0,26 | 0,17 | 0,09 | 0,06 |
| 1,10 | 0,29 | 0,20 | 0,11 | 0,07 |
| 1,20 | 0,33 | 0,22 | 0,13 | 0,08 |
| 1,30 | 0,37 | 0,26 | 0,15 | 0,10 |
| 1,40 | 0,40 | 0,29 | 0,18 | 0,12 |
| 1,50 | 0,44 | 0,32 | 0,20 | 0,14 |
| 1,60 | 0,48 | 0,36 | 0,23 | 0,17 |
| 1,70 | 0,52 | 0,40 | 0,27 | 0,19 |
| 1,80 | 0,56 | 0,44 | 0,30 | 0,22 |
| 1,90 | 0,60 | 0,48 | 0,34 | 0,25 |
| 2,00 | 0,64 | 0,52 | 0,37 | 0,28 |
| 2,10 | 0,68 | 0,56 | 0,41 | 0,32 |
| 2,20 | 0,71 | 0,60 | 0,45 | 0,35 |
| 2,30 | 0,74 | 0,63 | 0,49 | 0,39 |
| 2,40 | 0,78 | 0,67 | 0,53 | 0,43 |
| 2,50 | 0,80 | 0,71 | 0,57 | 0,47 |
| 2,60 | 0,83 | 0,74 | 0,61 | 0,51 |
| 2,70 | 0,85 | 0,77 | 0,65 | 0,55 |
| 2,80 | 0,88 | 0,80 | 0,68 | 0,59 |
| 2,90 | 0,90 | 0,83 | 0,72 | 0,63 |
| 3,00 | 0,91 | 0,85 | 0,75 | 0,66 |
| 3,10 | 0,93 | 0,87 | 0,78 | 0,70 |
| 3,20 | 0,94 | 0,89 | 0,81 | 0,73 |
| 3,30 | 0,95 | 0,91 | 0,84 | 0,77 |
| 3,40 | 0,96 | 0,93 | 0,86 | 0,80 |
| 3,50 | 0,97 | 0,94 | 0,88 | 0,82 |
| 3,60 | 0,98 | 0,95 | 0,90 | 0,85 |
| 3,70 | 0,98 | 0,96 | 0,92 | 0,87 |
| 3,80 | 0,98 | 0,97 | 0,93 | 0,89 |
| 3,90 | 0,99 | 0,97 | 0,94 | 0,91 |
| 4,00 | 0,99 | 0,98 | 0,95 | 0,92 |
| 4,10 | 0,99 | 0,98 | 0,96 | 0,94 |
| 4,20 | — | 0,99 | 0,97 | 0,95 |
| 4,30 | — | 0,99 | 0,98 | 0,96 |
| 4,40 | — | 0,99 | 0,98 | 0,97 |
| 4,50 | — | 0,99 | 0,99 | 0,97 |
| 4,60 | — | — | 0,99 | 0,98 |
| 4,70 | — | — | 0,99 | 0,98 |
| 4,80 | — | — | 0,99 | 0,99 |
| 4,90 | — | — | — | 0,99 |
| 5,00 | — | — | — | 0,99 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Ek*q*: Standartlaştırılmış Ranj İstatistiğinin Kritik Değerleri (*q*)

**Tablo 1** $\alpha = 0{,}05$

| *Hata sd* | r = Standartlaştırılmış Ortalamalar Arasındaki Adımların Sayısı | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* |
| 1 | 17,97 | 26,98 | 32,82 | 37,08 | 40,41 | 43,12 | 45,40 | 47,36 | 49,07 | 50,59 | 51,96 | 53,20 | 54,33 | 55,36 |
| 2 | 6,08 | 8,33 | 9,80 | 10,88 | 11,74 | 12,44 | 13,03 | 13,54 | 13,99 | 14,39 | 14,75 | 15,08 | 15,38 | 15,65 |
| 3 | 4,50 | 5,91 | 6,82 | 7,50 | 8,04 | 8,48 | 8,85 | 9,18 | 9,46 | 9,72 | 9,95 | 10,15 | 10,35 | 10,53 |
| 4 | 3,93 | 5,04 | 5,76 | 6,29 | 6,71 | 7,05 | 7,35 | 7,60 | 7,33 | 8,03 | 8,21 | 8,37 | 8,52 | 8,66 |
| 5 | 3,64 | 4,60 | 5,22 | 5,67 | 6,03 | 6,33 | 6,58 | 6,80 | 7,00 | 7,17 | 7,32 | 7,47 | 7,60 | 7,72 |
| 6 | 3,46 | 4,34 | 4,90 | 5,31 | 5,63 | 5,90 | 6,12 | 6,32 | 6,49 | 6,65 | 6,79 | 6,92 | 7,03 | 7,14 |
| 7 | 3,34 | 4,16 | 4,68 | 5,06 | 5,36 | 5,61 | 5,82 | 6,00 | 6,16 | 6,30 | 6,43 | 6,55 | 6,66 | 6,76 |
| 8 | 3,26 | 4,04 | 4,53 | 4,89 | 5,17 | 5,40 | 5,60 | 5,77 | 5,92 | 6,05 | 6,18 | 6,29 | 6,39 | 6,48 |
| 9 | 3,20 | 3,95 | 4,42 | 4,76 | 5,02 | 5,24 | 5,43 | 5,60 | 5,74 | 5,87 | 5,98 | 6,09 | 6,19 | 6,28 |
| 10 | 3,15 | 3,88 | 4,33 | 4,65 | 4,91 | 5,12 | 5,30 | 5,46 | 5,60 | 5,72 | 5,83 | 5,94 | 6,03 | 6,11 |
| 11 | 3,11 | 3,82 | 4,26 | 4,57 | 4,82 | 5,03 | 5,20 | 5,35 | 5,49 | 5,60 | 5,71 | 5,81 | 5,90 | 5,98 |
| 12 | 3,08 | 3,77 | 4,20 | 4,51 | 4,75 | 4,95 | 5,12 | 5,26 | 5,40 | 5,51 | 5,62 | 5,71 | 5,79 | 5,88 |
| 13 | 3,06 | 3,74 | 4,15 | 4,45 | 4,69 | 4,88 | 5,05 | 5,19 | 5,32 | 5,43 | 5,53 | 5,63 | 5,71 | 5,79 |
| 14 | 3,03 | 3,70 | 4,11 | 4,41 | 4,64 | 4,83 | 4,99 | 5,13 | 5,25 | 5,36 | 5,46 | 5,55 | 5,64 | 5,71 |
| 15 | 3,01 | 3,67 | 4,08 | 4,37 | 4,60 | 4,78 | 4,94 | 5,08 | 5,20 | 5,31 | 5,40 | 5,49 | 5,57 | 5,65 |
| 16 | 3,00 | 3,65 | 4,05 | 4,33 | 4,56 | 4,74 | 4,90 | 5,03 | 5,15 | 5,26 | 5,35 | 5,44 | 5,52 | 5,59 |
| 17 | 2,98 | 3,63 | 4,02 | 4,30 | 4,52 | 4,70 | 4,86 | 4,99 | 5,11 | 5,21 | 5,31 | 5,39 | 5,47 | 5,54 |
| 18 | 2,97 | 3,61 | 4,00 | 4,28 | 4,50 | 4,67 | 4,82 | 4,96 | 5,07 | 5,17 | 5,27 | 5,35 | 5,43 | 5,50 |
| 19 | 2,96 | 3,59 | 3,98 | 4,25 | 4,47 | 4,64 | 4,79 | 4,92 | 5,04 | 5,14 | 5,23 | 5,32 | 5,39 | 5,46 |
| 20 | 2,95 | 3,58 | 3,96 | 4,23 | 4,44 | 4,62 | 4,77 | 4,90 | 5,01 | 5,11 | 5,20 | 5,28 | 5,36 | 5,43 |
| 24 | 2,92 | 3,53 | 3,90 | 4,17 | 4,37 | 4,54 | 4,68 | 4,81 | 4,92 | 5,01 | 5,10 | 5,18 | 5,25 | 5,32 |
| 30 | 2,89 | 3,49 | 3,84 | 4,10 | 4,30 | 4,46 | 4,60 | 4,72 | 4,82 | 4,92 | 5,00 | 5,08 | 5,15 | 5,21 |
| 40 | 2,86 | 3,44 | 3,79 | 4,04 | 4,23 | 4,39 | 4,52 | 4,64 | 4,74 | 4,82 | 4,90 | 4,98 | 5,04 | 5,11 |
| 60 | 2,83 | 3,40 | 3,74 | 3,98 | 4,16 | 4,31 | 4,44 | 4,55 | 4,65 | 4,73 | 4,81 | 4,88 | 4,94 | 5,00 |
| 120 | 2,80 | 3,36 | 3,69 | 3,92 | 4,10 | 4,24 | 4,36 | 4,47 | 4,56 | 4,64 | 4,71 | 4,78 | 4,84 | 4,90 |
| ∞ | 2,77 | 3,31 | 3,63 | 3,86 | 4,03 | 4,17 | 4,29 | 4,39 | 4,47 | 4,55 | 4,62 | 4,68 | 4,74 | 4,80 |

*Kaynak:* H. L. Harter (1960), Ranj ve Standartlaştırılmış Ranj Tabloları, *Annals of Mathematical Statistics, 31,* 1122–1147'den yazarın ve yayıncının izni ile kısaltılarak alınmıştır.

**Tablo 2** $\alpha = 0{,}05$

| Hata | r = Sıralanmış Ortalamalar Arasındaki Adımlar | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *sd* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* |
| 1 | 90,03 | 135,0 | 164,3 | 185,6 | 202,2 | 215,8 | 227,2 | 237,0 | 245,6 | 253,2 | 260,0 | 266,2 | 271,8 | 277,0 |
| 2 | 14,04 | 19,02 | 22,29 | 24,72 | 26,63 | 28,20 | 29,53 | 30,68 | 31,69 | 32,59 | 33,40 | 34,13 | 34,81 | 35,43 |
| 3 | 8,26 | 10,62 | 12,17 | 13,33 | 14,24 | 15,00 | 15,64 | 16,20 | 16,69 | 17,13 | 17,53 | 17,89 | 18,22 | 18,52 |
| 4 | 6,51 | 8,12 | 9,17 | 9,96 | 10,58 | 11,10 | 11,55 | 11,93 | 12,27 | 12,57 | 12,84 | 13,09 | 13,32 | 13,53 |
| 5 | 5,70 | 6,98 | 7,80 | 8,42 | 8,91 | 9,32 | 9,67 | 9,97 | 10,24 | 10,48 | 10,70 | 10,89 | 11,08 | 11,24 |
| 6 | 5,24 | 6,33 | 7,03 | 7,56 | 7,97 | 8,32 | 8,62 | 8,87 | 9,10 | 9,30 | 9,48 | 9,65 | 9,81 | 9,95 |
| 7 | 4,95 | 5,92 | 6,54 | 7,00 | 7,37 | 7,68 | 7,94 | 8,17 | 8,37 | 8,55 | 8,71 | 8,86 | 9,00 | 9,12 |
| 8 | 4,75 | 5,64 | 6,20 | 6,62 | 6,96 | 7,24 | 7,47 | 7,68 | 7,86 | 8,03 | 8,18 | 8,31 | 8,44 | 8,55 |
| 9 | 4,60 | 5,43 | 5,96 | 6,35 | 6,66 | 6,92 | 7,13 | 7,32 | 7,50 | 7,65 | 7,78 | 7,91 | 8,02 | 8,13 |
| 10 | 4,48 | 5,27 | 5,77 | 6,14 | 6,43 | 6,67 | 6,88 | 7,06 | 7,21 | 7,36 | 7,48 | 7,60 | 7,71 | 7,81 |
| 11 | 4,39 | 5,15 | 5,62 | 5,97 | 6,25 | 6,48 | 6,67 | 6,84 | 6,99 | 7,13 | 7,25 | 7,36 | 7,46 | 7,56 |
| 12 | 4,32 | 5,05 | 5,50 | 5,84 | 6,10 | 6,32 | 6,51 | 6,67 | 6,81 | 6,94 | 7,06 | 7,17 | 7,26 | 7,36 |
| 13 | 4,26 | 4,96 | 5,40 | 5,73 | 5,98 | 6,19 | 6,37 | 6,53 | 6,67 | 6,79 | 6,90 | 7,01 | 7,10 | 7,19 |
| 14 | 4,21 | 4,90 | 5,32 | 5,63 | 5,88 | 6,08 | 6,26 | 6,41 | 6,54 | 6,66 | 6,77 | 6,87 | 6,96 | 7,05 |
| 15 | 4,17 | 4,84 | 5,25 | 5,56 | 5,80 | 5,99 | 6,16 | 6,31 | 6,44 | 6,56 | 6,66 | 6,76 | 6,84 | 6,93 |
| 16 | 4,13 | 4,79 | 5,19 | 5,49 | 5,72 | 5,92 | 6,08 | 6,22 | 6,35 | 6,46 | 6,56 | 6,66 | 6,74 | 6,82 |
| 17 | 4,10 | 4,74 | 5,14 | 5,43 | 5,66 | 5,85 | 6,01 | 6,15 | 6,27 | 6,38 | 6,48 | 6,57 | 6,66 | 6,73 |
| 18 | 4,07 | 4,70 | 5,09 | 5,38 | 5,60 | 5,79 | 5,94 | 6,08 | 6,20 | 6,31 | 6,41 | 6,50 | 6,58 | 6,66 |
| 19 | 4,05 | 4,67 | 5,05 | 5,33 | 5,55 | 5,74 | 5,89 | 6,02 | 6,14 | 6,25 | 6,34 | 6,43 | 6,51 | 6,58 |
| 20 | 4,02 | 4,64 | 5,02 | 5,29 | 5,51 | 5,69 | 5,84 | 5,97 | 6,09 | 6,19 | 6,28 | 6,37 | 6,45 | 6,52 |
| 24 | 3,96 | 4,55 | 4,91 | 5,17 | 5,37 | 5,54 | 5,69 | 5,81 | 5,92 | 6,02 | 6,11 | 6,19 | 6,26 | 6,33 |
| 30 | 3,89 | 4,46 | 4,80 | 5,05 | 5,24 | 5,40 | 5,54 | 5,65 | 5,76 | 5,85 | 5,93 | 6,01 | 6,08 | 6,14 |
| 40 | 3,82 | 4,37 | 4,70 | 4,93 | 5,11 | 5,26 | 5,39 | 5,50 | 5,60 | 5,69 | 5,76 | 5,84 | 5,90 | 5,96 |
| 60 | 3,76 | 4,28 | 4,60 | 4,82 | 4,99 | 5,13 | 5,25 | 5,36 | 5,45 | 5,53 | 5,60 | 5,67 | 5,73 | 5,78 |
| 120 | 3,70 | 4,20 | 4,50 | 4,71 | 4,87 | 5,01 | 5,12 | 5,21 | 5,30 | 5,38 | 5,44 | 5,51 | 5,56 | 5,61 |
| | 3,64 | 4,12 | 4,40 | 4,60 | 4,76 | 4,88 | 4,99 | 5,08 | 5,16 | 5,23 | 5,29 | 5,35 | 5,40 | 5,45 |

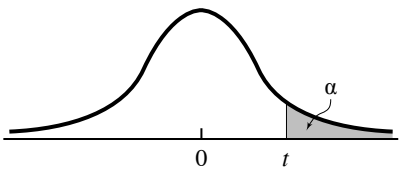
*Kaynak:* H. L. Harter (1960), Ranj ve Standartlaştırılmış Ranj Tabloları, *Annals of Mathematical Statistics, 31,* 1122–1147'den yazarın ve yayıncının izni ile kısaltılarak alınmıştır.

# Ek r´: r'den r´ 'ne Fisher Transformasyonu Tablosu

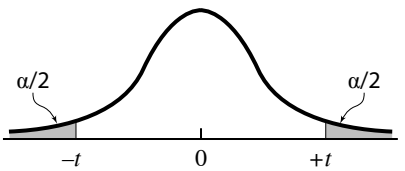
| *r* | *r´* | *r* | *r´* | *r* | *r´* | *r* | *r´* | *r* | *r´* |
|---|---|---|---|---|---|---|---|---|---|
| 0,000 | 0,000 | 0,200 | 0,203 | 0,400 | 0,424 | 0,600 | 0,693 | 0,800 | 1,099 |
| 0,005 | 0,005 | 0,205 | 0,208 | 0,405 | 0,430 | 0,605 | 0,701 | 0,805 | 1,113 |
| 0,010 | 0,010 | 0,210 | 0,213 | 0,410 | 0,436 | 0,610 | 0,709 | 0,810 | 1,127 |
| 0,015 | 0,015 | 0,215 | 0,218 | 0,415 | 0,442 | 0,615 | 0,717 | 0,815 | 1,142 |
| 0,020 | 0,020 | 0,220 | 0,224 | 0,420 | 0,448 | 0,620 | 0,725 | 0,820 | 1,157 |
| 0,025 | 0,025 | 0,225 | 0,229 | 0,425 | 0,454 | 0,625 | 0,733 | 0,825 | 1,172 |
| 0,030 | 0,030 | 0,230 | 0,234 | 0,430 | 0,460 | 0,630 | 0,741 | 0,830 | 1,188 |
| 0,035 | 0,035 | 0,235 | 0,239 | 0,435 | 0,466 | 0,635 | 0,750 | 0,835 | 1,204 |
| 0,040 | 0,040 | 0,240 | 0,245 | 0,440 | 0,472 | 0,640 | 0,758 | 0,840 | 1,221 |
| 0,045 | 0,045 | 0,245 | 0,250 | 0,445 | 0,478 | 0,645 | 0,767 | 0,845 | 1,238 |
| 0,050 | 0,050 | 0,250 | 0,255 | 0,450 | 0,485 | 0,650 | 0,775 | 0,850 | 1,256 |
| 0,055 | 0,055 | 0,255 | 0,261 | 0,455 | 0,491 | 0,655 | 0,784 | 0,855 | 1,274 |
| 0,060 | 0,060 | 0,260 | 0,266 | 0,460 | 0,497 | 0,660 | 0,793 | 0,860 | 1,293 |
| 0,065 | 0,065 | 0,265 | 0,271 | 0,465 | 0,504 | 0,665 | 0,802 | 0,865 | 1,313 |
| 0,070 | 0,070 | 0,270 | 0,277 | 0,470 | 0,510 | 0,670 | 0,811 | 0,870 | 1,333 |
| 0,075 | 0,075 | 0,275 | 0,282 | 0,475 | 0,517 | 0,675 | 0,820 | 0,875 | 1,354 |
| 0,080 | 0,080 | 0,280 | 0,288 | 0,480 | 0,523 | 0,680 | 0,829 | 0,880 | 1,376 |
| 0,085 | 0,085 | 0,285 | 0,293 | 0,485 | 0,530 | 0,685 | 0,838 | 0,885 | 1,398 |
| 0,090 | 0,090 | 0,290 | 0,299 | 0,490 | 0,536 | 0,690 | 0,848 | 0,890 | 1,422 |
| 0,095 | 0,095 | 0,295 | 0,304 | 0,495 | 0,543 | 0,695 | 0,858 | 0,895 | 1,447 |
| 0,100 | 0,100 | 0,300 | 0,310 | 0,500 | 0,549 | 0,700 | 0,867 | 0,900 | 1,472 |
| 0,105 | 0,105 | 0,305 | 0,315 | 0,505 | 0,556 | 0,705 | 0,877 | 0,905 | 1,499 |
| 0,110 | 0,110 | 0,310 | 0,321 | 0,510 | 0,563 | 0,710 | 0,887 | 0,910 | 1,528 |
| 0,115 | 0,116 | 0,315 | 0,326 | 0,515 | 0,570 | 0,715 | 0,897 | 0,915 | 1,557 |
| 0,120 | 0,121 | 0,320 | 0,332 | 0,520 | 0,576 | 0,720 | 0,908 | 0,920 | 1,589 |
| 0,125 | 0,126 | 0,325 | 0,337 | 0,525 | 0,583 | 0,725 | 0,918 | 0,925 | 1,623 |
| 0,130 | 0,131 | 0,330 | 0,343 | 0,530 | 0,590 | 0,730 | 0,929 | 0,930 | 1,658 |
| 0,135 | 0,136 | 0,335 | 0,348 | 0,535 | 0,597 | 0,735 | 0,940 | 0,935 | 1,697 |
| 0,140 | 0,141 | 0,340 | 0,354 | 0,540 | 0,604 | 0,740 | 0,950 | 0,940 | 1,738 |
| 0,145 | 0,146 | 0,345 | 0,360 | 0,545 | 0,611 | 0,745 | 0,962 | 0,945 | 1,783 |
| 0,150 | 0,151 | 0,350 | 0,365 | 0,550 | 0,618 | 0,750 | 0,973 | 0,950 | 1,832 |
| 0,155 | 0,156 | 0,355 | 0,371 | 0,555 | 0,626 | 0,755 | 0,984 | 0,955 | 1,886 |
| 0,160 | 0,161 | 0,360 | 0,377 | 0,560 | 0,633 | 0,760 | 0,996 | 0,960 | 1,946 |
| 0,165 | 0,167 | 0,365 | 0,383 | 0,565 | 0,640 | 0,765 | 1,008 | 0,965 | 2,014 |
| 0,170 | 0,172 | 0,370 | 0,388 | 0,570 | 0,648 | 0,770 | 1,020 | 0,970 | 2,092 |
| 0,175 | 0,177 | 0,375 | 0,394 | 0,575 | 0,655 | 0,775 | 1,033 | 0,975 | 2,185 |
| 0,180 | 0,182 | 0,380 | 0,400 | 0,580 | 0,662 | 0,780 | 1,045 | 0,980 | 2,298 |
| 0,185 | 0,187 | 0,385 | 0,406 | 0,585 | 0,670 | 0,785 | 1,058 | 0,985 | 2,443 |
| 0,190 | 0,192 | 0,390 | 0,412 | 0,590 | 0,678 | 0,790 | 1,071 | 0,990 | 2,647 |
| 0,195 | 0,198 | 0,395 | 0,418 | 0,595 | 0,685 | 0,795 | 1,085 | 0,995 | 2,994 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Ek *t*: *t* Dağılımının Yüzdelik Noktaları




Tek-kuyruklu test




İki-kuyruklu test

| | Tek-Kuyruklu Test İçin Manidarlık Düzeyi | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *0,25* | *0,20* | *0,15* | *0,10* | *0,05* | *0,025* | *0,01* | *0,005* | *0,0005* |
| | **İki-Kuyruklu Test İçin Manidarlık Düzeyi** | | | | | | | | |
| *sd* | *0,50* | *0,40* | *0,30* | *0,20* | *0,10* | *0,05* | *0,02* | *0,01* | *0,001* |
| 1 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 | 636,620 |
| 2 | 0,816 | 1,061 | 1,386 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 31,599 |
| 3 | 0,765 | 0,978 | 1,250 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 12,924 |
| 4 | 0,741 | 0,941 | 1,190 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 8,610 |
| 5 | 0,727 | 0,920 | 1,156 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 6,869 |
| 6 | 0,718 | 0,906 | 1,134 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 5,959 |
| 7 | 0,711 | 0,896 | 1,119 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 5,408 |
| 8 | 0,706 | 0,889 | 1,108 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 5,041 |
| 9 | 0,703 | 0,883 | 1,100 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,781 |
| 10 | 0,700 | 0,879 | 1,093 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,587 |
| 11 | 0,697 | 0,876 | 1,088 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 4,437 |
| 12 | 0,695 | 0,873 | 1,083 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 4,318 |
| 13 | 0,694 | 0,870 | 1,079 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 4,221 |
| 14 | 0,692 | 0,868 | 1,076 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 4,140 |
| 15 | 0,691 | 0,866 | 1,074 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 4,073 |
| 16 | 0,690 | 0,865 | 1,071 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 4,015 |
| 17 | 0,689 | 0,863 | 1,069 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,965 |
| 18 | 0,688 | 0,862 | 1,067 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 | 3,922 |
| 19 | 0,688 | 0,861 | 1,066 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,883 |
| 20 | 0,687 | 0,860 | 1,064 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,850 |
| 21 | 0,686 | 0,859 | 1,063 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,819 |
| 22 | 0,686 | 0,858 | 1,061 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,792 |
| 23 | 0,685 | 0,858 | 1,060 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,768 |
| 24 | 0,685 | 0,857 | 1,059 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 | 3,745 |
| 25 | 0,684 | 0,856 | 1,058 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,725 |
| 26 | 0,684 | 0,856 | 1,058 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,707 |
| 27 | 0,684 | 0,855 | 1,057 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,690 |
| 28 | 0,683 | 0,855 | 1,056 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,674 |
| 29 | 0,683 | 0,854 | 1,055 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,659 |
| 30 | 0,683 | 0,854 | 1,055 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,646 |
| 40 | 0,681 | 0,851 | 1,050 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 3,551 |
| 50 | 0,679 | 0,849 | 1,047 | 1,299 | 1,676 | 2,009 | 2,403 | 2,678 | 3,496 |
| 100 | 0,677 | 0,845 | 1,042 | 1,290 | 1,660 | 1,984 | 2,364 | 2,626 | 3,390 |
| ∞ | 0,674 | 0,842 | 1,036 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,291 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

# Ek T: Wilcoxon'ın Eşleştirilmiş-Çiftler İşaretli-Sıralar Testi için Kritik Alt-Kuyruk T Değerleri (ve Bunlarla İlgili Olasılıklar)

Nominal α (Tek-Kuyruklu)

| | *0,05* | | *0,025* | | *0,01* | | *0,005* | |
|---|---|---|---|---|---|---|---|---|
| *N* | *T* | α | *T* | α | *T* | α | *T* | α |
| 5 | 0 | 0,0313 | | | | | | |
| | 1 | 0,0625 | | | | | | |
| 6 | 2 | 0,0469 | 0 | 0,0156 | | | | |
| | 3 | 0,0781 | 1 | 0,0313 | | | | |
| 7 | 3 | 0,0391 | 2 | 0,0234 | 0 | 0,0078 | | |
| | 4 | 0,0547 | 3 | 0,0391 | 1 | 0,0156 | | |
| 8 | 5 | 0,0391 | 3 | 0,0195 | 1 | 0,0078 | 0 | 0,0039 |
| | 6 | 0,0547 | 4 | 0,0273 | 2 | 0,0117 | 1 | 0,0078 |
| 9 | 8 | 0,0488 | 5 | 0,0195 | 3 | 0,0098 | 1 | 0,0039 |
| | 9 | 0,0645 | 6 | 0,0273 | 4 | 0,0137 | 2 | 0,0059 |
| 10 | 10 | 0,0420 | 8 | 0,0244 | 5 | 0,0098 | 3 | 0,0049 |
| | 11 | 0,0527 | 9 | 0,0322 | 6 | 0,0137 | 4 | 0,0068 |
| 11 | 13 | 0,0415 | 10 | 0,0210 | 7 | 0,0093 | 5 | 0,0049 |
| | 14 | 0,0508 | 11 | 0,0269 | 8 | 0,0122 | 6 | 0,0068 |
| 12 | 17 | 0,0461 | 13 | 0,0212 | 9 | 0,0081 | 7 | 0,0046 |
| | 18 | 0,0549 | 14 | 0,0261 | 10 | 0,0105 | 8 | 0,0061 |
| 13 | 21 | 0,0471 | 17 | 0,0239 | 12 | 0,0085 | 9 | 0,0040 |
| | 22 | 0,0549 | 18 | 0,0287 | 13 | 0,0107 | 10 | 0,0052 |
| 14 | 25 | 0,0453 | 21 | 0,0247 | 15 | 0,0083 | 12 | 0,0043 |
| | 26 | 0,0520 | 22 | 0,0290 | 16 | 0,0101 | 13 | 0,0054 |
| 15 | 30 | 0,0473 | 25 | 0,0240 | 19 | 0,0090 | 15 | 0,0042 |
| | 31 | 0,0535 | 26 | 0,0277 | 20 | 0,0108 | 16 | 0,0051 |
| 16 | 35 | 0,0467 | 29 | 0,0222 | 23 | 0,0091 | 19 | 0,0046 |
| | 36 | 0,0523 | 30 | 0,0253 | 24 | 0,0107 | 20 | 0,0055 |
| 17 | 41 | 0,0492 | 34 | 0,0224 | 27 | 0,0087 | 23 | 0,0047 |
| | 42 | 0,0544 | 35 | 0,0253 | 28 | 0,0101 | 24 | 0,0055 |
| 18 | 47 | 0,0494 | 40 | 0,0241 | 32 | 0,0091 | 27 | 0,0045 |
| | 48 | 0,0542 | 41 | 0,0269 | 33 | 0,0104 | 28 | 0,0052 |
| 19 | 53 | 0,0478 | 46 | 0,0247 | 37 | 0,0090 | 32 | 0,0047 |
| | 54 | 0,0521 | 47 | 0,0273 | 38 | 0,0102 | 33 | 0,0054 |
| 20 | 60 | 0,0487 | 52 | 0,0242 | 43 | 0,0096 | 37 | 0,0047 |
| | 61 | 0,0527 | 53 | 0,0266 | 44 | 0,0107 | 38 | 0,0053 |
| 21 | 67 | 0,0479 | 58 | 0,0230 | 49 | 0,0097 | 42 | 0,0045 |
| | 68 | 0,0516 | 59 | 0,0251 | 50 | 0,0108 | 43 | 0,0051 |

*(devam ediyor)*

## Ek T *(devam ediyor)*

| | Nominal $\alpha$ (Tek-Kuyruklu) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *0,05* | | *0,025* | | *0,01* | | *0,005* | |
| *N* | *T* | $\alpha$ | *T* | $\alpha$ | *T* | $\alpha$ | *T* | $\alpha$ |
| 22 | 75 | 0,0492 | 65 | 0,0231 | 55 | 0,0095 | 48 | 0,0046 |
| | 76 | 0,0527 | 66 | 0,0250 | 56 | 0,0104 | 49 | 0,0052 |
| 23 | 83 | 0,0490 | 73 | 0,0242 | 62 | 0,0098 | 54 | 0,0046 |
| | 84 | 0,0523 | 74 | 0,0261 | 63 | 0,0107 | 55 | 0,0051 |
| 24 | 91 | 0,0475 | 81 | 0,0245 | 69 | 0,0097 | 61 | 0,0048 |
| | 92 | 0,0505 | 82 | 0,0263 | 70 | 0,0106 | 62 | 0,0053 |
| 25 | 100 | 0,0479 | 89 | 0,0241 | 76 | 0,0094 | 68 | 0,0048 |
| | 101 | 0,0507 | 90 | 0,0258 | 77 | 0,0101 | 69 | 0,0053 |
| 26 | 110 | 0,0497 | 98 | 0,0247 | 84 | 0,0095 | 75 | 0,0047 |
| | 111 | 0,0524 | 99 | 0,0263 | 85 | 0,0102 | 76 | 0,0051 |
| 27 | 119 | 0,0477 | 107 | 0,0246 | 92 | 0,0093 | 83 | 0,0048 |
| | 120 | 0,0502 | 108 | 0,0260 | 93 | 0,0100 | 84 | 0,0052 |
| 28 | 130 | 0,0496 | 116 | 0,0239 | 101 | 0,0096 | 91 | 0,0048 |
| | 131 | 0,0521 | 117 | 0,0252 | 102 | 0,0102 | 92 | 0,0051 |
| 29 | 140 | 0,0482 | 126 | 0,0240 | 110 | 0,0095 | 100 | 0,0049 |
| | 141 | 0,0504 | 127 | 0,0253 | 111 | 0,0101 | 101 | 0,0053 |
| 30 | 151 | 0,0481 | 137 | 0,0249 | 120 | 0,0098 | 109 | 0,0050 |
| | 152 | 0,0502 | 138 | 0,0261 | 121 | 0,0104 | 110 | 0,0053 |
| 31 | 163 | 0,0491 | 147 | 0,0239 | 130 | 0,0099 | 118 | 0,0049 |
| | 164 | 0,0512 | 148 | 0,0251 | 131 | 0,0105 | 119 | 0,0052 |
| 32 | 175 | 0,0492 | 159 | 0,0249 | 140 | 0,0097 | 128 | 0,0050 |
| | 176 | 0,0512 | 160 | 0,0260 | 141 | 0,0103 | 129 | 0,0053 |
| 33 | 187 | 0,0485 | 170 | 0,0242 | 151 | 0,0099 | 138 | 0,0049 |
| | 188 | 0,0503 | 171 | 0,0253 | 152 | 0,0104 | 139 | 0,0052 |
| 34 | 200 | 0,0488 | 182 | 0,0242 | 162 | 0,0098 | 148 | 0,0048 |
| | 201 | 0,0506 | 183 | 0,0252 | 163 | 0,0103 | 149 | 0,0051 |
| 35 | 213 | 0,0484 | 195 | 0,0247 | 173 | 0,0096 | 159 | 0,0048 |
| | 214 | 0,0501 | 196 | 0,0257 | 174 | 0,0100 | 160 | 0,0051 |
| 36 | 227 | 0,0489 | 208 | 0,0248 | 185 | 0,0096 | 171 | 0,0050 |
| | 228 | 0,0505 | 209 | 0,0258 | 186 | 0,0100 | 172 | 0,0052 |
| 37 | 241 | 0,0487 | 221 | 0,0245 | 198 | 0,0099 | 182 | 0,0048 |
| | 242 | 0,0503 | 222 | 0,0254 | 199 | 0,0103 | 183 | 0,0050 |

*Kaynak:* Bu sayfalardaki sayılar kitabın yazarı tarafından hesaplanmıştır.

## Ek *T* *(devam ediyor)*

| | Nominal $\alpha$ (Tek-Kuyruklu) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *0,05* | | *0,025* | | *0,01* | | *0,005* | |
| *N* | *T* | | *T* | $\alpha$ | *T* | $\alpha$ | *T* | $\alpha$ |
| 38 | 256 | 0,0493 | 235 | 0,0247 | 211 | 0,0099 | 194 | 0,0048 |
| | 257 | 0,0509 | 236 | 0,0256 | 212 | 0,0104 | 195 | 0,0050 |
| 39 | 271 | 0,0492 | 249 | 0,0246 | 224 | 0,0099 | 207 | 0,0049 |
| | 272 | 0,0507 | 250 | 0,0254 | 225 | 0,0103 | 208 | 0,0051 |
| 40 | 286 | 0,0486 | 264 | 0,0249 | 238 | 0,0100 | 220 | 0,0049 |
| | 287 | 0,0500 | 265 | 0,0257 | 239 | 0,0104 | 221 | 0,0051 |
| 41 | 302 | 0,0488 | 279 | 0,0248 | 252 | 0,0100 | 233 | 0,0048 |
| | 303 | 0,0501 | 280 | 0,0256 | 253 | 0,0103 | 234 | 0,0050 |
| 42 | 319 | 0,0496 | 294 | 0,0245 | 266 | 0,0098 | 247 | 0,0049 |
| | 320 | 0,0509 | 295 | 0,0252 | 267 | 0,0102 | 248 | 0,0051 |
| 43 | 336 | 0,0498 | 310 | 0,0245 | 281 | 0,0098 | 261 | 0,0048 |
| | 337 | 0,0511 | 311 | 0,0252 | 282 | 0,0102 | 262 | 0,0050 |
| 44 | 353 | 0,0495 | 327 | 0,0250 | 296 | 0,0097 | 276 | 0,0049 |
| | 354 | 0,0507 | 328 | 0,0257 | 297 | 0,0101 | 277 | 0,0051 |
| 45 | 371 | 0,0498 | 343 | 0,0244 | 312 | 0,0098 | 291 | 0,0049 |
| | 372 | 0,0510 | 344 | 0,0251 | 313 | 0,0101 | 292 | 0,0051 |
| 46 | 389 | 0,0497 | 361 | 0,0249 | 328 | 0,0098 | 307 | 0,0050 |
| | 390 | 0,0508 | 362 | 0,0256 | 329 | 0,0101 | 308 | 0,0052 |
| 47 | 407 | 0,0490 | 378 | 0,0245 | 345 | 0,0099 | 322 | 0,0048 |
| | 408 | 0,0501 | 379 | 0,0251 | 346 | 0,0102 | 323 | 0,0050 |
| 48 | 426 | 0,0490 | 396 | 0,0244 | 362 | 0,0099 | 339 | 0,0050 |
| | 427 | 0,0500 | 397 | 0,0251 | 363 | 0,0102 | 340 | 0,0051 |
| 49 | 446 | 0,0495 | 415 | 0,0247 | 379 | 0,0098 | 355 | 0,0049 |
| | 447 | 0,0505 | 416 | 0,0253 | 380 | 0,0100 | 356 | 0,0050 |
| 50 | 466 | 0,0495 | 434 | 0,0247 | 397 | 0,0098 | 373 | 0,0050 |
| | 467 | 0,0506 | 435 | 0,0253 | 398 | 0,0101 | 374 | 0,0051 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Ek t′: Bonferroni Çoklu Karşılaştırmalar Testi Kritik Değerleri

**Tablo 1** $\alpha = 0{,}05$

| | Karşılaştırmaların Sayısı | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *sd* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *15* | *20* | *25* | *30* | *35* | *40* | *45* | *50* | *55* |
| 5 | 3,16 | 3,53 | 3,81 | 4,03 | 4,22 | 4,38 | 4,53 | 4,66 | 4,77 | 5,25 | 5,60 | 5,89 | 6,14 | 6,35 | 6,54 | 6,71 | 6,87 | 7,01 |
| 6 | 2,97 | 3,29 | 3,52 | 3,71 | 3,86 | 4,00 | 4,12 | 4,22 | 4,32 | 4,70 | 4,98 | 5,21 | 5,40 | 5,56 | 5,71 | 5,84 | 5,96 | 6,07 |
| 7 | 2,84 | 3,13 | 3,34 | 3,50 | 3,64 | 3,75 | 3,86 | 3,95 | 4,03 | 4,36 | 4,59 | 4,79 | 4,94 | 5,08 | 5,20 | 5,31 | 5,41 | 5,50 |
| 8 | 2,75 | 3,02 | 3,21 | 3,36 | 3,48 | 3,58 | 3,68 | 3,76 | 3,83 | 4,12 | 4,33 | 4,50 | 4,64 | 4,76 | 4,86 | 4,96 | 5,04 | 5,12 |
| 9 | 2,69 | 2,93 | 3,11 | 3,25 | 3,36 | 3,46 | 3,55 | 3,62 | 3,69 | 3,95 | 4,15 | 4,30 | 4,42 | 4,53 | 4,62 | 4,71 | 4,78 | 4,85 |
| 10 | 2,63 | 2,87 | 3,04 | 3,17 | 3,28 | 3,37 | 3,45 | 3,52 | 3,58 | 3,83 | 4,00 | 4,14 | 4,26 | 4,36 | 4,44 | 4,52 | 4,59 | 4,65 |
| 11 | 2,59 | 2,82 | 2,98 | 3,11 | 3,21 | 3,29 | 3,37 | 3,44 | 3,50 | 3,73 | 3,89 | 4,02 | 4,13 | 4,22 | 4,30 | 4,37 | 4,44 | 4,49 |
| 12 | 2,56 | 2,78 | 2,93 | 3,05 | 3,15 | 3,24 | 3,31 | 3,37 | 3,43 | 3,65 | 3,81 | 3,93 | 4,03 | 4,12 | 4,19 | 4,26 | 4,32 | 4,37 |
| 13 | 2,53 | 2,75 | 2,90 | 3,01 | 3,11 | 3,19 | 3,26 | 3,32 | 3,37 | 3,58 | 3,73 | 3,85 | 3,95 | 4,03 | 4,10 | 4,16 | 4,22 | 4,27 |
| 14 | 2,51 | 2,72 | 2,86 | 2,98 | 3,07 | 3,15 | 3,21 | 3,27 | 3,33 | 3,53 | 3,67 | 3,79 | 3,88 | 3,96 | 4,03 | 4,09 | 4,14 | 4,19 |
| 15 | 2,49 | 2,69 | 2,84 | 2,95 | 3,04 | 3,11 | 3,18 | 3,23 | 3,29 | 3,48 | 3,62 | 3,73 | 3,82 | 3,90 | 3,96 | 4,02 | 4,07 | 4,12 |
| 16 | 2,47 | 2,67 | 2,81 | 2,92 | 3,01 | 3,08 | 3,15 | 3,20 | 3,25 | 3,44 | 3,58 | 3,69 | 3,77 | 3,85 | 3,91 | 3,96 | 4,01 | 4,06 |
| 17 | 2,46 | 2,65 | 2,79 | 2,90 | 2,98 | 3,06 | 3,12 | 3,17 | 3,22 | 3,41 | 3,54 | 3,65 | 3,73 | 3,80 | 3,86 | 3,92 | 3,97 | 4,01 |
| 18 | 2,45 | 2,64 | 2,77 | 2,88 | 2,96 | 3,03 | 3,09 | 3,15 | 3,20 | 3,38 | 3,51 | 3,61 | 3,69 | 3,76 | 3,82 | 3,87 | 3,92 | 3,96 |
| 19 | 2,43 | 2,63 | 2,76 | 2,86 | 2,94 | 3,01 | 3,07 | 3,13 | 3,17 | 3,35 | 3,48 | 3,58 | 3,66 | 3,73 | 3,79 | 3,84 | 3,88 | 3,93 |
| 20 | 2,42 | 2,61 | 2,74 | 2,85 | 2,93 | 3,00 | 3,06 | 3,11 | 3,15 | 3,33 | 3,46 | 3,55 | 3,63 | 3,70 | 3,75 | 3,80 | 3,85 | 3,89 |
| 21 | 2,41 | 2,60 | 2,73 | 2,83 | 2,91 | 2,98 | 3,04 | 3,09 | 3,14 | 3,31 | 3,43 | 3,53 | 3,60 | 3,67 | 3,73 | 3,78 | 3,82 | 3,86 |
| 22 | 2,41 | 2,59 | 2,72 | 2,82 | 2,90 | 2,97 | 3,02 | 3,07 | 3,12 | 3,29 | 3,41 | 3,50 | 3,58 | 3,64 | 3,70 | 3,75 | 3,79 | 3,83 |
| 23 | 2,40 | 2,58 | 2,71 | 2,81 | 2,89 | 2,95 | 3,01 | 3,06 | 3,10 | 3,27 | 3,39 | 3,48 | 3,56 | 3,62 | 3,68 | 3,72 | 3,77 | 3,81 |
| 24 | 2,39 | 2,57 | 2,70 | 2,80 | 2,88 | 2,94 | 3,00 | 3,05 | 3,09 | 3,26 | 3,38 | 3,47 | 3,54 | 3,60 | 3,66 | 3,70 | 3,75 | 3,78 |
| 25 | 2,38 | 2,57 | 2,69 | 2,79 | 2,86 | 2,93 | 2,99 | 3,03 | 3,08 | 3,24 | 3,36 | 3,45 | 3,52 | 3,58 | 3,64 | 3,68 | 3,73 | 3,76 |
| 30 | 2,36 | 2,54 | 2,66 | 2,75 | 2,82 | 2,89 | 2,94 | 2,99 | 3,03 | 3,19 | 3,30 | 3,39 | 3,45 | 3,51 | 3,56 | 3,61 | 3,65 | 3,68 |
| 40 | 2,33 | 2,50 | 2,62 | 2,70 | 2,78 | 2,84 | 2,89 | 2,93 | 2,97 | 3,12 | 3,23 | 3,31 | 3,37 | 3,43 | 3,47 | 3,51 | 3,55 | 3,58 |
| 50 | 2,31 | 2,48 | 2,59 | 2,68 | 2,75 | 2,81 | 2,85 | 2,90 | 2,94 | 3,08 | 3,18 | 3,26 | 3,32 | 3,38 | 3,42 | 3,46 | 3,50 | 3,53 |
| 75 | 2,29 | 2,45 | 2,56 | 2,64 | 2,71 | 2,77 | 2,81 | 2,86 | 2,89 | 3,03 | 3,13 | 3,20 | 3,26 | 3,31 | 3,35 | 3,39 | 3,43 | 3,45 |
| 100 | 2,28 | 2,43 | 2,54 | 2,63 | 2,69 | 2,75 | 2,79 | 2,83 | 2,87 | 3,01 | 3,10 | 3,17 | 3,23 | 3,28 | 3,32 | 3,36 | 3,39 | 3,42 |
| ∞ | 2,24 | 2,39 | 2,50 | 2,58 | 2,64 | 2,69 | 2,73 | 2,77 | 2,81 | 2,94 | 3,02 | 3,09 | 3,14 | 3,19 | 3,23 | 3,26 | 3,29 | 3,32 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

**Tablo 2** $\alpha = 0{,}01$

| | Karşılaştırmaların Sayısı | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *sd* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *15* | *20* | *25* | *30* | *35* | *40* | *45* | *50* | *55* |
| 5 | 4,77 | 5,25 | 5,60 | 5,89 | 6,14 | 6,35 | 6,54 | 6,71 | 6,87 | 7,50 | 7,98 | 8,36 | 8,69 | 8,98 | 9,24 | 9,47 | | |
| 6 | 4,32 | 4,70 | 4,98 | 5,21 | 5,40 | 5,56 | 5,71 | 5,84 | 5,96 | 6,43 | 6,79 | 7,07 | 7,31 | 7,52 | 7,71 | 7,87 | 8,02 | 8,16 |
| 7 | 4,03 | 4,36 | 4,59 | 4,79 | 4,94 | 5,08 | 5,20 | 5,31 | 5,41 | 5,80 | 6,08 | 6,31 | 6,50 | 6,67 | 6,81 | 6,94 | 7,06 | 7,17 |
| 8 | 3,83 | 4,12 | 4,33 | 4,50 | 4,64 | 4,76 | 4,86 | 4,96 | 5,04 | 5,37 | 5,62 | 5,81 | 5,97 | 6,11 | 6,23 | 6,34 | 6,44 | 6,53 |
| 9 | 3,69 | 3,95 | 4,15 | 4,30 | 4,42 | 4,53 | 4,62 | 4,71 | 4,78 | 5,08 | 5,29 | 5,46 | 5,60 | 5,72 | 5,83 | 5,92 | 6,01 | 6,09 |
| 10 | 3,58 | 3,83 | 4,00 | 4,14 | 4,26 | 4,36 | 4,44 | 4,52 | 4,59 | 4,85 | 5,05 | 5,20 | 5,33 | 5,44 | 5,53 | 5,62 | 5,69 | 5,76 |
| 11 | 3,50 | 3,73 | 3,89 | 4,02 | 4,13 | 4,22 | 4,30 | 4,37 | 4,44 | 4,68 | 4,86 | 5,00 | 5,12 | 5,22 | 5,31 | 5,38 | 5,45 | 5,52 |
| 12 | 3,43 | 3,65 | 3,81 | 3,93 | 4,03 | 4,12 | 4,19 | 4,26 | 4,32 | 4,55 | 4,72 | 4,85 | 4,96 | 5,05 | 5,13 | 5,20 | 5,26 | 5,32 |
| 13 | 3,37 | 3,58 | 3,73 | 3,85 | 3,95 | 4,03 | 4,10 | 4,16 | 4,22 | 4,44 | 4,60 | 4,72 | 4,82 | 4,91 | 4,98 | 5,05 | 5,11 | 5,17 |
| 14 | 3,33 | 3,53 | 3,67 | 3,79 | 3,88 | 3,96 | 4,03 | 4,09 | 4,14 | 4,35 | 4,50 | 4,62 | 4,71 | 4,79 | 4,87 | 4,93 | 4,99 | 5,04 |
| 15 | 3,29 | 3,48 | 3,62 | 3,73 | 3,82 | 3,90 | 3,96 | 4,02 | 4,07 | 4,27 | 4,42 | 4,53 | 4,62 | 4,70 | 4,77 | 4,83 | 4,88 | 4,93 |
| 16 | 3,25 | 3,44 | 3,58 | 3,69 | 3,77 | 3,85 | 3,91 | 3,96 | 4,01 | 4,21 | 4,35 | 4,45 | 4,54 | 4,62 | 4,68 | 4,74 | 4,79 | 4,84 |
| 17 | 3,22 | 3,41 | 3,54 | 3,65 | 3,73 | 3,80 | 3,86 | 3,92 | 3,97 | 4,15 | 4,29 | 4,39 | 4,47 | 4,55 | 4,61 | 4,66 | 4,71 | 4,76 |
| 18 | 3,20 | 3,38 | 3,51 | 3,61 | 3,69 | 3,76 | 3,82 | 3,87 | 3,92 | 4,10 | 4,23 | 4,33 | 4,42 | 4,49 | 4,55 | 4,60 | 4,65 | 4,69 |
| 19 | 3,17 | 3,35 | 3,48 | 3,58 | 3,66 | 3,73 | 3,79 | 3,84 | 3,88 | 4,06 | 4,19 | 4,28 | 4,36 | 4,43 | 4,49 | 4,54 | 4,59 | 4,63 |
| 20 | 3,15 | 3,33 | 3,46 | 3,55 | 3,63 | 3,70 | 3,75 | 3,80 | 3,85 | 4,02 | 4,15 | 4,24 | 4,32 | 4,39 | 4,44 | 4,49 | 4,54 | 4,58 |
| 21 | 3,14 | 3,31 | 3,43 | 3,53 | 3,60 | 3,67 | 3,73 | 3,78 | 3,82 | 3,99 | 4,11 | 4,20 | 4,28 | 4,34 | 4,40 | 4,45 | 4,49 | 4,53 |
| 22 | 3,12 | 3,29 | 3,41 | 3,50 | 3,58 | 3,64 | 3,70 | 3,75 | 3,79 | 3,96 | 4,08 | 4,17 | 4,24 | 4,31 | 4,36 | 4,41 | 4,45 | 4,49 |
| 23 | 3,10 | 3,27 | 3,39 | 3,48 | 3,56 | 3,62 | 3,68 | 3,72 | 3,77 | 3,93 | 4,05 | 4,14 | 4,21 | 4,27 | 4,33 | 4,37 | 4,42 | 4,45 |
| 24 | 3,09 | 3,26 | 3,38 | 3,47 | 3,54 | 3,60 | 3,66 | 3,70 | 3,75 | 3,91 | 4,02 | 4,11 | 4,18 | 4,24 | 4,29 | 4,34 | 4,38 | 4,42 |
| 25 | 3,08 | 3,24 | 3,36 | 3,45 | 3,52 | 3,58 | 3,64 | 3,68 | 3,73 | 3,88 | 4,00 | 4,08 | 4,15 | 4,21 | 4,27 | 4,31 | 4,35 | 4,39 |
| 30 | 3,03 | 3,19 | 3,30 | 3,39 | 3,45 | 3,51 | 3,56 | 3,61 | 3,65 | 3,80 | 3,90 | 3,98 | 4,05 | 4,11 | 4,15 | 4,20 | 4,23 | 4,27 |
| 40 | 2,97 | 3,12 | 3,23 | 3,31 | 3,37 | 3,43 | 3,47 | 3,51 | 3,55 | 3,69 | 3,79 | 3,86 | 3,92 | 3,98 | 4,02 | 4,06 | 4,09 | 4,13 |
| 50 | 2,94 | 3,08 | 3,18 | 3,26 | 3,32 | 3,38 | 3,42 | 3,46 | 3,50 | 3,63 | 3,72 | 3,79 | 3,85 | 3,90 | 3,94 | 3,98 | 4,01 | 4,04 |
| 75 | 2,89 | 3,03 | 3,13 | 3,20 | 3,26 | 3,31 | 3,35 | 3,39 | 3,43 | 3,55 | 3,64 | 3,71 | 3,76 | 3,81 | 3,85 | 3,88 | 3,91 | 3,94 |
| 100 | 2,87 | 3,01 | 3,10 | 3,17 | 3,23 | 3,28 | 3,32 | 3,36 | 3,39 | 3,51 | 3,60 | 3,66 | 3,72 | 3,76 | 3,80 | 3,83 | 3,86 | 3,89 |
| ∞ | 2,81 | 2,94 | 3,02 | 3,09 | 3,14 | 3,19 | 3,23 | 3,26 | 3,29 | 3,40 | 3,48 | 3,54 | 3,59 | 3,63 | 3,66 | 3,69 | 3,72 | 3,74 |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.

## Ek $t_d$: Dunnet İstatistiğinin ($t_d$) Kritik Değerleri

| Hata | | İki-Kuyruklu Karşılaştırmalar<br>$k$ = İşlem Ortalamaları Sayısı, Kontrol Dahil | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *sd* | $\alpha$ | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
| 5 | 0,05 | 2,57 | 3,03 | 3,29 | 3,48 | 3,62 | 3,73 | 3,82 | 3,90 | 3,97 |
| | 0,01 | 4,03 | 4,63 | 4,98 | 5,22 | 5,41 | 5,56 | 5,69 | 5,80 | 5,89 |
| 6 | 0,05 | 2,45 | 2,86 | 3,10 | 3,26 | 3,39 | 3,49 | 3,57 | 3,64 | 3,71 |
| | 0,01 | 3,71 | 4,21 | 4,51 | 4,71 | 4,87 | 5,00 | 5,10 | 5,20 | 5,28 |
| 7 | 0,05 | 2,36 | 2,75 | 2,97 | 3,12 | 3,24 | 3,33 | 3,41 | 3,47 | 3,53 |
| | 0,01 | 3,50 | 3,95 | 4,21 | 4,39 | 4,53 | 4,64 | 4,74 | 4,82 | 4,89 |
| 8 | 0,05 | 2,31 | 2,67 | 2,88 | 3,02 | 3,13 | 3,22 | 3,29 | 3,35 | 3,41 |
| | 0,01 | 3,36 | 3,77 | 4,00 | 4,17 | 4,29 | 4,40 | 4,48 | 4,56 | 4,62 |
| 9 | 0,05 | 2,26 | 2,61 | 2,81 | 2,95 | 3,05 | 3,14 | 3,20 | 3,26 | 3,32 |
| | 0,01 | 3,25 | 3,63 | 3,85 | 4,01 | 4,12 | 4,22 | 4,30 | 4,37 | 4,43 |
| 10 | 0,05 | 2,23 | 2,57 | 2,76 | 2,89 | 2,99 | 3,07 | 3,14 | 3,19 | 3,24 |
| | 0,01 | 3,17 | 3,53 | 3,74 | 3,88 | 3,99 | 4,08 | 4,16 | 4,22 | 4,28 |
| 11 | 0,05 | 2,20 | 2,53 | 2,72 | 2,84 | 2,94 | 3,02 | 3,08 | 3,14 | 3,19 |
| | 0,01 | 3,11 | 3,45 | 3,65 | 3,79 | 3,89 | 3,98 | 4,05 | 4,11 | 4,16 |
| 12 | 0,05 | 2,18 | 2,50 | 2,68 | 2,81 | 2,90 | 2,98 | 3,04 | 3,09 | 3,14 |
| | 0,01 | 3,05 | 3,39 | 3,58 | 3,71 | 3,81 | 3,89 | 3,96 | 4,02 | 4,07 |
| 13 | 0,05 | 2,16 | 2,48 | 2,65 | 2,78 | 2,87 | 2,94 | 3,00 | 3,06 | 3,10 |
| | 0,01 | 3,01 | 3,33 | 3,52 | 3,65 | 3,74 | 3,82 | 3,89 | 3,94 | 3,99 |
| 14 | 0,05 | 2,14 | 2,46 | 2,63 | 2,75 | 2,84 | 2,91 | 2,97 | 3,02 | 3,07 |
| | 0,01 | 2,98 | 3,29 | 3,47 | 3,59 | 3,69 | 3,76 | 3,83 | 3,88 | 3,93 |
| 15 | 0,05 | 2,13 | 2,44 | 2,61 | 2,73 | 2,82 | 2,89 | 2,95 | 3,00 | 3,04 |
| | 0,01 | 2,95 | 3,25 | 3,43 | 3,55 | 3,64 | 3,71 | 3,78 | 3,83 | 3,88 |
| 16 | 0,05 | 2,12 | 2,42 | 2,59 | 2,71 | 2,80 | 2,87 | 2,92 | 2,97 | 3,02 |
| | 0,01 | 2,92 | 3,22 | 3,39 | 3,51 | 3,60 | 3,67 | 3,73 | 3,78 | 3,83 |
| 17 | 0,05 | 2,11 | 2,41 | 2,58 | 2,69 | 2,78 | 2,85 | 2,90 | 2,95 | 3,00 |
| | 0,01 | 2,90 | 3,19 | 3,36 | 3,47 | 3,56 | 3,63 | 3,69 | 3,74 | 3,79 |
| 18 | 0,05 | 2,10 | 2,40 | 2,56 | 2,68 | 2,76 | 2,83 | 2,89 | 2,94 | 2,98 |
| | 0,01 | 2,88 | 3,17 | 3,33 | 3,44 | 3,53 | 3,60 | 3,66 | 3,71 | 3,75 |
| 19 | 0,05 | 2,09 | 2,39 | 2,55 | 2,66 | 2,75 | 2,81 | 2,87 | 2,92 | 2,96 |
| | 0,01 | 2,86 | 3,15 | 3,31 | 3,42 | 3,50 | 3,57 | 3,63 | 3,68 | 3,72 |
| 20 | 0,05 | 2,09 | 2,38 | 2,54 | 2,65 | 2,73 | 2,80 | 2,86 | 2,90 | 2,95 |
| | 0,01 | 2,85 | 3,13 | 3,29 | 3,40 | 3,48 | 3,55 | 3,60 | 3,65 | 3,69 |
| 24 | 0,05 | 2,06 | 2,35 | 2,51 | 2,61 | 2,70 | 2,76 | 2,81 | 2,86 | 2,90 |
| | 0,01 | 2,80 | 3,07 | 3,22 | 3,32 | 3,40 | 3,47 | 3,52 | 3,57 | 3,61 |
| 30 | 0,05 | 2,04 | 2,32 | 2,47 | 2,58 | 2,66 | 2,72 | 2,77 | 2,82 | 2,86 |
| | 0,01 | 2,75 | 3,01 | 3,15 | 3,25 | 3,33 | 3,39 | 3,44 | 3,49 | 3,52 |
| 40 | 0,05 | 2,02 | 2,29 | 2,44 | 2,54 | 2,62 | 2,68 | 2,73 | 2,77 | 2,81 |
| | 0,01 | 2,70 | 2,95 | 3,09 | 3,19 | 3,26 | 3,32 | 3,37 | 3,41 | 3,44 |
| 60 | 0,05 | 2,00 | 2,27 | 2,41 | 2,51 | 2,58 | 2,64 | 2,69 | 2,73 | 2,77 |
| | 0,01 | 2,66 | 2,90 | 3,03 | 3,12 | 3,19 | 3,25 | 3,29 | 3,33 | 3,37 |
| 120 | 0,05 | 1,98 | 2,24 | 2,38 | 2,47 | 2,55 | 2,60 | 2,65 | 2,69 | 2,73 |
| | 0,01 | 2,62 | 2,85 | 2,97 | 3,06 | 3,12 | 3,18 | 3,22 | 3,26 | 3,29 |
| ∞ | 0,05 | 1,96 | 2,21 | 2,35 | 2,44 | 2,51 | 2,57 | 2,61 | 2,65 | 2,69 |
| | 0,01 | 2,58 | 2,79 | 2,92 | 3,00 | 3,06 | 3,11 | 3,15 | 3,19 | 3,22 |

*Kaynak:* C. W. Dunnett (1964), New tables for multiple comparisons with a control, *Biometrics 20,* 482–491. Biometric Society'nin izniyle.

# Ek $W_S$: İki Bağımsız Örneklem ($N_1 \leq N_2$) İçin Sıra-Toplamı Testine Ait Kritik Alt-Sınır Değerleri

| | $N_1 = 1$ | | | | | | | $N_1 = 2$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | $N_2$ |
| 2 | | | | | | | 4 | | | | | | — | 10 | 2 |
| 3 | | | | | | | 5 | | | | | | 3 | 12 | 3 |
| 4 | | | | | | | 6 | | | | | — | 3 | 14 | 4 |
| 5 | | | | | | | 7 | | | | | 3 | 4 | 16 | 5 |
| 6 | | | | | | | 8 | | | | | 3 | 4 | 18 | 6 |
| 7 | | | | | | | 9 | | | | — | 3 | 4 | 20 | 7 |
| 8 | | | | | | — | 10 | | | | 3 | 4 | 5 | 22 | 8 |
| 9 | | | | | | 1 | 11 | | | | 3 | 4 | 5 | 24 | 9 |
| 10 | | | | | | 1 | 12 | | | | 3 | 4 | 6 | 26 | 10 |
| 11 | | | | | | 1 | 13 | | | | 3 | 4 | 6 | 28 | 11 |
| 12 | | | | | | 1 | 14 | | | — | 4 | 5 | 7 | 30 | 12 |
| 13 | | | | | | 1 | 15 | | | 3 | 4 | 5 | 7 | 32 | 13 |
| 14 | | | | | | 1 | 16 | | | 3 | 4 | 6 | 8 | 34 | 14 |
| 15 | | | | | | 1 | 17 | | | 3 | 4 | 6 | 8 | 36 | 15 |
| 16 | | | | | | 1 | 18 | | | 3 | 4 | 6 | 8 | 38 | 16 |
| 17 | | | | | | 1 | 19 | | | 3 | 5 | 6 | 9 | 40 | 17 |
| 18 | | | | | — | 1 | 20 | | — | 3 | 5 | 7 | 9 | 42 | 18 |
| 19 | | | | | 1 | 2 | 21 | | 3 | 4 | 5 | 7 | 10 | 44 | 19 |
| 20 | | | | | 1 | 2 | 22 | | 3 | 4 | 5 | 7 | 10 | 46 | 20 |
| 21 | | | | | 1 | 2 | 23 | | 3 | 4 | 6 | 8 | 11 | 48 | 21 |
| 22 | | | | | 1 | 2 | 24 | | 3 | 4 | 6 | 8 | 11 | 50 | 22 |
| 23 | | | | | 1 | 2 | 25 | | 3 | 4 | 6 | 8 | 12 | 52 | 23 |
| 24 | | | | | 1 | 2 | 26 | | 3 | 4 | 6 | 9 | 12 | 54 | 24 |
| 25 | — | — | — | — | 1 | 2 | 27 | — | 3 | 4 | 6 | 9 | 12 | 56 | 25 |

| | $N_1 = 3$ | | | | | | | $N_1 = 4$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | $N_2$ |
| 3 | | | | | 6 | 7 | 21 | | | | | | | | |
| 4 | | | | — | 6 | 7 | 24 | | | — | 10 | 11 | 13 | 36 | 4 |
| 5 | | | | 6 | 7 | 8 | 27 | | — | 10 | 11 | 12 | 14 | 40 | 5 |
| 6 | | | — | 7 | 8 | 9 | 30 | | 10 | 11 | 12 | 13 | 15 | 44 | 6 |
| 7 | | | 6 | 7 | 8 | 10 | 33 | | 10 | 11 | 13 | 14 | 16 | 48 | 7 |
| 8 | | — | 6 | 8 | 9 | 11 | 36 | | 11 | 12 | 14 | 15 | 17 | 52 | 8 |
| 9 | | 6 | 7 | 8 | 10 | 11 | 39 | — | 11 | 13 | 14 | 16 | 19 | 56 | 9 |
| 10 | | 6 | 7 | 9 | 10 | 12 | 42 | 10 | 12 | 13 | 15 | 17 | 20 | 60 | 10 |
| 11 | | 6 | 7 | 9 | 11 | 13 | 45 | 10 | 12 | 14 | 16 | 18 | 21 | 64 | 11 |
| 12 | | 7 | 8 | 10 | 11 | 14 | 48 | 10 | 13 | 15 | 17 | 19 | 22 | 68 | 12 |
| 13 | | 7 | 8 | 10 | 12 | 15 | 51 | 11 | 13 | 15 | 18 | 20 | 23 | 72 | 13 |
| 14 | | 7 | 8 | 11 | 13 | 16 | 54 | 11 | 14 | 16 | 19 | 21 | 25 | 76 | 14 |
| 15 | | 8 | 9 | 11 | 13 | 16 | 57 | 11 | 15 | 17 | 20 | 22 | 26 | 80 | 15 |

*(devam ediyor)*

## Ek *Ws* *(devamı)*

| | $N_1 = 3$ | | | | | | | $N_1 = 4$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 16 | — | 8 | 9 | 12 | 14 | 17 | 60 | 12 | 15 | 17 | 21 | 24 | 27 | 84 | 16 |
| 17 | 6 | 8 | 10 | 12 | 15 | 18 | 63 | 12 | 16 | 18 | 21 | 25 | 28 | 88 | 17 |
| 18 | 6 | 8 | 10 | 13 | 15 | 19 | 66 | 13 | 16 | 19 | 22 | 26 | 30 | 92 | 18 |
| 19 | 6 | 9 | 10 | 13 | 16 | 20 | 69 | 13 | 17 | 19 | 23 | 27 | 31 | 96 | 19 |
| 20 | 6 | 9 | 11 | 14 | 17 | 21 | 72 | 13 | 18 | 20 | 24 | 28 | 32 | 100 | 20 |
| 21 | 7 | 9 | 11 | 14 | 17 | 21 | 75 | 14 | 18 | 21 | 25 | 29 | 33 | 104 | 21 |
| 22 | 7 | 10 | 12 | 15 | 18 | 22 | 78 | 14 | 19 | 21 | 26 | 30 | 35 | 108 | 22 |
| 23 | 7 | 10 | 12 | 15 | 19 | 23 | 81 | 14 | 19 | 22 | 27 | 31 | 36 | 112 | 23 |
| 24 | 7 | 10 | 12 | 16 | 19 | 24 | 84 | 15 | 20 | 23 | 27 | 32 | 38 | 116 | 24 |
| 25 | 7 | 11 | 13 | 16 | 20 | 25 | 87 | 15 | 20 | 23 | 28 | 33 | 38 | 120 | 25 |

| | $N_1 = 5$ | | | | | | | $N_1 = 6$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 5 | | 15 | 16 | 17 | 19 | 20 | 55 | | | | | | | | |
| 6 | | 16 | 17 | 18 | 20 | 22 | 60 | — | 23 | 24 | 26 | 28 | 30 | 78 | 6 |
| 7 | — | 16 | 18 | 20 | 21 | 23 | 65 | 21 | 24 | 25 | 27 | 29 | 32 | 84 | 7 |
| 8 | 15 | 17 | 19 | 21 | 23 | 25 | 70 | 22 | 25 | 27 | 29 | 31 | 34 | 90 | 8 |
| 9 | 16 | 18 | 20 | 22 | 24 | 27 | 75 | 23 | 26 | 28 | 31 | 33 | 36 | 96 | 9 |
| 10 | 16 | 19 | 21 | 23 | 26 | 28 | 80 | 24 | 27 | 29 | 32 | 35 | 38 | 102 | 10 |
| 11 | 17 | 20 | 22 | 24 | 27 | 30 | 85 | 25 | 28 | 30 | 34 | 37 | 40 | 108 | 11 |
| 12 | 17 | 21 | 23 | 26 | 28 | 32 | 90 | 25 | 30 | 32 | 35 | 38 | 42 | 114 | 12 |
| 13 | 18 | 22 | 24 | 27 | 30 | 33 | 95 | 26 | 31 | 33 | 37 | 40 | 44 | 120 | 13 |
| 14 | 18 | 22 | 25 | 28 | 31 | 35 | 100 | 27 | 32 | 34 | 38 | 42 | 46 | 126 | 14 |
| 15 | 19 | 23 | 26 | 29 | 33 | 37 | 105 | 28 | 33 | 36 | 40 | 44 | 48 | 132 | 15 |
| 16 | 20 | 24 | 27 | 30 | 34 | 38 | 110 | 29 | 34 | 37 | 42 | 46 | 50 | 138 | 16 |
| 17 | 20 | 25 | 28 | 32 | 35 | 40 | 115 | 30 | 36 | 39 | 43 | 47 | 52 | 144 | 17 |
| 18 | 21 | 26 | 29 | 33 | 37 | 42 | 120 | 31 | 37 | 40 | 45 | 49 | 55 | 150 | 18 |
| 19 | 22 | 27 | 30 | 34 | 38 | 43 | 125 | 32 | 38 | 41 | 46 | 51 | 57 | 156 | 19 |
| 20 | 22 | 28 | 31 | 35 | 40 | 45 | 130 | 33 | 39 | 43 | 48 | 53 | 59 | 162 | 20 |
| 21 | 23 | 29 | 32 | 37 | 41 | 47 | 135 | 33 | 40 | 44 | 50 | 55 | 61 | 168 | 21 |
| 22 | 23 | 29 | 33 | 38 | 43 | 48 | 140 | 34 | 42 | 45 | 51 | 57 | 63 | 174 | 22 |
| 23 | 24 | 30 | 34 | 39 | 44 | 50 | 145 | 35 | 43 | 47 | 53 | 58 | 65 | 180 | 23 |
| 24 | 25 | 31 | 35 | 40 | 45 | 51 | 150 | 36 | 44 | 48 | 54 | 60 | 67 | 186 | 24 |
| 25 | 25 | 32 | 36 | 42 | 47 | 53 | 155 | 37 | 45 | 50 | 56 | 62 | 69 | 192 | 25 |

## Ek $W_S$ *(devamı)*

| | $N_1 = 7$ | | | | | | | $N_1 = 8$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | $N_2$ |
| 7 | 29 | 32 | 34 | 36 | 39 | 41 | 105 | | | | | | | | |
| 8 | 30 | 34 | 35 | 38 | 41 | 44 | 112 | 40 | 43 | 45 | 49 | 51 | 55 | 136 | 8 |
| 9 | 31 | 35 | 37 | 40 | 43 | 46 | 119 | 41 | 45 | 47 | 51 | 54 | 58 | 144 | 9 |
| 10 | 33 | 37 | 39 | 42 | 45 | 49 | 126 | 42 | 47 | 49 | 53 | 56 | 60 | 152 | 10 |
| 11 | 34 | 38 | 40 | 44 | 47 | 51 | 133 | 44 | 49 | 51 | 55 | 59 | 63 | 160 | 11 |
| 12 | 35 | 40 | 42 | 46 | 49 | 54 | 140 | 45 | 51 | 53 | 58 | 62 | 66 | 168 | 12 |
| 13 | 36 | 41 | 44 | 48 | 52 | 56 | 147 | 47 | 53 | 56 | 60 | 64 | 69 | 176 | 13 |
| 14 | 37 | 43 | 45 | 50 | 54 | 59 | 154 | 48 | 54 | 58 | 62 | 67 | 72 | 184 | 14 |
| 15 | 38 | 44 | 47 | 52 | 56 | 61 | 161 | 50 | 56 | 60 | 65 | 69 | 75 | 192 | 15 |
| 16 | 39 | 46 | 49 | 54 | 58 | 64 | 168 | 51 | 58 | 62 | 67 | 72 | 78 | 200 | 16 |
| 17 | 41 | 47 | 51 | 56 | 61 | 66 | 175 | 53 | 60 | 64 | 70 | 75 | 81 | 208 | 17 |
| 18 | 42 | 49 | 52 | 58 | 63 | 69 | 182 | 54 | 62 | 66 | 72 | 77 | 84 | 216 | 18 |
| 19 | 43 | 50 | 54 | 60 | 65 | 71 | 189 | 56 | 64 | 68 | 74 | 80 | 87 | 224 | 19 |
| 20 | 44 | 52 | 56 | 62 | 67 | 74 | 196 | 57 | 66 | 70 | 77 | 83 | 90 | 232 | 20 |
| 21 | 46 | 53 | 58 | 64 | 69 | 76 | 203 | 59 | 68 | 72 | 79 | 85 | 92 | 240 | 21 |
| 22 | 47 | 55 | 59 | 66 | 72 | 79 | 210 | 60 | 70 | 74 | 81 | 88 | 95 | 248 | 22 |
| 23 | 48 | 57 | 61 | 68 | 74 | 81 | 217 | 62 | 71 | 76 | 84 | 90 | 98 | 256 | 23 |
| 24 | 49 | 58 | 63 | 70 | 76 | 84 | 224 | 64 | 73 | 78 | 86 | 93 | 101 | 264 | 24 |
| 25 | 50 | 60 | 64 | 72 | 78 | 86 | 231 | 65 | 75 | 81 | 89 | 96 | 104 | 272 | 25 |

| | $N_1 = 9$ | | | | | | | $N_1 = 10$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | 0.001 | 0.005 | 0.010 | 0.025 | 0.05 | 0.10 | $2\overline{W}$ | $N_2$ |
| 9 | 52 | 56 | 59 | 62 | 66 | 70 | 171 | | | | | | | | |
| 10 | 53 | 58 | 61 | 65 | 69 | 73 | 180 | 65 | 71 | 74 | 78 | 82 | 87 | 210 | 10 |
| 11 | 55 | 61 | 63 | 68 | 72 | 76 | 189 | 67 | 73 | 77 | 81 | 86 | 91 | 220 | 11 |
| 12 | 57 | 63 | 66 | 71 | 75 | 80 | 198 | 69 | 76 | 79 | 84 | 89 | 94 | 230 | 12 |
| 13 | 59 | 65 | 68 | 73 | 78 | 83 | 207 | 72 | 79 | 82 | 88 | 92 | 98 | 240 | 13 |
| 14 | 60 | 67 | 71 | 76 | 81 | 86 | 216 | 74 | 81 | 85 | 91 | 96 | 102 | 250 | 14 |
| 15 | 62 | 69 | 73 | 79 | 84 | 90 | 225 | 76 | 84 | 88 | 94 | 99 | 106 | 260 | 15 |
| 16 | 64 | 72 | 76 | 82 | 87 | 93 | 234 | 78 | 86 | 91 | 97 | 103 | 109 | 270 | 16 |
| 17 | 66 | 74 | 78 | 84 | 90 | 97 | 243 | 80 | 89 | 93 | 100 | 106 | 113 | 280 | 17 |
| 18 | 68 | 76 | 81 | 87 | 93 | 100 | 252 | 82 | 92 | 96 | 103 | 110 | 117 | 290 | 18 |
| 19 | 70 | 78 | 83 | 90 | 96 | 103 | 261 | 84 | 94 | 99 | 107 | 113 | 121 | 300 | 19 |
| 20 | 71 | 81 | 85 | 93 | 99 | 107 | 270 | 87 | 97 | 102 | 110 | 117 | 125 | 310 | 20 |
| 21 | 73 | 83 | 88 | 95 | 102 | 110 | 279 | 89 | 99 | 105 | 113 | 120 | 128 | 320 | 21 |
| 22 | 75 | 85 | 90 | 98 | 105 | 113 | 288 | 91 | 102 | 108 | 116 | 123 | 132 | 330 | 22 |
| 23 | 77 | 88 | 93 | 101 | 108 | 117 | 297 | 93 | 105 | 110 | 119 | 127 | 136 | 340 | 23 |
| 24 | 79 | 90 | 95 | 104 | 111 | 120 | 306 | 95 | 107 | 113 | 122 | 130 | 140 | 350 | 24 |
| 25 | 81 | 92 | 98 | 107 | 114 | 123 | 315 | 98 | 110 | 116 | 126 | 134 | 144 | 360 | 25 |

*(devam ediyor)*

## Ek *Ws* *(devamı)*

| | $N_1 = 11$ | | | | | | | $N_1 = 12$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 11 | 81 | 87 | 91 | 96 | 100 | 106 | 253 | | | | | | | | |
| 12 | 83 | 90 | 94 | 99 | 104 | 110 | 264 | 98 | 105 | 109 | 115 | 120 | 127 | 300 | 12 |
| 13 | 86 | 93 | 97 | 103 | 108 | 114 | 275 | 101 | 109 | 113 | 119 | 125 | 131 | 312 | 13 |
| 14 | 88 | 96 | 100 | 106 | 112 | 118 | 286 | 103 | 112 | 116 | 123 | 129 | 136 | 324 | 14 |
| 15 | 90 | 99 | 103 | 110 | 116 | 123 | 297 | 106 | 115 | 120 | 127 | 133 | 141 | 336 | 15 |
| 16 | 93 | 102 | 107 | 113 | 120 | 127 | 308 | 109 | 119 | 124 | 131 | 138 | 145 | 348 | 16 |
| 17 | 95 | 105 | 110 | 117 | 123 | 131 | 319 | 112 | 122 | 127 | 135 | 142 | 150 | 360 | 17 |
| 18 | 98 | 108 | 113 | 121 | 127 | 135 | 330 | 115 | 125 | 131 | 139 | 146 | 155 | 372 | 18 |
| 19 | 100 | 111 | 116 | 124 | 131 | 139 | 341 | 118 | 129 | 134 | 143 | 150 | 159 | 384 | 19 |
| 20 | 103 | 114 | 119 | 128 | 135 | 144 | 352 | 120 | 132 | 138 | 147 | 155 | 164 | 396 | 20 |
| 21 | 106 | 117 | 123 | 131 | 139 | 148 | 363 | 123 | 136 | 142 | 151 | 159 | 169 | 408 | 21 |
| 22 | 108 | 120 | 126 | 135 | 143 | 152 | 374 | 126 | 139 | 145 | 155 | 163 | 173 | 420 | 22 |
| 23 | 111 | 123 | 129 | 139 | 147 | 156 | 385 | 129 | 142 | 149 | 159 | 168 | 178 | 432 | 23 |
| 24 | 113 | 126 | 132 | 142 | 151 | 161 | 396 | 132 | 146 | 153 | 163 | 172 | 183 | 444 | 24 |
| 25 | 116 | 129 | 136 | 146 | 155 | 165 | 407 | 135 | 149 | 156 | 167 | 176 | 187 | 456 | 25 |

| | $N_1 = 13$ | | | | | | | $N_1 = 14$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 13 | 117 | 125 | 130 | 136 | 142 | 149 | 351 | | | | | | | | |
| 14 | 120 | 129 | 134 | 141 | 147 | 154 | 364 | 137 | 147 | 152 | 160 | 166 | 174 | 406 | 14 |
| 15 | 123 | 133 | 138 | 145 | 152 | 159 | 377 | 141 | 151 | 156 | 164 | 171 | 179 | 420 | 15 |
| 16 | 126 | 136 | 142 | 150 | 156 | 165 | 390 | 144 | 155 | 161 | 169 | 176 | 185 | 434 | 16 |
| 17 | 129 | 140 | 146 | 154 | 161 | 170 | 403 | 148 | 159 | 165 | 174 | 182 | 190 | 448 | 17 |
| 18 | 133 | 144 | 150 | 158 | 166 | 175 | 416 | 151 | 163 | 170 | 179 | 187 | 196 | 462 | 18 |
| 19 | 136 | 148 | 154 | 163 | 171 | 180 | 429 | 155 | 168 | 174 | 183 | 192 | 202 | 476 | 19 |
| 20 | 139 | 151 | 158 | 167 | 175 | 185 | 442 | 159 | 172 | 178 | 188 | 197 | 207 | 490 | 20 |
| 21 | 142 | 155 | 162 | 171 | 180 | 190 | 455 | 162 | 176 | 183 | 193 | 202 | 213 | 504 | 21 |
| 22 | 145 | 159 | 166 | 176 | 185 | 195 | 468 | 166 | 180 | 187 | 198 | 207 | 218 | 518 | 22 |
| 23 | 149 | 163 | 170 | 180 | 189 | 200 | 481 | 169 | 184 | 192 | 203 | 212 | 224 | 532 | 23 |
| 24 | 152 | 166 | 174 | 185 | 194 | 205 | 494 | 173 | 188 | 196 | 207 | 218 | 229 | 546 | 24 |
| 25 | 155 | 170 | 178 | 189 | 199 | 211 | 507 | 177 | 192 | 200 | 212 | 223 | 235 | 560 | 25 |

| | $N_1 = 15$ | | | | | | | $N_1 = 16$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 15 | 160 | 171 | 176 | 184 | 192 | 200 | 465 | | | | | | | | |
| 16 | 163 | 175 | 181 | 190 | 197 | 206 | 480 | 184 | 196 | 202 | 211 | 219 | 229 | 528 | 16 |
| 17 | 167 | 180 | 186 | 195 | 203 | 212 | 495 | 188 | 201 | 207 | 217 | 225 | 235 | 544 | 17 |
| 18 | 171 | 184 | 190 | 200 | 208 | 218 | 510 | 192 | 206 | 212 | 222 | 231 | 242 | 560 | 18 |
| 19 | 175 | 189 | 195 | 205 | 214 | 224 | 525 | 196 | 210 | 218 | 228 | 237 | 248 | 576 | 19 |
| 20 | 179 | 193 | 200 | 210 | 220 | 230 | 540 | 201 | 215 | 223 | 234 | 243 | 255 | 592 | 20 |
| 21 | 183 | 198 | 205 | 216 | 225 | 236 | 555 | 205 | 220 | 228 | 239 | 249 | 261 | 608 | 21 |
| 22 | 187 | 202 | 210 | 221 | 231 | 242 | 570 | 209 | 225 | 233 | 245 | 255 | 267 | 624 | 22 |
| 23 | 191 | 207 | 214 | 226 | 236 | 248 | 585 | 214 | 230 | 238 | 251 | 261 | 274 | 640 | 23 |
| 24 | 195 | 211 | 219 | 231 | 242 | 254 | 600 | 218 | 235 | 244 | 256 | 267 | 280 | 656 | 24 |
| 25 | 199 | 216 | 224 | 237 | 248 | 260 | 615 | 222 | 240 | 249 | 262 | 273 | 287 | 672 | 25 |

## Ek $W_S$ *(devamı)*

| | $N_1 = 17$ | | | | | | | $N_1 = 18$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 17 | 210 | 223 | 230 | 240 | 249 | 259 | 595 | | | | | | | | |
| 18 | 214 | 228 | 235 | 246 | 255 | 266 | 612 | 237 | 252 | 259 | 270 | 280 | 291 | 666 | 18 |
| 19 | 219 | 234 | 241 | 252 | 262 | 273 | 629 | 242 | 258 | 265 | 277 | 287 | 299 | 684 | 19 |
| 20 | 223 | 239 | 246 | 258 | 268 | 280 | 646 | 247 | 263 | 271 | 283 | 294 | 306 | 702 | 20 |
| 21 | 228 | 244 | 252 | 264 | 274 | 287 | 663 | 252 | 269 | 277 | 290 | 301 | 313 | 720 | 21 |
| 22 | 233 | 249 | 258 | 270 | 281 | 294 | 680 | 257 | 275 | 283 | 296 | 307 | 321 | 738 | 22 |
| 23 | 238 | 255 | 263 | 276 | 287 | 300 | 697 | 262 | 280 | 289 | 303 | 314 | 328 | 756 | 23 |
| 24 | 242 | 260 | 269 | 282 | 294 | 307 | 714 | 267 | 286 | 295 | 309 | 321 | 335 | 774 | 24 |
| 25 | 247 | 265 | 275 | 288 | 300 | 314 | 731 | 273 | 292 | 301 | 316 | 328 | 343 | 792 | 25 |

| | $N_1 = 19$ | | | | | | | $N_1 = 20$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 19 | 267 | 283 | 291 | 303 | 313 | 325 | 741 | | | | | | | | |
| 20 | 272 | 289 | 297 | 309 | 320 | 333 | 760 | 298 | 315 | 324 | 337 | 348 | 361 | 820 | 20 |
| 21 | 277 | 295 | 303 | 316 | 328 | 341 | 779 | 304 | 322 | 331 | 344 | 356 | 370 | 840 | 21 |
| 22 | 283 | 301 | 310 | 323 | 335 | 349 | 798 | 309 | 328 | 337 | 351 | 364 | 378 | 860 | 22 |
| 23 | 288 | 307 | 316 | 330 | 342 | 357 | 817 | 315 | 335 | 344 | 359 | 371 | 386 | 880 | 23 |
| 24 | 294 | 313 | 323 | 337 | 350 | 364 | 836 | 321 | 341 | 351 | 366 | 379 | 394 | 900 | 24 |
| 25 | 299 | 319 | 329 | 344 | 357 | 372 | 855 | 327 | 348 | 358 | 373 | 387 | 403 | 920 | 25 |

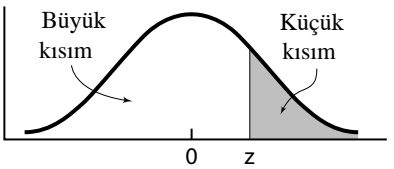
| | $N_1 = 21$ | | | | | | | $N_1 = 22$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 21 | 331 | 349 | 359 | 373 | 385 | 399 | 903 | | | | | | | | |
| 22 | 337 | 356 | 366 | 381 | 393 | 408 | 924 | 365 | 386 | 396 | 411 | 424 | 439 | 990 | 22 |
| 23 | 343 | 363 | 373 | 388 | 401 | 417 | 945 | 372 | 393 | 403 | 419 | 432 | 448 | 1012 | 23 |
| 24 | 349 | 370 | 381 | 396 | 410 | 425 | 966 | 379 | 400 | 411 | 427 | 441 | 457 | 1034 | 24 |
| 25 | 356 | 377 | 388 | 404 | 418 | 434 | 987 | 385 | 408 | 419 | 435 | 450 | 467 | 1056 | 25 |

| | $N_1 = 23$ | | | | | | | $N_1 = 24$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ | $N_2$ |
| 23 | 402 | 424 | 434 | 451 | 465 | 481 | 1081 | | | | | | | | |
| 24 | 409 | 431 | 443 | 459 | 474 | 491 | 1104 | 440 | 464 | 475 | 492 | 507 | 525 | 1176 | 24 |
| 25 | 416 | 439 | 451 | 468 | 483 | 500 | 1127 | 448 | 472 | 484 | 501 | 517 | 535 | 1200 | 25 |

| | $N_1 = 25$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $N_2$ | *0.001* | *0.005* | *0.010* | *0.025* | *0.05* | *0.10* | $2\overline{W}$ |
| 25 | 480 | 505 | 517 | 536 | 552 | 570 | 1275 |

*Source:* L. R. Verdooren (1963), (Tablo 1 içinde) Extended tables of critical values for Wilcoxon's test statistic, *Biometrika, 50,* 177–186, yazar ve editörün izni ile alınmıştır.

## Ek *z*: Normal Dağılım (*z*)




| *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* | *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* |
|---|---|---|---|---|---|---|---|---|---|
| 0,00 | 0,0000 | 0,5000 | 0,5000 | 0,3989 | 0,36 | 0,1406 | 0,6406 | 0,3594 | 0,3739 |
| 0,01 | 0,0040 | 0,5040 | 0,4960 | 0,3989 | 0,37 | 0,1443 | 0,6443 | 0,3557 | 0,3725 |
| 0,02 | 0,0080 | 0,5080 | 0,4920 | 0,3989 | 0,38 | 0,1480 | 0,6480 | 0,3520 | 0,3712 |
| 0,03 | 0,0120 | 0,5120 | 0,4880 | 0,3988 | 0,39 | 0,1517 | 0,6517 | 0,3483 | 0,3697 |
| 0,04 | 0,0160 | 0,5160 | 0,4840 | 0,3986 | 0,40 | 0,1554 | 0,6554 | 0,3446 | 0,3683 |
| 0,05 | 0,0199 | 0,5199 | 0,4801 | 0,3984 | 0,41 | 0,1591 | 0,6591 | 0,3409 | 0,3668 |
| 0,06 | 0,0239 | 0,5239 | 0,4761 | 0,3982 | 0,42 | 0,1628 | 0,6628 | 0,3372 | 0,3653 |
| 0,07 | 0,0279 | 0,5279 | 0,4721 | 0,3980 | 0,43 | 0,1664 | 0,6664 | 0,3336 | 0,3637 |
| 0,08 | 0,0319 | 0,5319 | 0,4681 | 0,3977 | 0,44 | 0,1700 | 0,6700 | 0,3300 | 0,3621 |
| 0,09 | 0,0359 | 0,5359 | 0,4641 | 0,3973 | 0,45 | 0,1736 | 0,6736 | 0,3264 | 0,3605 |
| 0,10 | 0,0398 | 0,5398 | 0,4602 | 0,3970 | 0,46 | 0,1772 | 0,6772 | 0,3228 | 0,3589 |
| 0,11 | 0,0438 | 0,5438 | 0,4562 | 0,3965 | 0,47 | 0,1808 | 0,6808 | 0,3192 | 0,3572 |
| 0,12 | 0,0478 | 0,5478 | 0,4522 | 0,3961 | 0,48 | 0,1844 | 0,6844 | 0,3156 | 0,3555 |
| 0,13 | 0,0517 | 0,5517 | 0,4483 | 0,3956 | 0,49 | 0,1879 | 0,6879 | 0,3121 | 0,3538 |
| 0,14 | 0,0557 | 0,5557 | 0,4443 | 0,3951 | 0,50 | 0,1915 | 0,6915 | 0,3085 | 0,3521 |
| 0,15 | 0,0596 | 0,5596 | 0,4404 | 0,3945 | 0,51 | 0,1950 | 0,6950 | 0,3050 | 0,3503 |
| 0,16 | 0,0636 | 0,5636 | 0,4364 | 0,3939 | 0,52 | 0,1985 | 0,6985 | 0,3015 | 0,3485 |
| 0,17 | 0,0675 | 0,5675 | 0,4325 | 0,3932 | 0,53 | 0,2019 | 0,7019 | 0,2981 | 0,3467 |
| 0,18 | 0,0714 | 0,5714 | 0,4286 | 0,3925 | 0,54 | 0,2054 | 0,7054 | 0,2946 | 0,3448 |
| 0,19 | 0,0753 | 0,5753 | 0,4247 | 0,3918 | 0,55 | 0,2088 | 0,7088 | 0,2912 | 0,3429 |
| 0,20 | 0,0793 | 0,5793 | 0,4207 | 0,3910 | 0,56 | 0,2123 | 0,7123 | 0,2877 | 0,3410 |
| 0,21 | 0,0832 | 0,5832 | 0,4168 | 0,3902 | 0,57 | 0,2157 | 0,7157 | 0,2843 | 0,3391 |
| 0,22 | 0,0871 | 0,5871 | 0,4129 | 0,3894 | 0,58 | 0,2190 | 0,7190 | 0,2810 | 0,3372 |
| 0,23 | 0,0910 | 0,5910 | 0,4090 | 0,3885 | 0,59 | 0,2224 | 0,7224 | 0,2776 | 0,3352 |
| 0,24 | 0,0948 | 0,5948 | 0,4052 | 0,3876 | 0,60 | 0,2257 | 0,7257 | 0,2743 | 0,3332 |
| 0,25 | 0,0987 | 0,5987 | 0,4013 | 0,3867 | 0,61 | 0,2291 | 0,7291 | 0,2709 | 0,3312 |
| 0,26 | 0,1026 | 0,6026 | 0,3974 | 0,3857 | 0,62 | 0,2324 | 0,7324 | 0,2676 | 0,3292 |
| 0,27 | 0,1064 | 0,6064 | 0,3936 | 0,3847 | 0,63 | 0,2357 | 0,7357 | 0,2643 | 0,3271 |
| 0,28 | 0,1103 | 0,6103 | 0,3897 | 0,3836 | 0,64 | 0,2389 | 0,7389 | 0,2611 | 0,3251 |
| 0,29 | 0,1141 | 0,6141 | 0,3859 | 0,3825 | 0,65 | 0,2422 | 0,7422 | 0,2578 | 0,3230 |
| 0,30 | 0,1179 | 0,6179 | 0,3821 | 0,3814 | 0,66 | 0,2454 | 0,7454 | 0,2546 | 0,3209 |
| 0,31 | 0,1217 | 0,6217 | 0,3783 | 0,3802 | 0,67 | 0,2486 | 0,7486 | 0,2514 | 0,3187 |
| 0,32 | 0,1255 | 0,6255 | 0,3745 | 0,3790 | 0,68 | 0,2517 | 0,7517 | 0,2483 | 0,3166 |
| 0,33 | 0,1293 | 0,6293 | 0,3707 | 0,3778 | 0,69 | 0,2549 | 0,7549 | 0,2451 | 0,3144 |
| 0,34 | 0,1331 | 0,6331 | 0,3669 | 0,3765 | 0,70 | 0,2580 | 0,7580 | 0,2420 | 0,3123 |
| 0,35 | 0,1368 | 0,6368 | 0,3632 | 0,3752 | 0,71 | 0,2611 | 0,7611 | 0,2389 | 0,3101 |

## Ek *z* *(devamı)*

| *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* | *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* |
|---|---|---|---|---|---|---|---|---|---|
| 0,72 | 0,2642 | 0,7642 | 0,2358 | 0,3079 | 1,14 | 0,3729 | 0,8729 | 0,1271 | 0,2083 |
| 0,73 | 0,2673 | 0,7673 | 0,2327 | 0,3056 | 1,15 | 0,3749 | 0,8749 | 0,1251 | 0,2059 |
| 0,74 | 0,2704 | 0,7704 | 0,2296 | 0,3034 | 1,16 | 0,3770 | 0,8770 | 0,1230 | 0,2036 |
| 0,75 | 0,2734 | 0,7734 | 0,2266 | 0,3011 | 1,17 | 0,3790 | 0,8790 | 0,1210 | 0,2012 |
| 0,76 | 0,2764 | 0,7764 | 0,2236 | 0,2989 | 1,18 | 0,3810 | 0,8810 | 0,1190 | 0,1989 |
| 0,77 | 0,2794 | 0,7794 | 0,2206 | 0,2966 | 1,19 | 0,3830 | 0,8830 | 0,1170 | 0,1965 |
| 0,78 | 0,2823 | 0,7823 | 0,2177 | 0,2943 | 1,20 | 0,3849 | 0,8849 | 0,1151 | 0,1942 |
| 0,79 | 0,2852 | 0,7852 | 0,2148 | 0,2920 | 1,21 | 0,3869 | 0,8869 | 0,1131 | 0,1919 |
| 0,80 | 0,2881 | 0,7881 | 0,2119 | 0,2897 | 1,22 | 0,3888 | 0,8888 | 0,1112 | 0,1895 |
| 0,81 | 0,2910 | 0,7910 | 0,2090 | 0,2874 | 1,23 | 0,3907 | 0,8907 | 0,1093 | 0,1872 |
| 0,82 | 0,2939 | 0,7939 | 0,2061 | 0,2850 | 1,24 | 0,3925 | 0,8925 | 0,1075 | 0,1849 |
| 0,83 | 0,2967 | 0,7967 | 0,2033 | 0,2827 | 1,25 | 0,3944 | 0,8944 | 0,1056 | 0,1826 |
| 0,84 | 0,2995 | 0,7995 | 0,2005 | 0,2803 | 1,26 | 0,3962 | 0,8962 | 0,1038 | 0,1804 |
| 0,85 | 0,3023 | 0,8023 | 0,1977 | 0,2780 | 1,27 | 0,3980 | 0,8980 | 0,1020 | 0,1781 |
| 0,86 | 0,3051 | 0,8051 | 0,1949 | 0,2756 | 1,28 | 0,3997 | 0,8997 | 0,1003 | 0,1758 |
| 0,87 | 0,3078 | 0,8078 | 0,1922 | 0,2732 | 1,29 | 0,4015 | 0,9015 | 0,0985 | 0,1736 |
| 0,88 | 0,3106 | 0,8106 | 0,1894 | 0,2709 | 1,30 | 0,4032 | 0,9032 | 0,0968 | 0,1714 |
| 0,89 | 0,3133 | 0,8133 | 0,1867 | 0,2685 | 1,31 | 0,4049 | 0,9049 | 0,0951 | 0,1691 |
| 0,90 | 0,3159 | 0,8159 | 0,1841 | 0,2661 | 1,32 | 0,4066 | 0,9066 | 0,0934 | 0,1669 |
| 0,91 | 0,3186 | 0,8186 | 0,1814 | 0,2637 | 1,33 | 0,4082 | 0,9082 | 0,0918 | 0,1647 |
| 0,92 | 0,3212 | 0,8212 | 0,1788 | 0,2613 | 1,34 | 0,4099 | 0,9099 | 0,0901 | 0,1626 |
| 0,93 | 0,3238 | 0,8238 | 0,1762 | 0,2589 | 1,35 | 0,4115 | 0,9115 | 0,0885 | 0,1604 |
| 0,94 | 0,3264 | 0,8264 | 0,1736 | 0,2565 | 1,36 | 0,4131 | 0,9131 | 0,0869 | 0,1582 |
| 0,95 | 0,3289 | 0,8289 | 0,1711 | 0,2541 | 1,37 | 0,4147 | 0,9147 | 0,0853 | 0,1561 |
| 0,96 | 0,3315 | 0,8315 | 0,1685 | 0,2516 | 1,38 | 0,4162 | 0,9162 | 0,0838 | 0,1539 |
| 0,97 | 0,3340 | 0,8340 | 0,1660 | 0,2492 | 1,39 | 0,4177 | 0,9177 | 0,0823 | 0,1518 |
| 0,98 | 0,3365 | 0,8365 | 0,1635 | 0,2468 | 1,40 | 0,4192 | 0,9192 | 0,0808 | 0,1497 |
| 0,99 | 0,3389 | 0,8389 | 0,1611 | 0,2444 | 1,41 | 0,4207 | 0,9207 | 0,0793 | 0,1476 |
| 1,00 | 0,3413 | 0,8413 | 0,1587 | 0,2420 | 1,42 | 0,4222 | 0,9222 | 0,0778 | 0,1456 |
| 1,01 | 0,3438 | 0,8438 | 0,1562 | 0,2396 | 1,43 | 0,4236 | 0,9236 | 0,0764 | 0,1435 |
| 1,02 | 0,3461 | 0,8461 | 0,1539 | 0,2371 | 1,44 | 0,4251 | 0,9251 | 0,0749 | 0,1415 |
| 1,03 | 0,3485 | 0,8485 | 0,1515 | 0,2347 | 1,45 | 0,4265 | 0,9265 | 0,0735 | 0,1394 |
| 1,04 | 0,3508 | 0,8508 | 0,1492 | 0,2323 | 1,46 | 0,4279 | 0,9279 | 0,0721 | 0,1374 |
| 1,05 | 0,3531 | 0,8531 | 0,1469 | 0,2299 | 1,47 | 0,4292 | 0,9292 | 0,0708 | 0,1354 |
| 1,06 | 0,3554 | 0,8554 | 0,1446 | 0,2275 | 1,48 | 0,4306 | 0,9306 | 0,0694 | 0,1334 |
| 1,07 | 0,3577 | 0,8577 | 0,1423 | 0,2251 | 1,49 | 0,4319 | 0,9319 | 0,0681 | 0,1315 |
| 1,08 | 0,3599 | 0,8599 | 0,1401 | 0,2227 | 1,50 | 0,4332 | 0,9332 | 0,0668 | 0,1295 |
| 1,09 | 0,3621 | 0,8621 | 0,1379 | 0,2203 | 1,51 | 0,4345 | 0,9345 | 0,0655 | 0,1276 |
| 1,10 | 0,3643 | 0,8643 | 0,1357 | 0,2179 | 1,52 | 0,4357 | 0,9357 | 0,0643 | 0,1257 |
| 1,11 | 0,3665 | 0,8665 | 0,1335 | 0,2155 | 1,53 | 0,4370 | 0,9370 | 0,0630 | 0,1238 |
| 1,12 | 0,3686 | 0,8686 | 0,1314 | 0,2131 | 1,54 | 0,4382 | 0,9382 | 0,0618 | 0,1219 |
| 1,13 | 0,3708 | 0,8708 | 0,1292 | 0,2107 | 1,55 | 0,4394 | 0,9394 | 0,0606 | 0,1200 |

*(devam ediyor)*

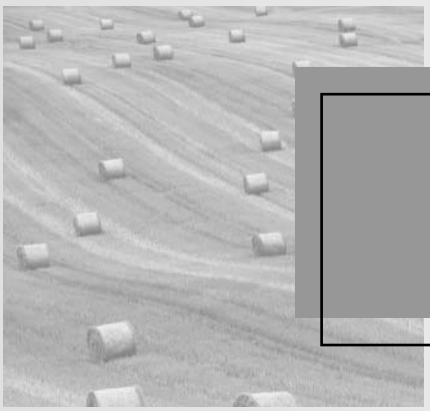
## Ek *z* *(devamı)*

| *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* | *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* |
|---|---|---|---|---|---|---|---|---|---|
| 1,56 | 0,4406 | 0,9406 | 0,0594 | 0,1182 | 1,98 | 0,4761 | 0,9761 | 0,0239 | 0,0562 |
| 1,57 | 0,4418 | 0,9418 | 0,0582 | 0,1163 | 1,99 | 0,4767 | 0,9767 | 0,0233 | 0,0551 |
| 1,58 | 0,4429 | 0,9429 | 0,0571 | 0,1145 | 2,00 | 0,4772 | 0,9772 | 0,0228 | 0,0540 |
| 1,59 | 0,4441 | 0,9441 | 0,0559 | 0,1127 | 2,01 | 0,4778 | 0,9778 | 0,0222 | 0,0529 |
| 1,60 | 0,4452 | 0,9452 | 0,0548 | 0,1109 | 2,02 | 0,4783 | 0,9783 | 0,0217 | 0,0519 |
| 1,61 | 0,4463 | 0,9463 | 0,0537 | 0,1092 | 2,03 | 0,4788 | 0,9788 | 0,0212 | 0,0508 |
| 1,62 | 0,4474 | 0,9474 | 0,0526 | 0,1074 | 2,04 | 0,4793 | 0,9793 | 0,0207 | 0,0498 |
| 1,63 | 0,4484 | 0,9484 | 0,0516 | 0,1057 | 2,05 | 0,4798 | 0,9798 | 0,0202 | 0,0488 |
| 1,64 | 0,4495 | 0,9495 | 0,0505 | 0,1040 | 2,06 | 0,4803 | 0,9803 | 0,0197 | 0,0478 |
| 1,65 | 0,4505 | 0,9505 | 0,0495 | 0,1023 | 2,07 | 0,4808 | 0,9808 | 0,0192 | 0,0468 |
| 1,66 | 0,4515 | 0,9515 | 0,0485 | 0,1006 | 2,08 | 0,4812 | 0,9812 | 0,0188 | 0,0459 |
| 1,67 | 0,4525 | 0,9525 | 0,0475 | 0,0989 | 2,09 | 0,4817 | 0,9817 | 0,0183 | 0,0449 |
| 1,68 | 0,4535 | 0,9535 | 0,0465 | 0,0973 | 2,10 | 0,4821 | 0,9821 | 0,0179 | 0,0440 |
| 1,69 | 0,4545 | 0,9545 | 0,0455 | 0,0957 | 2,11 | 0,4826 | 0,9826 | 0,0174 | 0,0431 |
| 1,70 | 0,4554 | 0,9554 | 0,0446 | 0,0940 | 2,12 | 0,4830 | 0,9830 | 0,0170 | 0,0422 |
| 1,71 | 0,4564 | 0,9564 | 0,0436 | 0,0925 | 2,13 | 0,4834 | 0,9834 | 0,0166 | 0,0413 |
| 1,72 | 0,4573 | 0,9573 | 0,0427 | 0,0909 | 2,14 | 0,4838 | 0,9838 | 0,0162 | 0,0404 |
| 1,73 | 0,4582 | 0,9582 | 0,0418 | 0,0893 | 2,15 | 0,4842 | 0,9842 | 0,0158 | 0,0396 |
| 1,74 | 0,4591 | 0,9591 | 0,0409 | 0,0878 | 2,16 | 0,4846 | 0,9846 | 0,0154 | 0,0387 |
| 1,75 | 0,4599 | 0,9599 | 0,0401 | 0,0863 | 2,17 | 0,4850 | 0,9850 | 0,0150 | 0,0379 |
| 1,76 | 0,4608 | 0,9608 | 0,0392 | 0,0848 | 2,18 | 0,4854 | 0,9854 | 0,0146 | 0,0371 |
| 1,77 | 0,4616 | 0,9616 | 0,0384 | 0,0833 | 2,19 | 0,4857 | 0,9857 | 0,0143 | 0,0363 |
| 1,78 | 0,4625 | 0,9625 | 0,0375 | 0,0818 | 2,20 | 0,4861 | 0,9861 | 0,0139 | 0,0355 |
| 1,79 | 0,4633 | 0,9633 | 0,0367 | 0,0804 | 2,21 | 0,4864 | 0,9864 | 0,0136 | 0,0347 |
| 1,80 | 0,4641 | 0,9641 | 0,0359 | 0,0790 | 2,22 | 0,4868 | 0,9868 | 0,0132 | 0,0339 |
| 1,81 | 0,4649 | 0,9649 | 0,0351 | 0,0775 | 2,23 | 0,4871 | 0,9871 | 0,0129 | 0,0332 |
| 1,82 | 0,4656 | 0,9656 | 0,0344 | 0,0761 | 2,24 | 0,4875 | 0,9875 | 0,0125 | 0,0325 |
| 1,83 | 0,4664 | 0,9664 | 0,0336 | 0,0748 | 2,25 | 0,4878 | 0,9878 | 0,0122 | 0,0317 |
| 1,84 | 0,4671 | 0,9671 | 0,0329 | 0,0734 | 2,26 | 0,4881 | 0,9881 | 0,0119 | 0,0310 |
| 1,85 | 0,4678 | 0,9678 | 0,0322 | 0,0721 | 2,27 | 0,4884 | 0,9884 | 0,0116 | 0,0303 |
| 1,86 | 0,4686 | 0,9686 | 0,0314 | 0,0707 | 2,28 | 0,4887 | 0,9887 | 0,0113 | 0,0297 |
| 1,87 | 0,4693 | 0,9693 | 0,0307 | 0,0694 | 2,29 | 0,4890 | 0,9890 | 0,0110 | 0,0290 |
| 1,88 | 0,4699 | 0,9699 | 0,0301 | 0,0681 | 2,30 | 0,4893 | 0,9893 | 0,0107 | 0,0283 |
| 1,89 | 0,4706 | 0,9706 | 0,0294 | 0,0669 | 2,31 | 0,4896 | 0,9896 | 0,0104 | 0,0277 |
| 1,90 | 0,4713 | 0,9713 | 0,0287 | 0,0656 | 2,32 | 0,4898 | 0,9898 | 0,0102 | 0,0270 |
| 1,91 | 0,4719 | 0,9719 | 0,0281 | 0,0644 | 2,33 | 0,4901 | 0,9901 | 0,0099 | 0,0264 |
| 1,92 | 0,4726 | 0,9726 | 0,0274 | 0,0632 | 2,34 | 0,4904 | 0,9904 | 0,0096 | 0,0258 |
| 1,93 | 0,4732 | 0,9732 | 0,0268 | 0,0620 | 2,35 | 0,4906 | 0,9906 | 0,0094 | 0,0252 |
| 1,94 | 0,4738 | 0,9738 | 0,0262 | 0,0608 | 2,36 | 0,4909 | 0,9909 | 0,0091 | 0,0246 |
| 1,95 | 0,4744 | 0,9744 | 0,0256 | 0,0596 | 2,37 | 0,4911 | 0,9911 | 0,0089 | 0,0241 |
| 1,96 | 0,4750 | 0,9750 | 0,0250 | 0,0584 | 2,38 | 0,4913 | 0,9913 | 0,0087 | 0,0235 |
| 1,97 | 0,4756 | 0,9756 | 0,0244 | 0,0573 | 2,39 | 0,4916 | 0,9916 | 0,0084 | 0,0229 |

## Ek *z* *(devamı)*

| *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* | *z* | Ortalamadan *z*'ye kadar | Büyük kısım | Küçük kısım | *y* |
|---|---|---|---|---|---|---|---|---|---|
| 2,40 | 0,4918 | 0,9918 | 0,0082 | 0,0224 | 2,75 | 0,4970 | 0,9970 | 0,0030 | 0,0091 |
| 2,41 | 0,4920 | 0,9920 | 0,0080 | 0,0219 | 2,76 | 0,4971 | 0,9971 | 0,0029 | 0,0088 |
| 2,42 | 0,4922 | 0,9922 | 0,0078 | 0,0213 | 2,77 | 0,4972 | 0,9972 | 0,0028 | 0,0086 |
| 2,43 | 0,4925 | 0,9925 | 0,0075 | 0,0208 | 2,78 | 0,4973 | 0,9973 | 0,0027 | 0,0084 |
| 2,44 | 0,4927 | 0,9927 | 0,0073 | 0,0203 | 2,79 | 0,4974 | 0,9974 | 0,0026 | 0,0081 |
| 2,45 | 0,4929 | 0,9929 | 0,0071 | 0,0198 | 2,80 | 0,4974 | 0,9974 | 0,0026 | 0,0079 |
| 2,46 | 0,4931 | 0,9931 | 0,0069 | 0,0194 | 2,81 | 0,4975 | 0,9975 | 0,0025 | 0,0077 |
| 2,47 | 0,4932 | 0,9932 | 0,0068 | 0,0189 | 2,82 | 0,4976 | 0,9976 | 0,0024 | 0,0075 |
| 2,48 | 0,4934 | 0,9934 | 0,0066 | 0,0184 | 2,83 | 0,4977 | 0,9977 | 0,0023 | 0,0073 |
| 2,49 | 0,4936 | 0,9936 | 0,0064 | 0,0180 | 2,84 | 0,4977 | 0,9977 | 0,0023 | 0,0071 |
| 2,50 | 0,4938 | 0,9938 | 0,0062 | 0,0175 | 2,85 | 0,4978 | 0,9978 | 0,0022 | 0,0069 |
| 2,51 | 0,4940 | 0,9940 | 0,0060 | 0,0171 | 2,86 | 0,4979 | 0,9979 | 0,0021 | 0,0067 |
| 2,52 | 0,4941 | 0,9941 | 0,0059 | 0,0167 | 2,87 | 0,4979 | 0,9979 | 0,0021 | 0,0065 |
| 2,53 | 0,4943 | 0,9943 | 0,0057 | 0,0163 | 2,88 | 0,4980 | 0,9980 | 0,0020 | 0,0063 |
| 2,54 | 0,4945 | 0,9945 | 0,0055 | 0,0158 | 2,89 | 0,4981 | 0,9981 | 0,0019 | 0,0061 |
| 2,55 | 0,4946 | 0,9946 | 0,0054 | 0,0154 | 2,90 | 0,4981 | 0,9981 | 0,0019 | 0,0060 |
| 2,56 | 0,4948 | 0,9948 | 0,0052 | 0,0151 | 2,91 | 0,4982 | 0,9982 | 0,0018 | 0,0058 |
| 2,57 | 0,4949 | 0,9949 | 0,0051 | 0,0147 | 2,92 | 0,4982 | 0,9982 | 0,0018 | 0,0056 |
| 2,58 | 0,4951 | 0,9951 | 0,0049 | 0,0143 | 2,93 | 0,4983 | 0,9983 | 0,0017 | 0,0055 |
| 2,59 | 0,4952 | 0,9952 | 0,0048 | 0,0139 | 2,94 | 0,4984 | 0,9984 | 0,0016 | 0,0053 |
| 2,60 | 0,4953 | 0,9953 | 0,0047 | 0,0136 | 2,95 | 0,4984 | 0,9984 | 0,0016 | 0,0051 |
| 2,61 | 0,4955 | 0,9955 | 0,0045 | 0,0132 | 2,96 | 0,4985 | 0,9985 | 0,0015 | 0,0050 |
| 2,62 | 0,4956 | 0,9956 | 0,0044 | 0,0129 | 2,97 | 0,4985 | 0,9985 | 0,0015 | 0,0048 |
| 2,63 | 0,4957 | 0,9957 | 0,0043 | 0,0126 | 2,98 | 0,4986 | 0,9986 | 0,0014 | 0,0047 |
| 2,64 | 0,4959 | 0,9959 | 0,0041 | 0,0122 | 2,99 | 0,4986 | 0,9986 | 0,0014 | 0,0046 |
| 2,65 | 0,4960 | 0,9960 | 0,0040 | 0,0119 | 3,00 | 0,4987 | 0,9987 | 0,0013 | 0,0044 |
| 2,66 | 0,4961 | 0,9961 | 0,0039 | 0,0116 | . . . | . . . | . . . | . . . | . . . |
| 2,67 | 0,4962 | 0,9962 | 0,0038 | 0,0113 | 3,25 | 0,4994 | 0,9994 | 0,0006 | 0,0020 |
| 2,68 | 0,4963 | 0,9963 | 0,0037 | 0,0110 | . . . | . . . | . . . | . . . | . . . |
| 2,69 | 0,4964 | 0,9964 | 0,0036 | 0,0107 | 3,50 | 0,4998 | 0,9998 | 0,0002 | 0,0009 |
| 2,70 | 0,4965 | 0,9965 | 0,0035 | 0,0104 | . . . | . . . | . . . | . . . | . . . |
| 2,71 | 0,4966 | 0,9966 | 0,0034 | 0,0101 | 3,75 | 0,4999 | 0,9999 | 0,0001 | 0,0004 |
| 2,72 | 0,4967 | 0,9967 | 0,0033 | 0,0099 | . . . | . . . | . . . | . . . | . . . |
| 2,73 | 0,4968 | 0,9968 | 0,0032 | 0,0096 | 4,00 | 0,5000 | 1,0000 | 0,0000 | 0,0001 |
| 2,74 | 0,4969 | 0,9969 | 0,0031 | 0,0093 | . . . | . . . | . . . | . . . | . . . |

*Kaynak:* Bu tablodaki sayılar kitabın yazarı tarafından üretilmiştir.



# Kaynaklar


Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile.* Burlington. VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., Howell, C. T., Aoki, M. F., & Rauh, V. A. (1993). Nine-year outcome of the Vermont Intervention Program for low-birthweight infants. *Pediatrics, 91*, 45–55.

Adams, H. E., Wright, L. W. (Jr.), & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology, 105*, 440–445.

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York: Wiley.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods, 10*, 317–328.

Allison, P. D. (1999). *Logistic regression using the SAS® System: Theory and Application*. Cary, NC: SAS Institute, Inc.

Anderson, N. H. (1963). Comparison of different populations: Resistance to extinction and transfer. *Psychological Review, 70,* 162–179.

Appelbaum, M. I., & Cramer, E. M. (1974). Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin, 81*, 335–343.

Appleton, D. R., French, J. M., & Vanderpump, M. P. J. (1996). Ignoring a covariate: An example of Simpson's paradox. *The American Statistician, 50,* 340–341.

Arnholt, A. T. (2007). Resampling with R. *Teaching Statistics, 29*(1), 21–26.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1998). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.

Baker, S. P., & Lew, R. (1987). A Monte Carlo comparison of Tukey's and Bonferroni's methods with an optimized Bonferroni multiple comparison procedure in repeated measures experiments. Paper presented at the annual meeting of the American Statistical Association. San Francisco. CA.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bauer, M., & Dopfmer, S. (1999). Lithium augmentation in treatment-resistant depression, meta-analysis of placebo-controlled studies. *Journal of Clinical Psychopharmacology*, 19, 427–434.

Beaujean, A. A. (2008) Mediation, moderation, and the study of individual differences. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 341–357). Thousand Oaks, CA, Sage, 2008.

Bell, R. A., Buerkel-Rothfuss, N. L., & Gore, K. E. (1987). "Did you bring the yarmulke for the Cabbage Patch kid?": The idiomatic communication among young lovers. *Human Communication Research, 14*, 47–67.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57*, 289–300.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics, 25*, 60–83.

Berger, V. I. (2005). Mid-p values. In B. S. Everitt & D. C Howell, *Encyclopedia of Statistics in Behavioral Sciences*. Chichester, England: Wiley.

Bhoj, D. S. (1978). Testing equality of means of correlated variates with missing data on both responses. *Biometrika 65,* 225–228.

Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley, *Science, 187*, 398–403.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge. MA: MIT Press.

Bisson J., & Andrew, M. (2007). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews* 2007, Issue 3. Art. No.: CD003388. DOI: 10.1002/14651858.CD003388.pub3.m

Blair, R. C. (1978). I've been testing some statistical hypotheses. . . . Can you guess what they are? *Journal of Educational Research, 72,* 116–118.

Blair, R. C., & Higgins, J. J. (1978). Tests of hypotheses for unbalanced factorial designs under various regression/coding method combinations. *Educational and Psychological Measurement, 38*, 621–631.

Blair, R. C., & Higgins, J. J. (1980). A comparison of the power of Wilcoxon's rank-sum statistic to that of Student's *t* statistic under various nonnormal distributions. *Journal of Educational Statistics, 5*, 309–335.

Blair, R. C., & Higgins, J. J. (1985). Comparison of the power of the paired samples *t* test to that of Wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin, 97,* 119–128.

Blanchard, E. B., Theobald, D. E., Williamson, D. A., Silver, B. V., & Brown, D. A. (1978). Temperature biofeedback in the treatment of migraine headaches. *Archives of General Psychiatry, 35*, 581–588.

Bloch, M. H., Panza, K. E., Landeros-Weisenberger, A., & Leckman, J. F. (2009). Meta-analysis: treatment of attention-deficit/hyperactivity disorder in children with comorbid tic disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*, 884.

Block J. J., & Crain B. R. (2007). Omissions and errors in "media violence and the American public." *The American Psychologist, 62*, 252–253.

Boik, R. J. (1981). A priori tests in repeated measures designs: Effects of nonsphericity. *Psychometrika, 46*, 241–255.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the *t* test. *Psychological Bulletin, 57,* 49–64.

Borenstein, M., & Cohen, J. (1988). *Statistical power analysis: A computer program.* Hillsdale, NJ: Lawrence Erlbaum.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons, Ltd.

Bouton, M., & Swartzentruber, D. (1985). Unpublished raw data, University of Vermont.

Box, G. E. P. (1953). Non-normality and tests on variance. *Biometrika, 40*, 318–335.

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics, 25,* 290–302.

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and of correlation of errors in the two-way classification. *Annals of Mathematical Statistics, 25,* 484–498.

Bradley, D. R., Bradley, T. D., McGrath, S. G., & Cutcomb, S. D. (1979). Type I error rate of the chi-square test of independence in $R \times C$ tables that have small expected frequencies. *Psychological Bulletin, 86*, 1290–1297.

Bradley, D. R. (1988). *DATASIM*. Lewiston ME: Desktop Press.

Bradley, D. R., Russell, R. L., & Reeve, C. P. (1996). Statistical power in complex experimental designs. *Behavioral Research Methods, Instruments, and Computers*, *28,* 319–326.

Bradley, J. V. (1963). *Studies in research methodology: IV A sampling study of the central limit theorem and the robustness of one sample parametric tests* (AMRL Tech. Rep. No. 63–29). Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, OH.

Bradley, J. V. (1964). *Studies in research methodology: VI. The central limit effect for a variety of populations and the robustness of z, t, and F* (AMRL Tech. Rep. No. 64–123). Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, OH.

Bradley, J. V. (1968). *Distribution free statistical tests.* Englewood Cliffs, NJ: Prentice-Hall.

Brand, A., Bradley, M. T., Best, L. A., & Stoica, G. (2008). Accuracy of effect size estimates from published psychological research. *Perceptual Motor Skills, 106*, 645–649.

Brooks, L., & Perot, A. R. (1991). Reporting sexual harassment. *Psychology of Women Quarterly, 15,* 31–47.

Brown, M. B. (1976). Screening effects in multidimensional contingency tables. *Applied Statistics, 25,* 37.

Brown, M. B., & Forsythe, A. B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics, 30,* 719–724.

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods, 40*, 467–478.

Cai, L., & Hayes, A. F. (2008). A new test of linear hypotheses in OLS regression under heteroscedasticity of unknown form. *Journal of Educational and Behavioral Statistics, 33*, 21–40.

Camilli, G., & Hopkins, K. D. (1978). Applicability of chi-square to 2 × 2 contingency tables with small expected cell frequencies. *Psychological Bulletin, 85,* 163–167.

Camilli, G., & Hopkins, K. D. (1979). Testing for association in 2 × 2 contingency tables with very small sample sizes. *Psychological Bulletin, 86,* 1011–1014.

Campbell, A., Converse, P. E., & Rogers, W. L. (1976). *The quality of American life*. New York: Russell Sage Foundation.

Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine, 26*, 3661–3675.

Carmer, S. G., & Swanson, M. R. (1973). An evaluation of ten multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association, 68,* 66–74.

Center, B., Skiba, R., & Casey, A. (1985–1986). Quantitative synthesis of intra-subject design literature. *The Journal of Special Education, 19*, 387–400.

Chen, R., & Dunlap, W. P. (1994). A Monte Carlo study of the performance of a corrected formula for $\tilde{\varepsilon}$ suggested by Lecoutre. *Journal of Educational Statistics, 19*, 119–126.

Christensen, R. (1997). *Log-linear models and logistic regression*. New York: Springer-Verlag.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self in the emergence of racial identification in Negro pre-school children. *Journal of Social Psychology, 10,* 591–599.

Cochran, W. G. (1954). Some methods of strengthening the common chi-square test. *Biometrics, 10*, 417–451.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wiley.

Cochrane, A. I., St. Leger, A. S., & Moore, F. (1978). Health service "input" and mortality "output" in developed countries. *Journal of Epidemiology and Community Health, 32,* 200–205.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65,* 145–153.

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology.* New York: McGraw-Hill.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,* 426–443.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences* (1st ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33,* 107–112.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1992a). A power primer. *Psychological Bulletin, 112,* 155–159.

Cohen, J. (1992b). Statistical power analysis. *Current Directions in Psychological Science, 1,* 98–101.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Cohen, S., Kaplan, J. R., Cunnick, J. E., Manuck, S. B., & Rabin, B. S. (1992). Chronic social stress, affiliation, and cellular immune response in nonhuman primates. *Psychological Science, 3,* 301–304.

Collier, R. O., Jr., Baker, F. B., & Mandeville, G. K. (1967). Tests of hypotheses in a repeated measures design from a permutation viewpoint. *Psychometrika, 32,* 15–24.

Collier, R. O., Jr., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika, 32,* 339–353.

Compas, B. E., Howell, D. C., Phares, V., Williams, R. A., & Giunta, C. (1989). Risk factors for emotional/behavioral problems in young adolescents: A prospective analysis of adolescent and parental stress symptoms. *Journal of Consulting and Clinical Psychology, 57,* 732–740.

Compas, B. E., Worsham, N. S., Grant, K., Mireault, G., Howell, D. C., & Malcarne, V. L. (1994). When mom or dad has cancer: I. Symptoms of depression and anxiety in cancer patients, spouses, and children. *Health Psychology, 13*, 507–515.

Conti, L., & Musty, R. E. (1984). The effects of delta-9tetrahydrocannabinol injections to the nucleus accumbens on the locomotor activity of rats. In S. Aquell et al. (Eds.), *The cannabinoids: Chemical, pharmacologic, and therapeutic Aspects.* New York: Academic Press.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis.* New York: Wiley.

Cooper, H. (2009). *Research synthesis and meta-analysis* (4th ed.). Los Angeles, Sage.

Cooper, H. M., Hedges, L. V., & Valentine J. (Eds.). (2009). *The Handbook of Research Synthesis and Meta-Analysis, 2nd Edition.* New York: The Russell Sage Foundation.

Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs.* Thousand Oaks, CA: Sage.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research, *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Cramer, E. M., & Appelbaum, M. I. (1980). Nonorthogonal analysis of variance—Once again. *Psychological Bulletin, 87,* 51–57.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.

Crawford, J. R., Garthwaite, P. H., & Howell, D. C. (2009). On comparing a single case with a control sample: An alternative perspective. *Neuropsychologica, 47*, 2690–2695.

Crawford, J. R., Garthwaite, P. H., Howell, D. C., & Venneri, A. (2003). Intra–individual measures of association in neuropsychology: Inferential methods for comparing a single case with a control or normative sample. *Journal of the International Neuropsychological Society, 9*, 989–1000.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12,* 482–486.

Crawley, M. J. (2007). *The R book.* Chichester, England: Wiley.

Cumming, G. (2008). Replication and *p* intervals. *Perspectives on Psychological Science*, 3, 286–300.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532–574.

Czitrom, V. (1999). One-factor-at-a-time versus designed experiments, *American Statistician, 53*, 126–131.

Dabbs, J. M., Jr., & Morris, R. (1990). Testosterone, social class, and antisocial behavior in a sample of 4462 men. *Psychological Science, 1*, 209–211.

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit techniques*. New York: Marcel Dekker.

Darley, J. M., & Latané B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*, 377–383.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin, 69*, 161–182.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Davey, G. C. L., Startup, H. M., Zara, A., MacDonald, C. B., & Field, A. P. (2003). The perseveration of checking thoughts and mood-as-input hypothesis. *Journal of Behavior Therapy & Experimental Psychiatry, 34*, 141–160.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95–106.

Delucchi, K. L. (1983). The use and misuse of chi-square: Lewis and Burke revisited. *Psychological Bulletin, 94*, 166–176.

Derogatis, L. R. (1983). *SCL90-R: Administration, scoring, and procedures manual* (vol. 1.). Towson, MD: Clinical Psychometric Research.

Dodd, D. H., & Schultz, R. F., Jr. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance designs. *Psychological Bulletin, 79*, 391–395.

Doob, A. N., & Gross, A. E. (1968). Status of frustrator as an inhibitor of horn-honking responses. *Journal of Social Psychology, 76*, 213–218.

Dracup, C. (2005). Confidence intervals. In B. S. Everitt & D. C. Howell (Eds,), *Encyclopedia of statistics in behavioral sciences*. Chichester, England: Wiley.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: Wiley.

Dunlap, G., & Fox, L. (1999). A demonstration of *behavioral support* for young children with *autism. Journal of Positive Behavioral Intervention, 1*, 77–87.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*, 52–64.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association, 50*, 1096–1121.

Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics, 20*, 482–491.

Dunning, T., & Freedman, D. A. (2008). Modeling selection effects. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp 225–231). London, Sage.

Eckelberry, E. (2009). *Using social stories with children with autism, learning disabilities, and ADHA.* Unpublished Master's Thesis. Ohio University.

Edgington, E. S. (1995). *Randomization tests*. New York: Marcel Dekker.

Edwards, A. L. (1985). *Experimental design in psychological research* (5th ed.). New York: Harper & Row.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the bootstrap.* New York: Chapman and Hall.

Einot, I., & Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association, 70*, 574–583.

Epping-Jordan, J. E., Compas, B. E., & Howell, D. C. (1994). Predictors of cancer progression in young adult men and women: Avoidance, intrusive thoughts, and psychological symptoms. *Health Psychology, 13*, 539–547.

Evans, S. H., & Anastasio, E. J. (1968). Misuse of analysis of covariance when treatment effect and covariate are confounded. Psy*chological Bulletin, 69*, 225–234.

Everitt, B. (1994). Cited in Hand et al. (1994), p. 229.

Everitt, B. S. (2005). *An R and S-PLUS companion to multivariate analysis.* London: Springer-Verlag

Everitt, B., & Hothorn, T. (2006). *A handbook of statistical analyses using R.* London: Chapman & Hall/CRC.

Eysenck, M. W. (1974). Age differences in incidental learning. *Developmental Psychology, 10,* 936–941.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Federer, W. T. (1955). *Experimental design: Theory and application.* New York: Macmillan.

Feinberg, M., & Willer, R. (2011). Apocalypse soon? Dire messages reduce belief in global warming by contradicting just-world beliefs. *Psychological Science, 22,* 34–38.

Fidalgo, A. M. (2005). Mantel-Haenszel methods. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science.* Chichester, England: Wiley.

Finney, J. W., Mitchell, R. E., Cronkite, R. C., & Moos, R. H. (1984). Methodological issues in estimating main and interactive effects: Examples from coping/social support and stress field. *Journal of Health and Social Behavior, 25*, 85–98.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1,* 3–32.

Fisher, R. A. (1935). *The design of experiments.* Edinburgh: Oliver & Boyd.

Fisher, R. A., & Yates, F. (1953). *Statistical tables for biological, agricultural, and medical research* (4th ed.). Edinburgh: Oliver & Boyd.

Fleiss, J. L. (1969). Estimating the magnitude of experimental effects. *Psychological Bulletin, 72*, 273–276.

Fleiss, J. L., Nee, J. C. M., & Landis, J. R. (1979). Large sample variance of kappa in the case of different sets of raters. *Psychological Bulletin, 86,* 974–977.

Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology, 59*, 715–723.

Fowler, R. L. (1985). Point estimates and confidence intervals in measures of association. *Psychological Bulletin, 98*, 160–165.

Fox, J. (2002). *An R and S-PLUS companion to applied regression.* Thousand Oaks, CA: Sage Publications.

Franklin, R. D., Allison, D. B., & Gorman, B. S. (Eds.). (1996). *Design and analysis of single-case Research.* Mahwah, NJ: Lawrence Erlbaum.

Frigge, M., Hoaglin, D. C., & Iglewicz B. (1989). Some implementations of the boxplot. *American Statistician, 43*, 50–54.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15*, 246–263.

Games, P. A. (1978a). A three-factor model encompassing many possible statistical tests on independent groups. *Psychological Bulletin, 85*, 168–182.

Games, P. A. (1978b). A four-factor structure for parametric tests on independent groups. *Psychological Bulletin, 85*, 661–672.

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal *n's* and/or variances: A Monte Carlo study. *Journal of Educational Statistics, 1,* 113–125.

Games, P. A., Keselman, H. J., & Rogan, J. C. (1981). Simultaneous pairwise multiple comparison procedures for means when sample sizes are unequal. *Psychological Bulletin, 90,* 594–598.

Garland, C. F., Garland, F. C., Gorham, E. D., Lipkin, M., Newmark, H., Mohr, S. B., & Holick, M. F. (2006). The role of vitamin D in cancer prevention. *American Journal of Public Health, 96,* 252–261.

Gartrell, N., & Bos, H. (2010). U.S. National Longitudinal Lesbian Family Study: Psychological adjustment of 17-year-old adolescents. *Pediatrics, 126*, 28–36.

Geffen, G., Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B. (1990). Performance measure of 16- to 86-year-old males and females on the Auditory Verbal Learning Test. *The Clinical Neuropsychologist, 4,* 45–63.

Geller, E. S., Witmer, J. F., & Orebaugh, A. L. (1976). Instructions as a determinant of paper disposal behaviors. *Environment and Behavior, 8,* 417–439.

Geyer, C. J. (1991). Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association*, *86,* 717–724.

Giancola, P. R., & Corman, M. D. (2007). Alcohol and aggression: A test of the attention-allocation model. *Psychological Science, 18*, 649–655.

Gibson, L., & Leitenberg, H. (2000). Child sexual abuse prevention programs: Do they decrease the occurrence of child sexual abuse? *Child Abuse & Neglect*, *24*, 1115–1125.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance*. Cambridge: Cambridge University Press.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8.

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills: Sage Publications.

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Goldberg, L. R. (1965). Diagnosticians versus diagnostic signs: The diagnosis of psychosis versus neurosis from the MMPI. *Psychological Monographs, 79* (9, Whole No. 602).

Goldstein, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician, 43,* 253–260.

Good, P. I. (1999). Resampling methods: A practical guide to data analysis. Boston: Birkhäuser.

Good, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses* (2nd ed.). New York: Springer-Verlag.

Gouzoulis-Mayfrank, E., Daumann, J., Tuchtenhagen, F., Pelz, S., Becker, S., Kunert, H. -J., Fimm, B., & Sasa, H. (2000). Impaired cognitive performance in drug free users of recreational ecstasy (MDMA). *Journal of Neurology, Neurosurgery, and Psychiatry, 68*, 719–725.

Green, J. A. (1988). Loglinear analysis of cross-classified ordinal data: Applications in developmental research. *Child Development, 59,* 1–25.

Green, S. B., Marquis, J. G., Hershberger, S. L., Thompson, M. S., & McCollam, K. M. (1999). The overparameterized analysis of variance model. *Psychological Methods, 4*, 214–233.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psvchometrika, 24,* 95–112.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68*, 155–165.

Gross, J. S. (1985). Weight modification and eating disorders in adolescent boys and girls. Unpublished doctoral dissertation, University of Vermont.

Guber, D. L. (1999) Getting what you pay for: The debate over equity in public school expenditures. *Journal of Statistics Education, 7*, No. 2.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., & Ostrowski, E. (1994). *Handbook of small data sets.* London: Chapman & Hall.

Hansen, J. C., & Swanson, J. L. (1983). Stability of interests and the predictive and concurrent validity of the 1981 Strong-Campbell Interest Inventory for college majors. *Journal of Counseling Psychology, 30*, 194–201.

Harris, C. W. (Ed.). (1963). *Problems in measuring change.* Madison, WI: University of Wisconsin Press.

Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). New York: Academic Press.

Harris, R. J. (2005) Classical statistical inference: Practice versus presentation. In B. E. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 268–278). London: Wiley.

Harter, H. L. (1960). Tables of range and Studentized range. *Annals of Mathematical Statistics, 31*, 1122–1147.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.

Hays, W. L. (1994). *Statistics* (5th ed.). New York: Harcourt, Brace.

Hayes, A. F., & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *British Journal of Mathematical and Statistical Psychology, 60,* 217–244.

Hector, A., von Felten, S., & Schmid, B. (2010). Analysis of variance with unbalanced data: An update for ecology and evolution. *Journal of Animal Ecology, 9,* 308–316.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92,* 490–499.

Hedges, L. V. (2009). Statistical considerations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985) *Statistical methods for meta-analysis* Orlando: Academic Press.

Henderson, D. A., & Denison, D. R. (1989). Stepwise regression in social and psychological research. *Psychological Reports, 64,* 251–257.

Hindley, C. B., Filliozat, A. M., Klackenberg, G., Nicolet-Meister, D., & Sand, E. A. (1966). Differences in age of walking for five European longitudinal samples. *Human Biology, 38,* 364–379.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures.* New York: Wiley.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician, 55*, 19–24.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65–70.

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research, 11,* 213.

Holway, A. H., & Boring, E. G. (1940). The moon illusion and the angle of regard. *American Journal of Psychology,* 53, 509–516.

Horowitz, J., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 74*, 401–415.

Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of event scale: A measure of subjective stress. *Psychosomatic Medicine, 41,* 209–218.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics, 2,* 360–378.

Hout, M., Duncan, O. D., & Sobel, M. E. (1987). Association and heterogeneity: Structural models of similarities and differences. In C. C. Clegg (Ed.), *Sociological methodology, 17,* 145ff.

Howard, B. L., & Kendall, P. C. (1996). Cognitive-behavioral family therapy for anxiety-disordered children: A multiple-baseline evaluation. *Cognitive Therapy and Research, 5*, 423–443.

Howell, D. C. (1987). *Statistical methods for psychology* (2nd ed.). Boston: PWS-KENT.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Pacific Grove, CA: Duxbury.

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Wadsworth Cengage Learning.

Howell, D. C. (2008a). *Fundamental statistics for the behavioral sciences* (6th ed.). Pacific Grove, CA: Duxbury.

Howell, D. C. (2008b). The treatment of missing data, Chapter 1. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 208–224). London: Sage.

Howell, D. C. (2008c). Best practices in the analysis of variance. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 341–357). Thousand Oaks, CA: Sage.

Howell, D. C. (2010). Scheffé test. In N. J. Salkind, D. M. Dougherty, & B. Frey (Eds.), *Encyclopedia of research design*. Thousand Oaks, CA: Sage.

Howell, D. C., & Gordon, L. R. (1976). Computing the exact probability of an R by C contingency table with fixed marginal totals. *Behavior Research Methods and Instrumentation, 8*, 317.

Howell, D. C., & Huessy, H. R. (1981). Hyperkinetic behavior followed from 7 to 21 years of age. In M. Gittelman (Ed.), *Intervention strategies with hyperactive children* (pp. 201–214). White Plains, NY: M. E. Sharpe.

Howell, D. C., & Huessy, H. R. (1985). A fifteen-year follow-up of a behavioral history of Attention Deficit Disorder (ADD). *Pediatrics, 76,* 185–190.

Howell, D. C., & McConaughy, S. H. (1982). Nonorthogonal analysis of variance: Putting the question before the answer. *Educational and Psychological Measurement, 42,* 9–24.

Hraba, J., & Grant, G. (1970). Black is beautiful: A reexamination of racial preference and identification. *Journal of Personality and Social Psychology, 16,* 398–402.

Huberty, C. J. (1989). Problems with stepwise methods—better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 1) (pp. 43–70). Greenwich, CT: JAI Press.

Huesmann, L. R., Moise-Titus, J., Podolski, C., & Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology, 39,* 201–221.

Huitema, B. E. (1980). *The analysis of covariance and alternatives.* New York: Wiley.

Huitema, B. E. (in press). *The analysis of covariance and alternatives* (2nd ed.). New York: Wiley.

Huitema, B. E. (2005). Analysis of covariance. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Sciences*. Chichester, England: Wiley.

Huitema, B. E., & McKean, J. W. (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin, 110,* 291–304.

Huitema, B. E., & McKean, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Bulletin, 3*, 104–116.

Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60*, 38–58.

Hunter, J. E. (1997). Needed: a ban on the significance test. *Psychological Science, 8*, 3–7.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact *F* distributions. *Journal of the American Statistical Association, 65,* 1582–1589.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split plot designs. *Journal of Educational Statistics, l ,* 69–82.

Huynh, H., & Mandeville, G. K. (1979). Validity conditions in repeated measures designs. *Psychological Bulletin, 86,* 964–973.

Introini-Collison, I., & McGaugh, J. L. (1986). Epinephrine modulates long-term retention of an aversively-motivated discrimination task. *Behavioral and Neural Biology, 45,* 358–365.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.

Jankowski, M. K., Leitenberg, H., Henning, K., & Coffey, P. (2002). Parental caring as a possible buffer against sexual revictimization in young adult survivors of child sexual abuse. *Journal of Traumatic Stress, 15*, 235–244.

Jones, L.V., & Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods, 5*, 411–414.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: a model comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology, 46*, 433–455.

Kapoor, P., Rajkumar, S. V., Dispenzieri, A., Gertz, M. A., Lacy, M. Q., Dingli, D., Mikhael, J. R., Roy, V., Kyle, R. A., Greipp, P. R., Kumar, S., & Mandrekar, S. (2011). Melphalan and prednisone versus melphalan, prednisone and thalidomide for elderly and/or transplant ineligible patients with multiple myeloma: a meta-analysis. *Leukemia 25*(*4*), 689–696.

Kapp, B., Frysinger, R., Gallagher, M., & Hazelton, J. (1979). Amygdala central nucleus lesions: Effects on heart rate conditioning in the rabbit. *Physiology and Behavior, 23,* 1109–1117.

Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science, 1,* 122–127.

Kaufman, L., & Rock, I. (1962). The moon illusion, I. *Science, 136*, 953–961.

Kaufman, L., & Kaufman, J. (2000). Explaining the Moon illusion. *Proceedings of the National Academy of Sciences, 97,* 500–504.

Keen, K. J. (2010). *Graphics for Statistics and Data Analysis with R*. New York: Chapman and Hall.

Kelley, K. (2008). Methods for the behavioral, educational, and social sciences: An R package. *Behavioral Research Methods, 39,* 979–984.

Kendall, M. G. (1948). *Rank correlation methods.* London: Griffin.

Kennedy, J. J. (1983). *Analyzing qualitative data: Introductory loglinear analysis for behavioral research.* New York: Praeger.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99,* 422–431.

Keppel, G. (1973). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall.

Keselman, H. J., Cribbie, R., & Holland, B. (1999). The pairwise multiple comparison multiplicity problem: An alternative approach to familywise and comparisonwise Type I error control. *Psychological Methods, 4*, 58–69.

Keselman, H. J., Games, P. A., & Rogan, J. C. (1979). Protecting the overall rate of Type I errors of pairwise comparisons with an omnibus test statistic. *Psychological Bulletin, 86*, 884–888.

Keselman, H. J., Holland, B., & Cribbie, R. A. (2005). Multiple comparison procedures. In Everitt, B. S. and Howell, D. C. *Encyclopedia of Statistics in Behavioral Sciences*. Chichester, England: Wiley.

Keselman, H. J., & Keselman, J. C. (1988). Repeated measures multiple comparison procedures: Effects of violating multisample sphericity in unbalanced designs. *Journal of Educational Statistics, 13*, 215–226.

Keselman, H. J., & Rogan, J. C. (1977). The Tukey multiple comparison test: 1953–1976. *Psychological Bulletin, 84*, 1050–1056.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity conditions of repeated measures *F* tests. *Psychological Bulletin, 87*, 479–481.

King, D. A. (1986). Associative control of tolerance to the sedative effects of a short-acting benzodiazepine. Unpublished doctoral dissertation, University of Vermont.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences.* Belmont, CA: Brooks/Cole.

Kleinbaum, D. G., & Klein, M. (2002). *Logistic Regression: A self-learning text.* New York: Springer.

Klemchuk, H. P., Bond, L. A., & Howell, D. C. (1990). Coherence and correlates of level 1 perspective taking in young children. *Merrill-Palmer Quarterly, 36*, 369–387.

Kline, R. B. (2004). *Beyond significance testing.* Washington, D. C.: American Psychological Association.

Koele, P. (1982). Calculating power in analysis of variance. *Psychological Bulletin, 92,* 513–516.

Kohr, R. L., & Games, P. A. (1974). Robustness of the analysis of variance, the Welch procedure, and a Box procedure to heterogeneous variances. *The Journal of Experimental Education, 43*, 61–69.

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Landwehr, J. M., & Watkins, A. E. (1987). *Exploring data.* Palo Alto, CA: Dale Seymour Publications.

Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology, 31*, 107–112.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121.

Lappe, J. M., Davies, K. M., Travers-Gustafson, D., & Heaney, R. P. (2006). Vitamin D status in a rural postmenopausal female population. *Journal of the American College of Nutrition, 25*, 395–402.

Larzelere, R. E., & Mulaik, S. A. (1977). Single-sample tests for many correlations. *Psychological Bulletin, 84*, 557–569.

Latané, B., & Dabbs, J. M., Jr. (1975). Sex, group size, and helping in three cities. *Sociometry, 38*, 180–194.

Lecoutre, B. (1991). A correction for the $\tilde{\varepsilon}$ approximate test in repeated measures designs with two or more independent groups. *Journal of Educational Statistics, 16*, 371–372.

Leerkes, E., & Crockenberg, S. (1999). The development of maternal self-efficacy and its impact on maternal behavior. Poster presentation at the Biennial Meetings of the Society for Research in Child Development, Albuquerque, NM, April.

Lehman, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association, 88,* 1242–1249.

Lenth, R. V. (2001) Some practical guidelines for effective sample size determination. *American Statistician, 55*, 187–193.

Levene, H. (1960). Robust tests for the equality of variance. In I. Olkin (Ed.), *Contributions to probability and statistics.* Palo Alto, CA: Stanford University Press.

Levine, R. (1990). The pace of life and coronary heart disease. *American Scientist, 78,* 450–459.

Lewis, C., & Keren, G. (1977). You can't have your cake and eat it too: Some considerations of the error term. *Psychological Bulletin, 84*,1150–1154.

Lewis, D., & Burke, C. J. (1949). The use and misuse of the chi square test. *Psychological Bulletin, 46*, 433–489.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York, Wiley.

Lock, J., Le Grange, D., Agras, S., Moye, A., Bryson, S., & Jo, B. (2010). Randomized clinical trial comparing family-based treatment with adolescent-focused individual therapy for adolescents with anorexia nervosa. *Archives of General Psychiatry, 67*, 1025 –1032.

Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist, 8*, 750–751.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304–305.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin, 72,* 336–337.

Lunneborg, C. E. (1994). *Modeling experimental and observational data.* Belmont, CA: Duxbury.

Lunneborg, C. E. (2000). *Data analysis by resampling: Concepts and applications*. Pacific Grove, CA: Duxbury.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

Macnaughton, D. B. (1998). Which sums of squares are best in unbalanced analysis of variance. Unpublished paper available at http://www.matstat.com/ss/.

Maimaris, C., Summer, C. L., Browning, C., & Palmer, C. R. (1994). Injury patterns in cyclists attending an accident and emergency department: A comparison of helmet wearers and non-wearers. *British Medical Journal, 308*, 1537–1540.

Maindonald, J., & Braun, J. (2007). *Data analysis and graphics using R: An example-based approach* (2nd ed.). Cambridge, England: Cambridge University Press.

Malcarne, V. L., Compas, B. E., Epping-Jordan, J. E., & Howell, D. C. (1995). Cognitive factors in adjustment to cancer: Attributions of self-blame and perceptions of control. *Journal of Behavioral Medicine, 18*, 401–417.

Manly, B. F. J. (1997). *Randomization, Bootstrap, and Monte Carlo Methods in Biology* (2nd ed.). London: Chapman & Hall.

Mann-Jones, J. M., Ettinger, R. H., Baisden, J., & Baisden, K. (Unpublished). Dextromethorphan modulation of context-dependent morphine tolerance. Retrieved December 12, 2007 at www.eou.edu/psych/re/morphine-tolerance.doc.

Manning, C. A., Hall, J. L., & Gold, P. E. (1990). Glucose effects on memory and other neuropsychological tests in elderly humans. *Psychological Science, 1,* 307–311.

Manolov, R., & Solanas, A. (2008). Comparing $N = 1$ effect size indices in presence of autocorrelation. *Behavior Modification, 32*, 860–875.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the American Statistical Association, 22*, 719–748.

Marascuilo, L. A., & Busk, P. L. (1987) Log-linear models: A way to study main effects and interactions for multidimensional contingency tables with categorical data. *Journal of Counseling Psychology, 34,* 443–455.

Marascuilo, L. A., & Serlin, R. C. (1979). Tests and contrasts for comparing change parameters for a multiple sample McNemar data model. *British Journal of Mathematical and Statistical Psychology, 32*, 105–112.

Marascuilo, L. A., & Serlin, R. C. (1990). *Statistical methods for the social and behavioral sciences.* New York: Freeman.

Maris, E. (1998). Covariance adjustment versus gain scores—revisited. *Psychological Methods, 3*, 309–327.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

Mauchly, J. W. (1940). Significance test for sphericity of a normal *n*-variate distribution. *Annals of Mathematical Statistics, 11,* 204–209.

Maxwell, A. E. (1961). *Analyzing Quantitative Data*. London, Methuen.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics, 5,* 269–287.

Maxwell, S. E., & Cramer, E. M. (1975). A note on analysis of covariance. *Psychological Bulletin, 82,* 187–190.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison approach* (2nd ed.). Belmont, CA: Wadsworth.

Maxwell, S. E., Delaney, H. D., & Mannheimer, J. M. (1985). ANOVA of residuals and ANCOVA: Correcting an illusion by using model comparisons and graphs. *Journal of Educational Statistics, 10,* 197–209.

Mazzucchelli, T., Kane, R. T., & Rees, C. S. (2010). Behavioral activation interventions for well-being: A meta-analysis. *Journal of Positive Psychology, 5, 105–121.*

McClelland, G. H. (1997). Optimal design in psychological research. *Psychological Methods, 2*, 3–19.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376–390.

McConaughy, S. H. (1980). Cognitive structures for reading comprehension: Judging the relative importance of ideas in short stories. Unpublished doctoral dissertation, University of Vermont.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Methods, 11*, 386–401.

McIntyre, S. H., Montgomery, D. B., Srinivasan, V., & Weitz, B. A. (1983). "Evaluating the statistical significance of models developed by stepwise regression," *Journal of Marketing Research, 20*, 1–11.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153–157.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: Wiley.

Meijs, N., Cillessen, A. H. N., Scholte, R. H. J., Segers, E., & Spijkerman, R. (2010). Social intelligence and academic achievement as predictors of adolescent popularity. *Journal of Youth and Adolescence, 39,* 9373 –9379.

Miller, R. G., Jr. (1981). *Simultaneous statistical inference* (2nd ed.). New York: McGraw-Hill.

Mireault, G. C., & Bond, L. A. (1992). Parent death in childhood, perceived vulnerability, and adult depression and anxiety. *American Journal of Orthopsychiatry, 62*, 517–524.

Mood, A. M. (1950). *Introduction to the theory of statistics.* New York: McGraw-Hill.

Mood, A. M., & Graybill, F. A. (1963). *Introduction to the theory of statistics* (2nd ed.). New York: McGraw-Hill.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Newbury Park, CA: Sage.

Moore, D. S., & McCabe, G. P. (1989). *Introduction to the practice of statistics*. New York: Freeman.

Murphy, K. R., & Myors, B. (1998). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Mahwah, NJ: Lawrence Erlbaum.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon.

Neter, J., & Wasserman, W. (1974). *Applied linear statistical models.* Homewood, IL: Richard D. Irwin.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophic Transactions of the Royal Society of London (Series A), 231,* 289–337.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241–301.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). *SPSS*: *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill.

Norton, D. W. (1953). Study reported in E. F. Lindquist, *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin.

Norusis, M. J. (1985). SPSS$^X$ *advanced statistics guide.* New York: McGraw-Hill.

Nurcombe, B., & Fitzhenry-Coor, I. (1979). Decision making in the mental health interview: I. An introduction to an education and research problem. Paper delivered at the Conference on Problem Solving in Medicine, Smuggler's Notch, Vermont.

Nurcombe, B., Howell, D. C., Rauh, V. A., Teti, D. M., Ruoff, P., & Brennan, J. (1984). An intervention program for mothers of low birthweight infants: Preliminary results. *Journal of the American Academy of Child Psychiatry, 23,* 319–325.

Oakes, M. (1990) *Statistical inference*. Chestnut Hill, MA: Epidemiology Resources, Inc.

O'Brien, R. G. (1976). Comment on "Some problems in the nonorthogonal analysis of variance." *Psychological Bulletin, 83,* 72–74.

O'Brien, R. G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin, 89,* 570–574.

O'Grady, K. E. (1982). Measures of explained variation: Cautions and limitations. *Psychological Bulletin, 92,* 766–777.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286.

O'Neil, R., & Wetherill, G. B. (1971). The present state of multiple comparison methods. *Journal of the Royal Statistical Society (Series B), 33,* 218–250.

Osbourne, J. W. (2008). *Best practices in quantitative methods*. Thousand Oaks, CA: Sage.

Overall, J. E. (1972). Computers in behavioral science: Multiple covariance analysis by the general least squares regression method. *Behavioral Science, 17,* 313–320.

Overall, J. E. (1980). Power of chi-square tests for $2 \times 2$ contingency tables with small expected frequencies. *Psychological Bulletin, 87,* 132–135.

Overall, J. E., & Klett, C. J. (1972). *Applied multivariate analysis.* New York: McGraw-Hill.

Overall, J. E., & Shivakumar, C. (1999). Testing differences in response trends across a normalized time domain. *Journal of Clinical Psychology, 55*, 857–867.

Overall, J. E., & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin, 72,* 311–322.

Overall, J. E., Spiegel, D. K., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin, 82,* 182–186.

Overall, J. E., & Tonidandel, S. (2007). Analysis of data from a controlled repeated measurements design with baseline-dependent dropouts. *Methodology, 3*, 58–66.

Owens, C. M., Farmer, J. L., Ferron, J. M., & Allsopp, D. H. (2010). The correspondence among alternative approaches to summarizing effects in single-case studies. Paper presented conference of the American Educational Research Association, Denver, CO.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen in random sampling. *Philosophical Magazine, 50,* 157–175.

Peterson, W. P. (2001). Topics for discussion from current newspapers and journals. *Journal of Statistics Education, 9* (pages not numbered).

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717–737.

Preacher, K. J., & Hayes, A. F. (2009). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediation models. *Behavior Research Methods, 40*, 871–891.

Preacher, K. J., & Leonardelli, G. J. (2001, March). Calculation for the Sobel test: An interactive calculation tool for mediation tests [Computer software]. Available from http://www.people.ku.edu/~preacher/sobel/sobel.htm.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112,* 160–164.

Pugh, M. D. (1983). Contributory fault and rape convictions: Loglinear models for blaming the victim. *Social Psychology Quarterly, 46*, 233–242.

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but non-overlapping correlations, *Psychological Methods*, *1*, 178–183.

Ramsey, F. L., & Schafer, D. W. (1997). *The statistical sleuth: A course in methods of data analysis.* Belmont, CA: Duxbury Press.

Rasmussen, J. L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin, 101*, 136–139.

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent control group designs. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Boston: Houghton Mifflin.

Reilly, T. P., Drudge, O. W., Rosen, J. C., Loew, D. E., & Fischer, M. (1985). Concurrent and predictive validity of the WISC-R, McCarthy Scales, Woodcock-Johnson, and academic achievement. *Psychology in the Schools, 22,* 380–382.

Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6,* 271–280.

Robson, D. S. (1959). A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics, 15,* 187–191.

Rogers, R. W., & Prentice-Dunn, S. (1981). Deindividuation and anger-mediated aggression: Unmasking regressive racism. *Journal of Personality and Social Psychology, 41*, 63–73.

Rosa, L., Rosa, E., Sarner, L., & Barrett, S. (1998) A close look at Therapeutic Touch. *Journal of the American Medical Association, 279*, 1005–1010.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychologocal Bulletin, 86*, 638–641.

Rosenthal, R. (1990). How are we doing in soft psychology? Comment in *American Psychologist, 45,* 775–777.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.

Rosenthal, R., & Rubin, D. B. (1984). Multiple contrasts and ordered Bonferroni procedures. *Journal of Educational Psychology, 76,* 1028–1034.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105,* 143–146.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection.* New York: Wiley.

Ruback, R. B., & Juieng, D. (1997). Territorial defense in parking lots: Retaliation against waiting drivers. *Journal of Applied Social Psychology, 27*, 821–834.

Ryan, B. F., Joiner, B. L., & Ryan, T. A. (1985). *Minitab handbook* (2nd ed.). Boston: Duxbury Press.

Ryan, T. A. (1959). Multiple comparisons in psychological research. *Psychological Bulletin, 56,* 26–47.

Ryan, T. A. (1960). Significance tests for multiple comparisons of proportions, variances, and other statistics. *Psychological Bulletin, 57,* 318–328.

Sackett, D. L., Deeks, J. J., & Altman, D. G. (1996). Down with odds ratios! *Evidence-Based Medicine*, *1*, 164–166.

Saint-Exupery, A. de. (1943). *The little prince.* Tr. by K. Woods. New York: Harcourt Brace Jovanovich.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114.

Sauerland, S., Lefering, R., Bayer-Sandow, T., & Neugebauer, F. A. M. (2003). Fingers, hands, or patients? The concept of independent observations. *The Journal of Hand Surgery: Journal of the British Society of Surgery of the Hand, 28,* 102–105.

Scheffé, H. A. (1953). A method for judging all possible contrasts in the analysis of variance. *Biometrika, 40,* 87–104.

Scheffé, H. A. (1959). *The analysis of variance.* New York: Wiley.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316.

Seligman, M. E. P., Nolen-Hoeksema, S., Thornton, N., & Thornton, K. M. (1990). Explanatory style as a mechanism of disappointing athletic performance. *Psychological Science, 1,* 143–146.

Sgro, J. A., & Weinstock, S. (1963). Effects of delay on subsequent running under immediate reinforcement. *Journal of Experimental Psychology, 66*, 260–263.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Sĭdák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62*, 623–633.

Siegel, S. (1975). Evidence from rats that morphine tolerance is a learned response. *Journal of Comparative and Physiological Psychology, 80*, 498–506.

Simon, J. L., & Bruce, P. *Resampling Stats*. Software available at http://www.resample.com.

Smith, H. F. (1957). Interpretation of adjusted treatment means and regression in analysis of covariance. *Biometrics, 13*, 282–308.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Spilich, G. J., June, L., & Renner, J. (1992). Cigarette smoking and cognitive performance. *British Journal of Addiction 87*, 1313–1326.

St. Lawrence, J. S., Brasfield, T. L., Shirley, A., Jefferson, K. W., Alleyne, E., & O'Bannon, R. E. III. (1995). Cognitive-behavioral intervention to reduce African American adolescents' risk for HIV infection. *Journal of Consulting and Clinical Psychology, 63*, 221–237.

Stead, L. F., Perera, R., Bullen, C., Mant, D., & Lancaster, T. (2008). Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews* 2008, Issue 1. Art. No.: CD000146. DOI: 10.1002/14651858.CD000146.pub3.

St. Leger, A. S., Cochrane, A. L., & Moore, F. (1978). The anomaly that wouldn't go away. *Lancet, ii,* 1153.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164–182.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652–654.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Stigler, S. M. (1999). *Statistics on the table: A history of statistical concepts and methods.* Harvard University Press: Cambridge, MA.

Stone, W. S., Rudd, R. J., Ragozzino, M. E., & Gold, P. E. (1992). Glucose attenuation of deficits in memory retrieval in altered light/dark cycles. *Psychobiology*, *20*, 47–50.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human Factors, 48*, 381–391.

Stuetzle, W. (2005). Cross-validation. In B. S. Everitt & D. C. Howell, *Encyclopedia of statistics in behavioral sciences*. Chichester, England: Wiley.

Sullivan, C. M., & Bybee, D. I. (1999). Reducing violence using community-based advocacy for women with abusive partners. *Journal of Consulting and Clinical Psychology, 67,* 43–53.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Harper & Row.

Teri, L., Logdon, R. G., Uomoto, J., & McCurry, S. M. (1997). Behavioral treatment of depression in dementia patients: A controlled clinical trial. *Journals of Gerontology, 52B*, 159–166.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*, 77–83.

Thompson, B. (2000). A suggested revision to the forthcoming 5th edition of the APA *Publication Manual.* Retrieved from http://people.cehd.tamu.edu/~bthompson/apaeffec.htm accessed on 6/1/2011.

Tiku, M. L. (1967). Tables of the power of the *F* test. *Journal of the American Statistical Association, 62*, 525–539.

Todman, J., & Dugard, P. (2007). *Approaching multivariate analysis: An introduction for psychology.* Psychology Press.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning: I. Orientation and the short cut. *Journal of Experimental Psychology, 36*, 13–24.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin, 99*, 90–99.

Toothaker, L. (1991). *Multiple comparisons for researchers.* Newbury Park, CA: Sage.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Tukey, J. W. (1949). One degree of freedom for nonadditivity. *Biometrics, 5*, 232–242.

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton University.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tversky, A., & Kahneman, C. (1980) Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology, Vol. 1*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Unah, I., & Boger, J. (2001). Race and the death penalty in North Carolina: An empirical analysis: 1993–1997. Available online at http://www.deathpenaltyinfo.org/article.php?did=246&scid=.

U.S. Department of Commerce. (1977). *Social indicators,* 1976. (Document #C3.2: S01/2/976). Washington, DC: U.S. Government Printing Office.

U.S. Department of Justice, Bureau of Justice Statistics. (1983). *Prisoners in 1982*, Bulletin NCJ-87933. Washington, DC: U.S. Government Printing Office.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18*, 325–346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavioral Research Methods, Instruments, and Computers*, *35*, 1–10.

Vaughan, G. M., & Corballis, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin, 72,* 204–223.

Venables, W. N. (2000). Exegeses on Linear Models. Paper presented to the S-PLUS User's Conference. Washington, DC, 8–9th October, 1998. Available at http://www.stats.ox.ac.uk/pub/MASS3/Exegeses.pdf.

Verdooren, L. R. (1963). Extended tables of critical values for Wilcoxon's test statistic. *Biometrika, 50,* 177–186.

Vermont Department of Health. (1982). *1981 annual report of vital statistics in Vermont.* Burlington, VT.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36*, 1–48. <URL: http://www.jstatsoft.org/v36/i03/>.

Visintainer, M. A., Volpicelli, J. R., & Seligman, M. E. P. (1982). Tumor rejection in rats after inescapable or escapable shock. *Science, 216*, 437–439.

Wagner, B. M., Compas, B. E., & Howell, D. C. (1988). Daily and major life events: A test of an integrative model of psychosocial stress. *American Journal of Community Psychology, 61*, 189–205.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213–217.

Wainer, H. (1978). On the sensitivity of regression and regressors. *Psychological Bulletin, 85,* 267–273.

Wainer, H. (1984). How to display data badly. *American Statistician, 38*, 137–147.

Walsh, T. B., Kaplan, A. S., Attia, E., Olmsted, M., Parides, M., Carter, J. C., Pike, K. M., Devlin, M. J., Woodside, B., Roberto, C. A., & Rockert, W. (2006). Fluoxetine after weight restoration in anorexia nervosa. *Journal of the American Medical Association, 295*, 2605–2612.

Watkins, A. E. (1995). The law of averages. *Chance, 8,* 28–32.

Weaver, K. A. (1999). The statistically marvelous medical growth chart: A tool for teaching variability. *Teaching of Psychology, 26,* 284–286.

Weinberg, C. R., & Gladen, B. C. (1986). The beta-geometric distribution applied to comparative fecundability studies. *Biometrics, 42*, 547–560.

Weisberg, H. I. (1979). Statistical adjustments and uncontrolled studies. *Psychological Bulletin, 86,* 1149–1164.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29,* 350–362.

Welch, B. L. (1947). The generalization of Student's problem when several difference population variances are involved. *Biometrika, 34,* 29–35.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38,* 330–336.

Welkowitz, J., Ewen, R. B., & Cohen, J. (2000). *Introductory statistics for the behavioral sciences* (5th ed.). New York: Harcourt/Academic Press.

Welsch, R. E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association, 72,* 566–575.

Werner, M., Stabenau, J. B., & Pollin, W. (1970). TAT methods for the differentiation of families of schizophrenics, delinquents, and normals. *Journal of Abnormal Psychology, 75*, 139–145.

Wickens, T. D. (1989). *Multiway contingency table analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Wilcox, R. R. (1986). Critical values for the correlated *t*-test when there are missing observations. *Communications in Statistics, Simulation and Computation, 15,* 709–714.

Wilcox, R. R. (1987a). New designs in analysis of variance. *Annual Review of Psychology, 38,* 29–60.

Wilcox, R. R. (1987b). *New statistical procedures for the social sciences.* Hillsdale, NJ: Erlbaum.

Wilcox, R. R. (1992). Why can methods for comparing means have relatively low power, and what can you do to correct the problem? *Current Directions in Psychological Science, 1,* 101–105.

Wilcox, R. R. (1993). Analyzing repeated measures or randomized block designs using trimmed means. *British Journal of Mathematical and Statistical Psychology, 46,* 63–76.

Wilcox, R. R. (1995). ANOVA: The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies. *British Journal of Mathematical and Statistical Psychology, 48, 99–114.*

Wilcox, R. R. (2005). Trimmed means. In Everitt, B. E. & Howell, D. C. *Encyclopedia of statistics in behavioral science.* Chichester, England: Wiley.

Wilcox, R. R. (2008). Robust methods for detecting and describing associations. In Osborne, J. R. *Best practices in quantitative methods*. Thousand Oaks, CA, Sage.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society (Series B), 21,* 396–399.

Wilkinson, L. (1994). Less is more: Two and three dimensional graphics for data display. *Behavior Research Methods, Instrumentation, and Computers, 26,* 172–176.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

Willer, R. (2005). "Overdoing Gender: A Test of the Masculine Overcompensation Thesis." Paper presented at the Annual Meeting of the American Sociological Association. Philadelphia, PA.

Williamson, J. A. (2008). Correlates of Coping Styles in Children of Depressed Parents: Observations of Positive and Negative Emotions in Parent-Child Interactions. Honors Thesis, Vanderbilt University.

Winer, B. J. (1962). *Statistical principles in experimental design.* New York: McGraw-Hill.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical Principles in Experimental Design* (3rd ed.). New York: McGraw-Hill.

Yates, F. (1934). Contingency tables involving small numbers and the $\chi^2$ test. Supplement. *Journal of the Royal Statistical Society (Series B),1,* 217–235.

Yuen, K. K., & Dixon, W. J. (1973). The approximate behavior and performance of the two-sample trimmed *t. Biometrika, 60,* 369–374.

Ziegler, S. G. (1994). The effects of attentional shift training on the execution of soccer skills: A preliminary investigation. *Journal of Applied Behavior Analysis, 27*, 545–552.

Zuckerman, M., Hodgins, H. S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data. *Psychological Science, 4,* 49–53.

Zumbo, B. D., & Zimmerman, D. W. (2000). Scales of measurement and the relation between parametric and nonparametric statistical tests. In B. Thompson (Ed.), *Advances in social science methodology, Vol. 6*. Greenwich, CT: JAI Press.



# Alıştırmalara Cevaplar

Tek-numaralı alıştırmaların çoğunun cevaplarını veriyorum. (Çift numaralı alıştırmaların cevapları verilmedi, çünkü öğretim elemanlarının çoğu bazı cevapların öğrenciye bırakılmasını istediler, ilerideki bölümlerde yer alan karmaşık cevaplar bu klitabın Web sitesinden (www.uvm.edu/~dhowell/methods8/) bulunabilir. Bazı tek numaralı cevaplar (soru grafik veya şekil çizmeyi veya bilgisayar cevabıyla el ile hesaplama arasında karşılaştırma gerektirdiğinden) atlanmıştır. Verilen bir veri kümesine bilgisayar yazılımı uygulanmasını gerektiren cevaplar da atlanmıştır. Ya sorunun cevabını verebilecek -doğru cevabı kesin olarak bulabileceksiniz- ya da başlangıçta problemi kuramayacaksınız (bu durumda sayısal cevabın hiçbir yardımı olmayacak). Bazen hayal kırıklığına uğrayacaksınız, çünkü Tartışma Sorularına sıklıkla cevap vermedim, bu soruların cevapları çoğunlukla basit değildir. Diğer durumlarda cevaplar kolaydır, fakat problem üzerinde bir süre düşünmenizi ve nereye varacağınızı görmenizi istiyorum. Çoğu zaman benim verebileceğim basit cevaptan daha enteresan cevaplar bulacaksınız. Alıştırmanın nereye doğru gittiğini görememenin sinir bozucu olduğunu hatırlarım; böyle durumlara ben de kızarım. Fakat tartışma sorularının doğası budur.

## BÖLÜM 1

1.1 Kolej veya üniversitenizin tüm öğrencilerinin kümesi herhangi bir koşul altında evren olarak düşünülebilir, burada sadece okulunuzun öğrenci evrenine genelleme yapabilirsiniz.

1.3 Kolejinizin veya üniversitenizin öğrencileri ABD öğrencilerinin tesadüfî olmayan bir örneklemidir, çünkü bütün örnekleme girmek için ABD öğrencilerinin hepsine eşit şans verilmemiştir.

1.5 Bağımsız değişkenler: anaokuluna gitmeyenlere karşı anaokulu birinci sınıf öğrencileri; üst sınıf öğrencileri, maraton koşucuları kategorileri olarak yüksek lisans öğrencileri, yüksek lisansa hazırlık öğrencileri ve gençler. Bağımlı değişkenler: birinci-sınıf öğretmenleri tarafından verilen sosyal-uyum puanları; 26,2 mil koşmak için geçen süre.

1.7 Sürekli değişkenler: hamilelik süresi; kelime/dakika olarak yazma süresi; kütüphanedeki kitapların sayısı.

1.9 Maraton yarışmasının planlayıcıları, yarışmayı buna göre planlayabilmek için, büyükler (seniors), ustalar (masters), ustaların altındakiler (submasters), ve genç yarışmacılar (juniors) için ortalama koşma sürelerini bilmek istiyor olabilirler.

1.11 (a) 1984 yılı Ekim ayında Brown Üniversitesi öğrencileri arasında yapılan oylamada, nükleer hastalıklara karşı intihar eylemleri için kullanılan ilaçları stoklama ve bu stoklamaya karşı olanların sayıları.

(b) Küçük bir Midwestern Üniversitesindeki beyaz, Afrika kökenli Amerikalı, İspanya kökenli Amerikalı, Asyalı ve diğer öğrencilerin sayıları.

(c) Alkolizmi tedavi etmek için uygulanan programdan bir yıl sonra, "hâlâ tedavide olanlar", "tedavi aramadan içmeye deam edenler" veya "tekrar tedavi altına alınanlar"

1.13 Bir ilkokuldaki öğrencilerin puanları sayısal olarak (ölçme ölçeği) rapor edilebilir, öğrenciler Bluebirds (Rating > 90), Robins (Rating 70-90) veya Kardinaller (Rating < 70) olarak sınıflandırılabilir.

1.15 Boy uzunlukları ve cinsiyetleri verilen yetişkinler için ağırlık, beden ağırlığının oran ölçeğidir; fakat güç için iyi bir sıralama ölçeğidir.

1.17 Muhtemelen, hız, motivasyon için öğrenmeye göre daha iyi bir indekstir.

1.19 (a) Genel başarı ortalaması düşük olan öğrencilerden bu durumla ilgili olanların ortalamaları ile olmayanların ortalamaları karşılaştırılabilir. (b) Cinsel birleşmeden mutlu olan eşlerle olmayan eşlerin frekansları karşılaştırılabilir.

1.21 Sigaranın sağlığa etkileri konusunda Çin'de yapılan ilginç bir çalışma http://berkeley.edu/news/media/releases/2005/09/04_smoking.shtml adresinden bulunabilir.

## BÖLÜM 2

2.1 (b) Tek modlu ve pozitif kayışlı.

2.3 Alıştırma 2.1'deki veri için gövde-ve-yaprak grafiğinin yapılmasındaki problem, onluk sayı sisteminin kullanılması durumunda değerlerin çoğunun iki yaprak arasına düşmesidir. Ve gövdeleri iki katına çıkarsak bile daha iyi olmayacaktır.

2.5 Belirli bir cevap yoktur.

2.7 Belirli bir cevap yoktur.

2.9 Birinci çeyrek kadınlar için 80 olduğu halde, erkekler için 77'dir. Üçüncü çeyrek erkek ve kadınlar için hemen hemen eşittir ve değeri 87'dir.

2.11 Karşılaşacağım sonraki 200 kişinin her ay gideceği flim sayısının dağılımının şekli tepe noktası "0"da olan pozitif çarpık ve başlangıca doğru keskin bir düşüşle ayda 5 film olacaktır.

2.13 Bu bir gövde-ve-yaprak grafiğidir.

2.15 (a) $Y_1 = 9$, $Y_{10} = 2$ (b) $Y = 9 + 9 + \ldots + 2 = 57$
$(\Sigma Y)^2 = (9 + 9 + \cdots + 2)^2 = 3249$

2.17 (a) $\Sigma Y^2 = 9^2 + 9^2 \cdots 2^2 = 377$

(b) $$\frac{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}{N-1} = \frac{377 - \frac{3249}{10}}{9} = 5{,}789$$

(c) $\sqrt{5{,}789} = 2{,}406$

(d) (b)'deki müzikalite puanlarının karelerinin ölçümünün birimi ve (c)'deki müzikalite puanlarının birimi.

2.19 (a) $\Sigma(X + Y) = (10 + 9) + (8 + 9) + \cdots + (7 + 2) = 134$
$\Sigma X + \Sigma Y = 77 + 57 = 134$

(b) $\Sigma XY = 10(9) + 3(8) + \cdots + 3(7) = 460$
$\Sigma X \Sigma Y = (77)(57) = 4389$

(c) $\Sigma CX = \Sigma 3X = 3(10) + 3(8) + \cdots + 3(7) = 231$
$C\Sigma X = 3(77) = 231$

(d) $\Sigma X^2 = 10^2 + 8^2 + \cdot\cdot 7^2 = 657$
$(\Sigma X)^2 = 77^2 = 5929$

2.21 Alıştırma 2.20'deki sonuç ardışık süreci destekler.

2.23 Veri kümesi gözlemlerin bağımsız olması durumunu desteklemiyor, çünkü katılımcı muhtemelen denemelerin başlarında öğreniyor ve daha sonra görev geliştikçe yoruluyor. Böylece cevaplar zaman ilerledikçe birbirinden uzaklaşmak yerine yakınlaşıyor ve birbirine benziyor.

2.25 Beyaz deneklere verilen şok miktarı araştırmacı tarafından hakaret edilip edilmemesine bağlı değildir. Diğer taraftan, siyah katılımcılar araştırmacının deneğe hakaret etmesinden acı çekmektedirler.

2.27 Soru grafik çizilmesini istemektedir.

2.29 Erkek ve kadınların arasındaki fark değişmeden evlenme yaşı güçlü bir şekilde yükseliyor. Sanki, çiftler evlilik öncesinde evlenmeden birlikte yaşıyor görünüyorlar.

2.31 Ortalama ortancadan daha büyük bir değer olur.

2.33 ortalama = 21,33; ortanca = 21

2.35 Eğer 3 ile çarpılırsa ortalama 4,783'ten 14,5'e yükselir, ortanca ve tepe değer 5 iken 15 olur.

2.37 Bunlar hoş bir şekilde, bana, önceki alıştırmalarla olan beklentilerimin uyum içinde olduğunu gösterdi.

2.39 Bu bir bilgisayar sorusu.

2.41 Açıklık (Ranj) = 16; varyans = 11,592; standart sapma = 3,405.

2.43 Aralık = 9,908 ile 27,892 arasında.

2.45 Orijinal veriyi sadece orijinal standart sapma olan 2,1'e bölünüz.

2.47 Bu bir grafik gösterimi.

2.49 CV (değişim katsayısı) = 0,351

2.51 %10 kırpılmış ortalama = 32,675.

2.53 Bu bir grafik gösterimidir.

2.55 Bu bir Internet araştırmasıdır.

## BÖLÜM 3

3.1 (b) -3, -2, -2, -1, -1, -1, 0, 0, 0, 0, 1, 1 ,1, 2, 2, 3

(c) -1,84, -1,23, -1,23, -0,61, -0,61, 0,61, 0, 0, 0, 0, 0,61, 0,61 0,61, 1,23, 1,23 1,84

3.3 (a) %68 (b) %50 (c) %84

3.5 $z$ = (950 - 975)/15 = -1,67; verilen dağılımdan bildiğimize göre, zamanın sadece %4,75'i kadar bir sürede 950'ye kadar düşük bir değer bekleyebiliriz. İki kuyruklu olasılık ,095 olacaktır.

3.7 İki dağılımın standart sapmaları eşit olduğunda Alıştırma 3.6 (b) ve (c)'deki soruların cevapları aynı olacaktır.

3.9 (a) $2512.68 (b) $1342.00

3.11 Standart sapmayı 10'a yükseltmek için ham puanları 7/10'a bölünüz ve daha sonra ortalamayı 80'e getirmek için her yeni puana 11,43 ekleyiniz.

3.13 $z$ = (600 - 489)/126 = 0,88. Bu nedenle puanların %88'i bunun altına düşer. Böylece 600, 81'inci yüzdeliği temsil eder.

3.15 $z$ = 0,79, $p$ = ,7852; $X$ = 586,591

İleri sınıf öğrencileri ve kayıtsız olan kolej mezunları için, 600 GBO puanı 79'uncu yüzdelik ve 587 puanı 75'inci yüzdeliğe düşer.

3.17 GBO için 75'inci yüzdelik 3,04'tür.

3.19 Ortalamanın altındaki puanlarda yapılabilecek anlamlı bir ayrım yoktur ve böylece bu aralıktaki bütün öğrencilerin $T$ puanları 50'dir.

3.21 Müdahele-sonrası ağırlıklar makul şekilde normaldir, fakat kazanılan puanlar ve kazanılan yüzde puanlar normalden uzaktır. Ancak örneklemimiz küçüktür.

3.23 İlk olarak $\mu = 0$ ve $\sigma = 1$ olan normal dağılıma sahip bir evrenden 16 puan çekerim. Bu değişkene $z1$ diyelim. Bu örneklemin ($z1$), hemen hemen kesinlikle, ortalaması "0" ve standart sapması "1" değildir. Sonra yeni bir değişken oluştururum, $z2 = z1$ - ortalama $z1$. Daha sonra ortalaması "0" ve standart sapması "1" olan yeni bir değişken ($z3$) oluşturmak için $z2$'yi std ($z1$)'e bölerim. Ardından bunu 4,25 ile çarparak Std'si 4,25 olan bir dağılım elde ederim, en sonda 16,2 (yeni ortalama) ile toplarım. Şimdi, ortalama tam 16,3 ve standart sapma 4,25'tir.

3.25 Birleşik veri kümesi çift modludur, fakat böyle küçük bir veri kümesi ile her ne kadar çok kesin olmak zor olsa da, düzeltilmiş veri kümesinin dağılımı normale daha yakındır.

## BÖLÜM 4

4.1 (a) Sıfır hipotezini, dün akşamki maç gerçekten Ulusal Hokey Ligi (NHL) maçı idi ifadesiyle kurarım.

(b) Bu hipotezi temel alarak her takımın 0 ile 6 arasında puan kazanmasını beklerim. Sonra kazanılan gerçek puanlara bakarım ve şu karara varırım: Eğer bu bir NHL maçı ise, onlar benim beklentimin dışında bir yoldalar. Böylece, sıfır hipotezini reddederim.

4.3 Öyle olmadığı halde paranın üstünü eksik aldığımı düşünmüştüm.

4.5 Kritik değer aşağıdaki değişimin bu miktarın olacaktır ki ben onu göz ardı etmeye karar verdim. Ret bölgesi, değişimin bütün miktarları kritik değerden küçük olacaktır - bu, bütün miktarların $H_0$'ın ret bölgesini oluşturacağı anlamına gelir.

4.7 z = (490 - 650)/50 = -3,2. Uygun olarak seçilenler arasından tesadüfî olarak çekilen bir öğrencinin GBO'nun 490 olması olasılığı ,0007'dir. Bu öğrencinin annesinin mütevelli heyet üyeleri arasında olmasının rolü olabileceğini zannederim.

4.9 Her zaman olduğu gibi, dağılımın aynı nedenden dolayı düzgün bir şekilde sağa düşeceği anlamına gelir. -az sayıda yüksek-puanlı öğrenci vardır.

4.11 $M$ test istatistiği olarak adlandırılır.

4.13 Alternatif hipotez, bu öğrenci ortalaması 650 olmayan öğrenciler arasından çekilmiştir.

4.15 "Dağılım" kelimesi herhangi bir gözlemler kümesinden elde edilmiştir anlamına gelir. "Örneklem dağılımı" örneklem istatistiklerinin elde edildiği (teorik veya deneysel) dağılımı belirtir.

4.17 *(a) Araştırma hipotezi* - Anaokuluna giden öğrenciler, gitmeyenlere göre birinci sınıfa daha hızlı uyum sağlarlar. *Sıfır hipotez* - Anaokuluna giden ve gitmeyen öğrenciler için birinci sınıfa uyum oranları eşittir.

(b) *Araştırma hipotezi* - Ortaokulda seks eğitimi verilmesi lisedeki evlenmemiş anneler arasında hamilelik oranını düşürür. *Sıfır hipotez* -Lisedeki evlenmemiş anneler arasında hamilelik oranları ortaokulda seks eğitimi almış olma veya olmama yönünden aynıdır.

4.19 $a = 0{,}1$ için $z = -2{,}327$ olmalıdır. Kesme puanı yaklaşık olarak 53,46'dır. $z$ için 53,46 kesme puanı $H_1$ eğrisine uygulandığında $z = -1{,}33$'tür. Ek $Z$'den $\beta = {,}9082$ buluruz.

4.21 Puanlar ile ders değerlendirmeleri arasında gerçek bir ilişkinin olup olmadığını belirlemek için, bu iki değişken arasındaki ilişkinin derecesini bulurum. Böyle bir istatistiği ($r$) Bölüm 9'da göreceksiniz. Bu istatistiği örneklem dağılımından hesaplarım, böyle bir durumda bu iki değişken arasında bir ilişki olmadığı görülür. Son olarak temsilci öğrenciler kümesi ve sınıflar arasında bu istatistiği hesaplarım ve örneklem değerimle örneklem dağılımına ait değerini karşılaştırırım.

4.23 (a) Sınıftan geniş bir erkek öğrenciler örneklemi ve geniş bir kız öğrenciler örneklemi çekersiniz ve bu iki grubun yemek paraları ortalamalarını hesaplarsınız. Sıfır hipotezi, erkek ve kız öğrencilerin harçlıkları ortalamalarının evrende eşit olduğudur.

(b) İki kuyruklu $t$ testi kullanırdım, çünkü kız öğrencilerin erkek öğrencilerden manidar olarak daha çok veya manidar olarak daha az harçlık aldıkları sıfır hipotezini reddetmek istiyorum.

(c) Eğer iki ortalama arasındaki fark şansa bağlı olarak bekleyebileceğimden büyükse sıfır hipotezini reddederim. Aksi halde sıfır hipotezini reddetmem.

(d) Yapılacak en önemli şey, bazı çocuklar tarafından rapor edilen doğruların dışındaki harcama miktarları bulmaktır.

4.25 Park yeri örneğinde hipotez testine geleneksel yaklaşım, bir yerin boşalması ortalamasının, bir kimsenin bekleyip beklemediği ortalaması sıfır hipotezidir. Eğer sıfır hipotezi reddedilemezse, $a = {,}05$ manidarlık düzeyinde iki yönlü test edeceğim. Diğer taraftan Jones ve Tukey muhtemelen evren ortalamalarının eşit olduğu sıfır hipotezinin doğru olmadığını düşünmektedirler.

Onlar hangi evrenin ortalamasının büyük olduğunun karar verilmesine odaklanmaktadırlar. "Manidar olmayan bir sonuç"un sadece herhangi bir karar çıkarmaya yeterli veri olmadığı anlamına geldiği sonucuna varılabileceği düşüncesindedirler. Jones ve Tukey aynı zamanda $a = {,}025$ düzeyinde tek yönlü test ile çalışılmasından yanadırlar, fakat iki yönlü test yapılması gerekir çünkü farkın yönü hakkında varsayımda bulunulmamıştır.

4.27 Yardım arayan kadınlar oranı

(a) Eğer p = 0,50 ise kadın oranının %61 olması olası değildir. Benim örneklem dağılımımda puanın 61 veya daha byük olması 16/1000 = %1,6 defa elde edildi.

(b) Aynı yöntemi binom dağılımında $p = {,}75$ için hariç olmak üzere tekrar uygulardım.

## BÖLÜM 5

5.1 (a) Analitik: Eğer iki tenis oyuncusu tamamen eşit beceriye sahipse -böylece maçlarının sonuçları tesaadüfîdir, A oyuncusu sonraki maçı ,50 olasılıkla kazanacaktır.

(b) Bağıl frekans: Eğer geçmiş maçlarda oynadıkları maçtan 13'ünde A oyuncusu B oyuncusunu yenmiş ise, A oyuncusu 13/17 = ,76 olasılıkla gelecekteki maçları kazanacaktır.

(c) Öznel/subjektif: A oyuncusunun antrenörü B'ye karşı, gelecekteki maçları ,90 olasılıkla kazanacağını hissetmektedir.

5.3 (a) Birinci ödülü kazanmadı ise ikinci ödülü kazanma olasılığı $p = 1/9 = {,}111$'dir.

(b) Birinci ödülü onun, ikinciyi sizin kazanma olasılığı $p = (2/10)(1/9) = (,20)(,111) = {,}022$'dir.

(c) Birinci ve ikinciyi sizin kazanmanız olasılığı $p = (1/10)(2/9) = 0{,}022$'dir.

(d) $p$ = (siz önce, o ikinci) [=,022)]) + (o önce, siz ikinci) [=,022)]) = $p$ (siz ve o birinci ve ikinci) = ,044.

5.5 Şartlı olasılıklar Alıştırma 5.3(a)'da dikkate alındı.

5.7 Şartlı olasılıklar: Kayakçılık şartlarının bugün iyi oldukları bilindiğine göre çarşamba günü de iyi olacağı olasılığı kaçtır?

5.9 $p = (2/13)(3/13) = (,154)(,231) = {,}036$.

5.11 Gözlemlerin belirlenen bir aralığa düştüğü ilgilendiğimiz sürekli dağılım, bebeğinizin tasarlanan gününde doğacağı güne göre örneklendirilmiştir.

5.13 Kesikli değişkene iki örnek: Bu akşam yemekte servisi yapılacak et yemekleri; sahip olduğu masa üstü bilgisayarın markası.

5.15 (a) Başvuranların %20 veya %60, 80'inci yüzdelikte veya üstünde ve bunların 10'u seçilecektir. Böylece, $p$ (yüksek puanla başvuran bir başvuranın kabul edilmesi) = 10/60 = 0,167.

(b) 80'nci yüzdeliğin altında hiç kimse kabul edilmedi. Böylece $p$(düşük puanla kabul edilme) = ,00.

5.17 (a) $z$ = -,33; $p$(büyük kısım) = ,6293.

(b) 29/88 = %53 > 50; 32/55 = %58 ≥ 50.

5.19 Okulu terk etme olasılığını, ADDSC puanını yok sayarak, ilkokuldaki ADDSC puanı belli bir değerin (örn: 66) üstünde olan öğrencilerin, okulu terk etme şartlı olasılığı ile karşılaştırırız.

5.21 5-seçenekli bir test maddesinde doğru seçimlerin olasılığı:

| | |
|---|---|
| $p(0)$ = ,1074 | $p(6)$ = ,0055 |
| $p(1)$ = ,2684 | $p(7)$ = ,0008 |
| $p(2)$ = ,3020 | $p(8)$ = ,0001 |
| $p(3)$ = ,2013 | $p(9)$ = ,0000 |
| $p(4)$ = ,0881 | $p(10)$ = ,0000 |
| $p(5)$ = ,0264 | |

5.23 4 doğru seçime kadar başarı $a$ = ,05 düzeyinde şans başarısını gösterir, fakat 5 veya daha fazla doğru seçim artık şans düzeyinde başarıyı göstermez.

5.25 Eğer konut ayrımcılığı yoksa, bireyin ırkı ile kendisine verilmeyen evin belirli bir birimi bağımsız olaylardır. Belirli bir birimin (veya şehrin belirli bölgesindeki birimin) belirli bir gelir grubundaki kimseye verilmesi olasılığını, bağımsız olduğunu varsayarak, heaplayabiliriz ve bu cevabı böyle bir birimin etnik azınlık grubuna verilmesiyle ilgili gerçek oranla karşılaştırabiliriz.

5.27 Alıştırma 5.26'daki sözel öğrenme deneyinde gerekli deneklerin sayısı, eğer her denek dört sınıftaki kelimelerden sadece 2 sınıftakini görebilirse 4 şeyden her defasında 2'sinin alınması permütasyonlarının sayısı = 4!/2! = 12'dir.

5.29 Toplam dondurma külâhları yapma sayısı = 63. (Dondurma olmadan dondurma külâhı yapamazsınız, böylece 6 şeyden her defasında "0" tanesinin alınmasıyla ilgili kombinasyonu çıkarmanız gerekir.)

5.31 Çünkü şansla 11 doğru cevap olasılığı ,16; 11 veya *daha fazla* doğru cevap olasılığı ,16'dan büyük olmalı. Böylece, $p$ = ,50 hipotezini $a$ = ,05 düzeyinde reddedemeyiz (öğrenci tahmin ediyor).

5.33 Sürücü testinde beklenen geçme oranı %60 iken, 30 sürücüden 22'si geçti:

$$z = \frac{22 - 30(,60)}{\sqrt{30(,60)(,40)}} = 1,49$$

$H_0$'ı, $a$ = ,05 düzeyinde reddedemeyiz.

5.35 Öğrenciler doğanın, olanların ortaya çıkmasında sorumluluğunun olmadığının ve geçmişte olanların hatırlanmasında korkunç bir hafızaya sahip olduğunu anlamalıdır. Herhangi bir "ortalamalar kanunu", uzun dönemdeki bir olaylar serisinin sonuçlarını belirtir ve ne görmeyi beklediğimizi açıklar. Herhangi bir kendi kendine düzeltme mekanizması yoktur.

5.37 Düşüktür çünkü göğüs kanserinin kendisine ait olasılık düşüktür. Fakat cesaretiniz fazla kırılmasın. Toplanan veriye (bir pozitif mamoğrafi) göre gözlendiğinden olasılık 7,8 kat daha fazladır. (Ve eğer bir kadınsanız, lütfen mamoğrafi çektirmeye son vermeyiniz).

## BÖLÜM 6

6.1 2 *sd*'de $\chi^2$ = 11,33; $H_0$ reddedilir ve öğrencilerin tesadüfî olarak kayıt yaptırmadıkları kararına varılır.

6.3 1 *sd*'de $\chi^2$ = 29,35; $H_0$ reddedilir ve çocukların bebekleri tesadüfî olarak seçmedikleri kararına varılır (en azından renklerine göre).

6.5 1 sd'de $\chi^2$ = 34,184; $H_0$ reddedilir ve bu iki çalışmada, siyah ve beyaz bebeklerin seçilmesi dağılımlarının farklı olduğu kararına varılır. Seçim çalışmadan bağımsız *değildir*. Bundan böyle bir bebek renginin diğerine tercih edilip edilmediğini soramayız. Fakat çalışmalar arasında model tercihinin sabit olup olmadığını sorabiliriz.

6.7 (a) Tesadüfî olarak bir grup denek seçiniz, bunları cinsiyete ve hayat tarzlarına (üç düzeyde sınıflayarak) göre tasnif ediniz.

(b) Kasıtlı olarak eşit sayıda erkek ve kadın alınız ve onlara üç hayat tarzından hangisini tercih ettiklerini sorunuz.

(c) Kasıtlı olarak 10 erkek ve 10 kadın alınız ve kendilerinin her birinde 10'ar oyuncu olacak şekilde iki takıma ayırmalarını isteyiniz.

6.9 (a) $\chi^2$ = 10,306

(b) Bu gösteri $\chi^2$'nin kritik değeri sabit kalmışken elde edilen değerinin tam olarak ikiye katlanmış olduğunu gösterir. Böylece örneklem büyüklüğü çok önemli rol oynar, büyük olasılıkla geniş örneklemlerde manidar sonuçlar elde edilir - bu durum diğer testler için de geçerlidir.

6.11 $\chi^2$ = ,50. $H_0$ reddedilir ve oylamada kadınlar erkeklerden farklı seçim yapmışlardır. Kadınların sivil toplum kuruluşlarına oy vermeleri büyük olasılıkla daha muhtemeldir -olasılık ratio (35/9)(60/41) = 3,89/1,46 = 2,66'dır. Bu sivil toplum kuruluşlarına oy verme oranının kadınlarda erkeklere göre 2,66 kat olduğu anlamına gelir. Bu dikkate değer bir farktır ve tutumlarda temel farktır.

6.13 (a) 2 *sd*'de $\chi^2$ = 37,141; $H_0$ reddedilir ve yetişkin kızların kilo verme tercihleri ırka göre değişmektedir.

(b) Kilo vermek isteyen kızların sayısı gerçekten fazla kilolu sayısından fazla oldu.

6.15 Olabilirlik oranı 7 *sd*'de $\chi^2 = 12{,}753$; $H_0$ reddedilmez.

6.17 (a) $\chi^2 = 9{,}0$

(b) Eğer Pazartesi Gecesi Futbolu insanların düşüncelerini gerçekten değiştiriyorsa (olumsuza doğru), daha sonra düşüncelerini değiştiren bu insanlar, ne oldu da daha çok olumludan olumsuza veya tam tersine doğru değiştirdiler.

(c) Analiz değişmeyenleri dikkate almaz. Sadece değişenleri dikkate alır.

6.19 (b) Satır yüzdeleri satır toplamlarını, sütun yüzdeleri sütun toplamlarını göz önüne alır.

(c) Bunlar $H_0$ altında $\chi^2 \geq \chi^2_{gözlenen}$ olasılıklardır (dört ondalıklı).

(d) Değişkenler arasındaki korelasyon yaklaşık olarak ,25'tir.

6.21 (a) Cramér $\phi_c = \sqrt{26{,}903/22{,}071} = ,0349$.

(b) Ölümcül | etkisiz (placebo) = 18/10,845 = ,00166 olasılığındadır. Ölümcül | aspirin olasılığı ise 5/10,933 = ,000453'tür. Olasılık oranı = ,00166/,000453 = 3,66 bulunur. Aspirin alınmazsa kalp krizinden ölüm, alınması durumuna göre 3,66 katıdır.

6.23 Tablo 6.4 için, ırkın fonksiyonu olarak ölüm cezası olasılık oranı (33,251/33/508) = 2,017'dir. Eğer bir kimse beyaz değilse, ölüm cezasına çarptırılması olasılığı beyazların iki katıdır.

6.25 Dabbs ve Morris'in (1990) çalışmasına göre 1 *sd*'de $\chi^2 = 64{,}08$'dir. $H_0$'ı reddedebiliriz ve antisosyal davranışın testesteron düzeyi ile ilişkili olduğu kararına varabiliriz.

6.27 (a) 1 *sd*'de $\chi^2 = 15{,}57$. $H_0$ ret.

(b) Yetişkin erkeklerde testeron düzeyi ve çocuklukta işledikleri suçlar arasında manidar bir ilişki vardır.

(c) Bu sonuç, bu iki değişkeni (suçluluk ve testeron) tarihsel olarak birbirine bağlayabileceğimizi gösterir.

6.29 (a) $\chi^2 = 9{,}79$. $H_0$ ret.

(b) Olasılık oranı = (43/457)/(50/268) = ,094/,186 = ,505. Programı alanların yaklaşık yarısı daha sonra tacize uğramışlardır.

6.31 (a) $\chi^2 = ,232$, $p = ,630$.

(b) Ailenin cinsiyeti ile bebeğin cinsiyeti arasında bir ilişki yoktur.

(c) Çocuğun cinsiyetinden gelen etkileri ailenin cinsiyetinden gelen etkilerden ayıramayacaktık. Tamamen karıştırılmış olacaktı.

6.33 Benzer bir dizi soru sorabiliriz ve dengeli bir şekilde "doğru" ve "yanlış" cevapları ayırabiliriz. Daha sonra cevapları pozitif ve negatif kategorilere ayırabiliriz ve fakültenin büyük olasılıkla öğrencilerden daha olumsuz cevaplar verip vermediğini sorabiliriz.

6.35 Eğer erkekler ve kadınlar için ölçek puanları ortalamaları farklı şeyler ifade ediyorsa ölçeğin kapalı-uçlu olması gereği olarak ilişki bozuk olabilir.

6.37 Fidalgo'nun işgücündeki zorbalık çalışması.

(a) İş kategorilerindeki daralma $\chi^2 = 4{,}20$, $p = ,05$

(b) **OR** = ,70.

(c) ve

(d) Manthel-Haentszel istatistiği 2,285'tir. İş kategorileri arasındaki şartları düzenlersek cinsiyet ve zorbalık arasında ilişki görülmez ve risk oranı 1,36'ya düşer.

(e) Erkekler için $\chi^2 = ,510$. İş kategorileri arttıkça zorbalık düşer, fakat kadınlar için durum farklıdır.

6.39 Appleton, French ve Vanderpump'ın (1990) çalışması.

Genellikle gençler arasında sigara içme eğilimi yaşlılara kıyasla daha fazladır. Çünkü gençler yaşlılara göre daha uzun hayat beklentisine sahiptir, bu onları sanki daha düşük ölüm riski olduğu düşüncesine götürür. Bu, sigara içme etkisini yaş etkisi gibi gösteriyor.

## BÖLÜM 7

7.1 Bu bir grafik örneğidir.

7.3 Örneklemin ortalama ve standart sapması 4,46 ve 2,69'dur. Ortalama ve standart sapma, örneklemin çekildiği evrenin diğer parametrelerine (sırasıyla 4,5 ve 2,7) çok yakındır. Ortalamaların dağılımının ortalaması 4,45, ki evren ortalamasına yakındır ve standart sapma 1,20'dir.

(a) Merkezî limit teoremine göre, örneklem dağılımının ortalamasını 4,45 ve standart sapmasını $2{,}69\sqrt{5} = 1{,}20$ olarak kestiriyorum.

(b) Bu değerler, beklediğimiz değerlere çok yakındır.

7.5 Büyüklükleri 15 olan 50 örneklem çekseydiniz, örneklem dağılımının ortalaması evren ortalamasının yine yaklaşık bir değeri olurdu, fakat dağılımın standart hatası sadece $2{,}69\sqrt{15} = 0{,}69$ olacaktı.

7.7 Son problemde iki yönlü test kullandım, ancak tek yönlü test kullanılabilir olsa da, öğrencilerin ortalamanın altında olduklarını düşünmeleri bizim ilgi alanımızda değildir; fakat ortalamanın üzerinde olduklarını düşünüyor olmaları yönünden tek yönlü test kullanılabilir.

7.9 Grup dibe yakın olsaydı, kesinlikle, performanslarını yüksek kestirmeye göre düşük kestirmek için daha az yer vardı ve gerçekte, çok manidar bulduğunu söyleyenler olduğundan yüksek kestirmektedirler. (Eğer en alt çeyrekte olsalardı, en iyi ihtimalle 25 yüzdelikte puanlamış olabilirler, ortalama kestirim 68. yüzdelikte.)

7.11 Ortalama kazanç = 3,01, standart sapma = 7,3. $t = 2{,}22$. 28 *sd*'de kritik değer = 2,048'dir, böylece sıfır hipotezini reddedeceğiz ve kızlar şans düzeyinden daha çok kazanmışlardır. Etki büyüklüğü 3,01/7,31 = 0,41'dir.

7.13 (a) 27 *sd*'de $t = 20,70$. Sıfır hipotezini reddedebiliriz.

(b) Bu SAT'ın geçerli bir ölçü olmadığı anlamına gelmez, fakat cevapları tahmin etmede iyi olan cevaplayıcılar da SAT'ta iyi puan alırlar. Bu çok sürpriz değildir.

7.15 $GA_{,95} = 3,51 \leq \mu \leq 5,27$. Ortalamayı bu şekilde kapsayan bir aralık ,95 olasılıkla varsayılan evren ortalamasını ihtiva eder; Alıştırma 7.14'teki sonuç ile tutarlıdır.

7.17 Beta- endorphinle ilgili güven aralığı $2,906 \leq \mu \leq 12,494$'tür.

7.19 Erkek ve kadınlar için ortalamalar sırasıyla 2,73 ve 2,79'dur. 90 *sd*'de $t = -0,485$ değeri elde edilir. Erkek ve kadınların eşit şekilde tatmin oldukları hipotezini reddedemeyiz. Eşleştirilmiş örneklem *t* uygundur, çünkü kocanın tatmin olma derecesinin karısının tatmin olma derecesinden bağımsız olduğunu varsaymak mantıklı görünmüyor.

7.21 Kocanın ve karısının puanları arasındaki korelasyon ,334 idi, bu değer manidardır ve puanların ilişkili oluğu varsayımını destekler.

7.23 Önemli soru, ortalamanın (ortalamaların arasındaki farkların) örneklem dağılımının neye benzediğidir ve 91 puanın oluşturduğu örneklemin dağılımı büyük ölçüde ortalamalarının dağılımı normal olacak şekilde süreklidir.

7.25 $t = 2,545$, $p < ,05$. Bu bize müdahele edilen grubun hayat kalitesinin daha iyi olduğunu söyler.

7.27 $GA_{,95} = 0,24 \leq \mu \leq 5,80$

7.29 ($\alpha$) Sıfır hipotezi - pasajı okuyanlarla okumayanların test puanları arasında manidar fark yoktur.

(b) Alternatif hipotez - bu iki durum arasında manidar fark vardır.

(c) Eğer varyansları birleştirirsek 43 *sd*'de $t = 8,89$'dur. Bu fark manidardır.

(d) Öğrencilern pasajları okumaları halinde bu testteki soruları daha iyi cevaplayacakları kararına varabiliriz.

7.31 Kontrol grubundaki kızlar ortalama 0,45 pount kaybettiler, Bilişsel Davranış Terapy grubundakiler de ortalama 3,01 pount kilo aldılar. İki bağımsız grup arasındaki 53 *sd*'de $t = -1,68$ olup manidar değildir. Bilişsel Davranış Terapisi kilo kaybına manidar şekilde yol açmamaktadır. (Varyanslar homojen idi.)

7.33 Bu ortalamalar gerçekten bağımsız örneklemlerden geliyorsa, çiftlerden kaynaklanan farkı kaldıramayız ve elde edilen *t* biraz daha küçük olurdu. Kazandılar.

7.35 Korelasyon oldukça düşüktür.

7.37 (a) Araştırmacının beklentisinin, Normal gruptaki anneler ve çocuklar arasında olumlu ebeveyn-çocuk ilişkilerini gösteren daha fazla hikâyeler olacağını varsayardım.

(b) Normal ve şizofrenik grupların ortalamaları sırasıyla 3,55 ve 2,10 olup 38 sd'de t = 2,66'dır. Bu değer manidardır. (a)'daki araştırma hipotezi desteklenmektedir.

7.39 Böyle şizofrenik çocuk sahibi olmak ebeveyn-çocuk ilişkilerinde bozulmaya neden olabilir. Çünkü biz çocukları gruplara tesadüfî olarak atayamayız, nedensellik hakkında güvenle konuşamayız.

7.41 Alıştırma 7.40'ta güvenlik sınırları: $GA_{,95} = 1,153 \pm (2,131)(1,195) = -3,04 \leq \mu \leq 5,34$. Güvenlik sınırları "0"ı içerdiğinden bu sonuçlar önceki manidar olmayan sonuçlarla tutarlıdır.

7.43 $t = 2,134$. Varyanslar birbirinden çok farklı olduğundan, $n_1 - 1$ ve $n_2 - 1$ serbestlik derecelerinde muhafazakâr *t* testini uygulayabiliriz. Bu test manidar olmayan sonuç üretir. $n_1 + n_2 - 2$ serbestlik derecesini de uygulasaydık manidar olmayan sonuç elde edecektik.

7.45 Eğer gözlemler arasındaki mutlak farkları ve bu iki gruba ait ortalamaların farklarının mutlak değerlerini alır ve bu iki ortalamayı *t* testi ile karşılaştırırsanız $t = 0,625$ elde edersiniz. Bunun karesini alarak $F = 0,391$ bulursunuz, bu sonuç SPSS'te Levene testinin farkların mutlak değerleriyle çalıştığını açık hale gelir. (Farkların karelerine ait *t* değeri 0,213'tür, bu $F = 0,045$ değerini verir.)

7.47 Bilgisayar alıştırması.

7.49 Terapiden önceki ölçümlerin standart sapması kullanılarak hesaplanan etki büyüklüğü (*d*) 0,62 idi. Bu, tedavi süresince standart sapmanın 2/3'ü kadar bir kazanç olduğunu gösterir .

7.51 (a) Ölçmenin ölçeği önemlidir; çünkü, örneğin kategorileri 1, 2, 4 ve 6 olarak tekrar ölçekleseydik çok farklı cevaplar bulurduk.

(b) Birinci alıştırma eşlerin tatmin olma dereceleri arasında bir ilişki olup olmadığını, ikincisi kocaların ortalama olarak eşlerinden daha fazla tatmin olup olmadığını sormaktadır.

(c) Kitapta yapılan tavsiyeyi, bağımsız gruplar ve eşleştirilmiş gruplar üzerine olan *t*'yi birleştirerek uyarlayabilirsiniz.

(d) *t* testi ile çok rahat değilim, çünkü ölçmenin ölçeği ile aram iyi değil. Bir alternatif sıra (rank) testi olabilir, fakat kuyrukların sayısı çok büyük ve muhtemelen bu beni daha çok endişelendirir.

## BÖLÜM 8

8.1 (a) 0,250 (b) 2,5 (c) 0,71

8.3 n = 99, 126 ve 169 (N tam sayı olacak şekilde yuvarladım.)

8.5 Bu bir grafiktir.

8.7 (a) Güç için = ,50, $\delta = 1,95$ ve $n = 15,21 = 16$

(b) Güç için = ,80, $\delta = 2,80$ ve $n = 31,36 = 32$

8.9 (d) = ,50, $\delta = 1,46$ güç = ,31

8.11 $t$ (= -1,98), $t$ istatistikten ve δ parametreden hesaplanmış olmasına rağmen sayısal olarak δ'ya eşittir. Diğer bir deyişle, eğer elde ettiğiniz veriye ait parametreler düşündüğünüz veriye ait parametrelere tam olarak uyuyorsa δ $t$'ye eşittir.

8.13 Bu bir grafik sorusudur.

8.15 PhD adayı, okulu terk edenler grubunu kullanmalıdır. Standart sapmalarının eşit olduğunu varsayarak 25 kişilik H.S okulu terk edenler grubuna ait δ daha yüksek ve böylece güç de daha yüksek olacaktır.

Her iki hesaplama da aynı olmak şartıyla $\sigma$ herhangi bir değer olabilir.

8.17 Güç = ,49.

8.19 $\mu$ = 104,935 olduğunda güç $\beta$'ya eşit olur.

8.21 (a) Katılımcıları gruplara tesadüfî olarak atamazdım, laboratuvarlar arasında sonuçları karıştırıcı farklılıklar olabilirdi. (b) Katılımcılarımı birleştirir ve onları koşullara tesadüfî olarak atardım. (c) Eğer varsa cinsiyet farklılıkları sonuçları değiştirici olacaktır. Cinsiyet farklılıklarını dikkate alan yöntemi (Bölüm 13'e bakınız) kullanırız ve katılımcıları cinsiyetlerine göre ayırır, erkekler ve kadınlar için sonuçlara bakarız.

8.23 Her iki soru dikkatli tasarım yapılmasını gerektirir, böylece sonuçlar açık ve anlaşılabilir olur.

## BÖLÜM 9

9.1 Bu bir grafiktir.

9.3 $n$ = 25 ve ,396 evren korelasyon katsayısına ihtiyacımız olacaktır.

9.5 Aile geliri bebek ölümleri için güçlü bir kestiricidir, bunu aile planlamasına başvuran kadınlar izler.

9.7 Bu değişken çok küçük varyansa sahiptir ve durumlar arasındaki farklar olmadan korelasyon çok büyük olamaz.

9.9 Psikologlar davranışları değiştirme yollarıyla çok ilgilidirler ve bebek ölümlerini azaltmak önemli amaçları arasındadır.

9.11 Pearson $r$ korelasyon katsayısı doğrusal ilişkilerle ilgilenir ve bu kesinlikle eğriseldir.

9.13 Güç = ,17.

9.15 Kestirilen ln(semptom) = 4,37'dir.

9.17 $$s'_{Y.X} = 0,1726\sqrt{1+\frac{1}{107}+\frac{(X_i-\overline{X}^2)}{(N-1)S^2_{\overline{X}}}}$$

$t_{a/2}$ = 1,984

$GA(Y) = \hat{Y} \pm (t_{\alpha/2})S'_{Y.X}$

$\hat{Y}$ *ve* $S'_{Y.X}$'i $X$'in farklı değerleri için hesaplayabilirsiniz ve sonuçların grafiğini çizebilirsiniz.

9.19 Galton'ın sonuçları:

(a) $r$ = ,459

(b) Çocuk = 0,646 x midparent + 23,942

(c) Çocuklar 67,12 68,02 68,71 70,18; Ebeveynler 66,66 68,50 69,50 71,18

(d) En yüksek çeyrekteki ebeveynler, ortalama olarak, çocuklardan daha uzundur. Bunun tersi en alttaki çeyrek için doğrudur.

9.21 Güç için ,80, δ = 2,80. Böylece $N$ = 50.

9.23 (a) $z$ = 0,797. Korelasyonlar manidar olarak farklı değildir. (b) Paragrafı okuyup okumamasının ön test puanları ve başarıyı etkileyip etkilemeyeceği ile ilgili korelasyonu tartışmamıza gerek yoktur.

9.25 Manidar farkın örneklemin büyüklüğüne mi yoksa (ve böylece daha çok fark) r'nin büyüklüğüne mi atfedilebileceğini söylemek zordur. Önceki olması muhtemeldir.

9.27 (a) $r$ = ,224, $p$ = ,509. $H_0$'ı ret etme. (b) İrlandalılar çok sigara içerler, fakat diğer bölgelerle karşılaştırıldığında kesinlikle çok içki içici değillerdir. (c) Kuzey İrlanda'nın dahil edilmesi veriyi bozar. Eğer onları çıkarırsak, $r$ = ,784, p = ,007; sigara içme ile içki içme arasında güçlü bir ilişki vardır.

9.29 (a) Aşağıdaki tabloya bakınız.

(b) Korelasyonların hepsi manidardır, semptomların biri diğerleriyle ilişkilidir.

9.31 (b) Boy uzunluğunda 1 inç artma için ağırlığın 4,356 pount artma beklenir. "0" inç uzunluğundaki bir kimsenin -149,934 pount ağırlık kazanması gerekir. Eğrisellik, uç değerlerde böyle mantıklı olmayan sonuçlar karşımıza çıkarır. Korelasyon ,904'tür, eğim ve korelasyonun ikisi de manidardır.

| | SomT | ObsessT | SensitT | DepressT | AnxT | HostT | PhobT | ParT | PsyT |
|---|---|---|---|---|---|---|---|---|---|
| ObsessT | 0.482 | | | | | | | | |
| SensitT | 0.377 | 0.539 | | | | | | | |
| DepressT | 0.400 | 0.599 | 0.654 | | | | | | |
| AnxT | 0.569 | 0.621 | 0.550 | 0.590 | | | | | |
| HostT | 0.420 | 0.470 | 0.451 | 0.508 | 0.475 | | | | |
| PhobT | 0.466 | 0.509 | 0.613 | 0.568 | 0.528 | 0.411 | | | |
| ParT | 0.400 | 0.524 | 0.677 | 0.621 | 0.547 | 0.494 | 0.540 | | |
| PsyT | 0.334 | 0.503 | 0.625 | 0.725 | 0.509 | 0.404 | 0.529 | 0.651 | |
| GSIT | 0.646 | 0.791 | 0.770 | 0.820 | 0.786 | 0.633 | 0.679 | 0.766 | 0.741 |

9.33 Boyu 5′58″ olan bir erkek olarak benim kestirilen ağırlığım $\hat{Y}$ = 4,356(boy) - 149,934 x 68 - 149,934 = 146,27 pount'tur.

(a) Ben tartıldım 146 paunt geldim. (Birkaç sene önce). Böylece kestirimden kalan $Y - \hat{Y}$ = 146 – 146,27 = -0,27'dir.

(b) Eğer bu denklem, öğrencilerin kendi boy ve ağırlıklarını düşük veya fazla kestiriyorsa, benim ağırlığım geçersiz veriye dayalı olacak ve sistematik olarak hatalı olacaktır.

9.35 Kadının ağırlığı 125,354 pount kestirilmişken erkeğinki 137,562 pount olarak kestirilecektir. Erkekle kadının kestirilen ağırlıkları arasındaki fark 12,712 pount olacaktır.

9.37 Regresyon doğrusu hafif positif eğime sahip olduğu halde, eğim "0"dan manidar olarak farklı değildir. Regresyon doğrusunun denklemi $\hat{Y}$ = 0,429X + 221,843'tür.

9.39 Eris grafiğe benim istediğim kadar uymuyor. Biraz çok uzakta kalıyor.

9.41 $z$ = 3,30. İki korelasyon arasındaki fark manidardır.

## BÖLÜM 10

10.1 (b) $r_{nç}$ = -0,540; $t$ = -2,72 (c) Sabah başarısı manidar şekilde onların algılarının en yüksek olduğu saatlerle ilgilidir.

10.3 Sabah saatlerinde insanların başarıları zaman içinde sanki değişir gibi görünüyor, fakat akşam saatlerinde insanlar düzgün bir şekilde düşük performans gösteriyorlar.

10.5 $t$ = 2,725. Bu $r_{nç}$ üzerinde yapılan $t$ testine eşittir.

10.7 $\hat{Y}$ = 0,202$X$ + 0,093; $X = \bar{X}$ = 2,903 olduğunda $\hat{Y}$ = 0,68 = $\bar{Y}$.

10.9 (b) $\phi$ = ,256 (c) $t$ = 1,27 manidar değil.

10.11 (a) $\phi$ = ,628 (b) $\chi^2$ =12,62, $p$ < ,05

10.13 $\tau$ ,886 (b) z = 4,60, $p$ <,05

10.15 $\tau$ =,735

10.17 $r^2$ = ,0512 $\chi^2$ = 10,24 değerine karşı gelecektir. Eğer katılımcıları, birinci durumda 61/39'u, ikinci durumda 39/61'i olacak şekilde ikiye bölerseniz (tam sayılara yuvarlayarak) gelebileceğiniz en yakın değer bu olur.

10.19 (a) $\chi^2$ = 2,815 [$p$ = ,245] $\phi_c$ = ,087 (c) Eğer depresyon puanlarındaki farklılıkların kliniksel kesme puanlarının altında olduğuna nedeniniz varsa ve önemsiz olduğuna inanıyorsanız göz ardı edilmelidir; bu yöntem Bölüm 7'de kullanılan yaklaşıma tercih edilecektir.

10.21 (b) Eğer bir istatistik manidar değilse, bu, güvenirliğin "0"dan (veya $H_0$ altındaki parametre ne ise) farklı olduğuna inanmak için bir nedenimizin olmadığı anlamına gelir. Burada değişkenler arasında ilişki olduğuna inanmak için bir nedenimiz yoktur. Böylece, bunun önemli olduğuna inanmak için bir nedenimiz yoktur. (c) Güç ile ilgili konular dışında örneklem büyüklüğünün olduğundan daha önemli etkisi olmayacaktır. Sadece manidarlık düzeyini artıracaktır.

## BÖLÜM 11

11.1 Özet Tablo

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| İşlemler | 4 | 351,52 | 87,88 | 9,08 |
| Hata | 45 | 435,30 | 9,67 | |
| Toplam | 49 | 786,82 | | |

11.3 (a)

ANOVA

HATIRLAMA

| | Kareler Toplamı | sd | Kareler Ortalaması | F | Mnd. |
|---|---|---|---|---|---|
| Gruplar arası | 1059,800 | 3 | 353,267 | 53,301 | ,000 |
| Gruplar içi | 238,600 | 36 | 6,628 | | |
| Toplam | 1298,400 | 39 | | | |

(b)

| *Grup* | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Grup | 1 | 792,10 | 792,10 | 59,451* |
| Hata | 38 | 506,30 | 13,324 | |
| Toplam | 39 | 1.298,40 | | |

*p < ,05 [$F_{,05}(1,38)$ = 4,10]

(c) Sonuçları yorumlamak zordur çünkü hata terimi genç ve yaşlı katılımcılar için varyans ihtiva etmektedir. Ayrıca, her iki yaş grubuna uygulanan işlem düzeylerinin etkisini bilmiyoruz.

11.5 (a)

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Grup | 1 | 224,583 | 224,583 | 18,8* |
| Hata | 20 | 238,917 | 11,946 | |
| Toplam | 21 | 463,500 | | |

*p < ,05 [$F_{,05}(1,20)$ = 4,35]

(b) Varyanslar birleştirilmeden $t$ = 4,27 $t^2$ = 18,2

(c) Varyanslar birleştirilerek $t$ = 4,34 $t^2$ = 18,8

(d) Varyanslar birleştirilerek $t$ (a)'daki $F$'e eşdeğerdir.

11.7 $\eta^2$ = ,816; $\omega^2$ = ,796.

11.9 $\eta^2$ = ,182; $\omega^2$ = ,120.

Eta-kare bize yaklaşık olarak semptomların şiddetindeki değişkenliğin %18'inin işlemlere atfedilebileceğini söyler, omega-kare de daha az yanlı kestirmenin %12 olacağını belirtir. Çünkü *F* manidar idi, her iki kestirim de şans düzeyinden iyidir.

11.11

ANOVA

HATIRLAMA

| | Kareler toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Gruplar Arası | 4,280 | 4 | 1,070 | 4,117 | ,005 |
| Gruplar İçi | 14,294 | 55 | ,260 | | |
| Toplam | 18,574 | 59 | | | |

11.13 $\mu$ genel ortalama, $\tau_j$ j'yinci işlemin etkisi ve $e_{ij}$ j'yinci işlemde *i*'yinci katılımcıya ait birim hata olmak üzere $X_{ij} = \mu + \tau_j + e_{ij}$'dir.

11.15 $\mu$ genel ortalama, $\tau_j$ j'yinci işlemin etkisi ve *eij* j'yinci işlemde *i*'yinci katılımcıya ait birim hata olmak üzere $X_{ij} = \mu + \tau_j + e_{ij}$'dir.

11.17

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Grup | 7 | 44,557 | 6,365 | 7,27* |
| Hata | 264 | 231,282 | 0,876 | |
| Toplam | 271 | 275,839 | | |

*p < ,05 [$F_{,05}(7,264) = 2,06$]

11.19 Buraya yazılabilecek bir şey yok.

11.21 $\eta^2 = ,16$; $\omega^2 = ,14$

11.23 Zamanı hıza dönüştürmek ters tarnsformasyon gerektirir. Transformasyonun etkisi büyük değerler arasındaki bağıl uzaklığı azaltır.

11.25 Kelime grupları bir sabit değişkendir. Çünkü kelime gruplarının hangi kısmını kullanacağımıza bir karar veririz. İsim kategorisi içindeki kelimeler çoğunlukla tesadüfîdir, çünkü biz muhtemelen isimleri tesadüfî olarak seçeriz (örneğin kelimedeki harflerin sayısı gibi bazı kısıtlamalarla). İsimleri tesadüfî olarak seçeriz, çünkü "ev", "otomobil", "ağaç" vb.'de olduğu gibi özel olarak değil, genel olarak insanların isimleri nasıl önemsediklerine önem veririz.

11.27 *F* = 4,88 ve sıfır hipotezini reddedebiliriz. Bu soru, ikili farklılıklar değil, üç grup arasında farklılık olup olmadığını belirtir.

11.29 Aralığı dikkate almayarak Epinenuneq .dat'ın analizi

ANOVA

HATALAR

| | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Gruplar Arası | 147,970 | 2 | 73,985 | 36,197 | ,000 |
| Gruplar İçi | 241,187 | 118 | 2,044 | | |
| Toplam | 389,157 | 120 | | | |

11.31 Ortalama üç hata terimi 2,40, 1,83 ve 2,26 olup bunların ortalaması 2,162'dir, bu değer yuvarlama hataları içinde 9 hücreye ait varyansların ortalamasıdır.

11.33 Gouzoulis-Mayfrank ve arkadaşlarının (2000) çalışması

(a)

ANOVA

PERFORMANS

| | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Gruplar Arası | 238,738 | 2 | 119,369 | 7,082 | ,001 |
| Gruplar İçi | 1365,214 | 81 | 16,854 | | |
| Toplam | 1603,952 | 83 | | | |

(b) İkili farklar 3,678, 3,464, 0,214 ve $KO_{hata}$'nın karekökü 4,104'tür. Bu ***d*** değerlerini 0,896, 0,844 ve 0,05 olarak verir.

(c) Mantıksal olarak Ecstacy'nin kontrol veya esrar koşullarının ikisinden de geçici olarak düşük puanlar ürettiği sonucuna varılabilir, ikisi birbirinden fark etmez.

11.35 Verinin büyüklüğü bir etki yapmaz, çünkü $\eta^2$ bağımsız değişkenin metriğinden bağımsızdır.

11.37

DEĞİŞİM

| | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Gruplar Arası | 91,391 | 3 | 130,464 | 12,922 | ,000 |
| Gruplar İçi | 686,570 | 68 | 10,097 | | |
| Toplam | 1077,961 | 71 | | | |

## BÖLÜM 12

12.1 (a)

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| İşlemler | 4 | 816,00 | 204,00 | 36,43* |
| 1,2 e karşı 3,4,5 | 1 | 682,67 | 682,67 | 121,90* |
| 1 e karşı 2 | 1 | 90,00 | 90,00 | 16,07* |
| 3,4 e karşı | 1 | 3,33 | 3,33 | <1 |
| 3 e karşı 4 | 1 | 40,00 | 40,00 | 7,14* |
| Hata | 20 | 112,00 | 5,60 | |
| Toplam | 24 | 928,00 | | |

*p < ,05 [$F_{,05}(1, 20) = 4.35$; $F_{,05}(4, 20) = 2,87$]
*p < ,01 [$F_{,01}(1, 20) = 8.10$; $F_{,01}(4, 20) = 4,43$]

(b) Karşılaştırmaların dikliği:

Katsayıların çapraz çarpımları:

$$\Sigma a_j b_j = (,5)(1) + (,5)(-1) + (,333)(0) + (,333)(0) + (,333)(0) = 0$$

$$\Sigma a_j c_j = (,5)(0) + (,5)(0) + (,333)(,5) + (,333)(,5) + (,333)(-1) = 0$$

$$\Sigma a_j d_j = (,5)(0) + (,5)(0) + (,333)(1) + (,333)(-1) + (,333)(0) = 0$$

$$\Sigma b_j c_j = (1)(0) + (-1)(0) + (0)(,5) + (0)(,5) + (0)(-1) = 0$$

$$c_j d_j = (0)(0) + (0)(0) + (,5)(1) + (,5)(-1) + (1)(0) = 0$$

(c) 682,67 + 90,00 + 3,33 + 40,00 = 816,00

12.3 $a$ = ,05 için: PC = $a$; FW = 1 - (1-a)$^2$ = ,0975

12.5 $q$ = 7,101; $t\sqrt{2}$ = 7,101

12.7 $t'_1$ = -5,34 $H_0$ ret, $t'_2$ = -1,72 $H_0$'ı reddetme.

12.9 Tukey veya REGWQ gibi bir planlı çoklu karşılaştırma testi (post hoc test) sıklıkla düşündüğümüz sorulara ve aynı zamanda ilgilenmediğimiz sorulara cevap vermez.

1211 $F_{\text{karşılaştırma \#1}}$ = 113,26; $F_{\text{karşılaştırma \#2}}$ = 61,57 $F_{\text{kritk}}$ = 4(2,69) = 10,76. Böylece her iki karşılaştırma da manidardır.

12.13 Grup 1 (kontrol), Grup 2, 3 ve 4'ten farklıdır.

12.15 Sırasıyla uyarlanır, çünkü her defasında kritik değeri değiştiriyorsunuz ve başka bir sıfır hipotezini reddediyorsunuz.

12.17 Conti ve Musti (1984) farelerde aktif beyin bölgesinde THC'nin enjeksiyonuna cevap olarak lokomotif davranışı kaydettiler. Ham veri açık bir şekilde grup ortalamalarıyla standart sapmaları arasında doğrusal bir ilişki olduğunu göstermiştir, fakat verinin logaritmik dönüşümü bu ilişkiden uzaklaşmıştır. Lokomotif davranış ortalaması 0,5 mg doza kadar artmış, fakat dozun bundan fazla arttırılması davranışta düşme ile sonuçlanmıştır. Polinomsal eğilim analizi doğrusal bir eğilim olmadığını fakat manidar ikinci dereceden (kuadratik) bir eğilim olduğunu göstermiştir.

12.19 Aralığa atfedilebilecek manidar farklar olsaydı ve aralıkları birleştirseydik bu farklılıklar hata terimi içinde birleşecekti ve güç azalacaktı.

12.21 Üç aralıkta da manidar şekilde doğrusal ve ikinci dereceden (kuadratik) eğilim vardır, bu, orta düzeyde epinephrine'nin hafızayı artırdığını fakat daha büyük dozun düşürdüğünü gösterir. Doğrusal eğilim yüksek dozda epinephrine etkisindeki hayvanların bile hiç epinehrine almadıkları durumuna göre daha kötüdür.

12.23 Güven aralığı:

$$GA_{,95} = (\psi) \pm t_{,05} s_{\text{hata}}$$
$$= (10,67) \pm (2,086)(2,366)$$
$$= 8,735 \le \mu_1 - \mu_2 \le 15,605$$

12.25 Davey ve arkadaşlarının çalışması (2003): Negatif ve pozitif duygu durumları arasındaki karşılaştırma manidardır. ($t(27) = 3,045$, $p < ,05$).
Bu,
$d = \psi / \sqrt{KO_{hata}}$ $5,6/\sqrt{16,907}$ = 5,6/4,11 = 1,36
etki büyüklüğünü verir. Bu iki grup birbirinden 1 tam 1/3 standart sapmadan fazla ayrılır.

12.27 Bu, öğrencilerin tamamlama çalışmaları yapmalarını gerektirir.

## BÖLÜM 13

13.1

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Benzerlik | 1 | 13,067 | 13,067 | 3,354 |
| Büyüklük/Yaş | 2 | 97,733 | 48,867 | 12,541* |
| Ben × *B* | *2* | 17,733 | 8,867 | 2,276 |
| Hata | 54 | 210,400 | 3,896 | |
| Toplam | 59 | 338,933 | | |

*$p < ,05$ $F_{,05(2,54)} = 3,17$

13.3 Daha önce bir doğum yapmış bu anneler, önceden doğum yapmış bütün anneleri temsil etmeyeceğinden, örneklem evreni temsil edici değildir. Örneğin, daha önce doğum yapmış annelerin evreninin %50'si düşük doğum yapmış (LBW) anneler olması beklenmez.

13.5 Alıştırma 13.3 bir faktör (etken) olarak

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Gecikme | 2 | 188,578 | 94,289 | 3,22 |
| Alan | 2 | 356,044 | 178,022 | 6,07* |
| *G* × *A* | 4 | 371,956 | 92,989 | 3,17* |
| Hata | 36 | 1055,200 | 29,311 | |
| Toplam | 44 | 1971,778 | | |

*$p < ,05$ $F_{,05(2,36)} = 3,27$; $F_{,05(4,36)} = 2,64$

13.7 Alıştırma 13.5'te A, Alanı belirttiğine göre:

$\hat{a}_1$ = Doğal alan için işlem etkisi

$\overline{X}_{.1} - \overline{X}_{..1}$ = 24,222 - 28,2 = 3,978

13.9 Grup N'ye karşı Grup *A*: $t = 3{,}03$, Grup *N*'ye karşı Grup *B*: $t = 3{,}00$. $KO_{\text{hata}}$ serbestlik dereceleri 36 olan her her iki $t$ testi için, $a = ,05/2 = 0{,}25$ (iki kuyruklu), kritik değer ± 2,339'dur. Her bir durumda $H_0$'ı reddederiz.

13.11 Alıştırma 11.3'te bir faktör (neden) olarak

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Yaş | 1 | 115,60 | 115,60 | 17,44* |
| Düzey | 1 | 792,10 | 792,10 | 119,51* |
| Yaş × Düzey | 1 | 152,10 | 152,10 | 22,95* |
| Hata | 36 | 238,60 | 6,6278 | |
| Toplam | 39 | 1298,40 | | |

*$p < ,05$ $F_{,05\,(1,36)} = 4{,}12$

13.13 Aşağıdaki tabloya uygun ana etkileri olan fakat etkileşim olmayan bir veri kümesi oluşturunuz.

| Hücre ortalamaları: | 8 | 12 |
|---|---|---|
| | 4 | 6 |

13.15 Nisbett taafından yapılan bir araştırmada başta gelen ilgi konusu insanların yedikleri yiyecek miktarına ilişkin etkileşim idi. Nisbett gösterdi ki obez insanların yediği yiyecek az veya çok göründüğüne göre değişiyordu, normal ağırlıktaki insanların yedikleri ise bu iki durumda aynı idi.

13.17

$$\eta^2_{\text{Benzerlik}} = \frac{KT_{\text{benzerlik}}}{KT_{\text{toplam}}} = \frac{13{,}067}{338{,}933} = ,04$$

$$\eta^2_{\text{büyüklük}} = \frac{KT_{\text{büyüklük}}}{KT_{\text{toplam}}} = \frac{97{,}733}{338{,}933} = ,29$$

$$\eta^2_{\text{Ben.xbüyük.}} = \frac{KT_{\text{Ben.xBüy.}}}{KT_{\text{toplam}}} = \frac{17{,}733}{338{,}933} = ,05$$

$$\omega^2_{\text{Benzerlik}} = \frac{KT_{\text{Benzerlik}} - (p-1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{13{,}067 - (1)3{,}896}{338{,}933 + 3{,}896} = ,03$$

$$\omega^2_{\text{Büyüklük}} = \frac{KT_{\text{büyüklük}} - (\text{büyüklük} - 1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{97{,}733 - (2)(3{,}896)}{338{,}933 + 3{,}896} = ,26$$

$$\omega^2_{\text{Ben.xbüyük.}} = \frac{KT_{\text{Ben.xBüy.}} - (\text{Ben.} - 1)(\text{Büy.} - 1)KO_{\text{hata}}}{KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{17{,}733 - (1)(2)(3{,}896)}{338{,}933 + 3{,}896} = ,03$$

13.19 Alıştırma 13.5'teki öğrenmeden kaçınma verisi için etki büyüklüğü:

$$\eta^2_{\text{GecikmexAlan}} = \frac{KT_{\text{Gecikme}}}{KT_{\text{toplam}}}\frac{188{,}578}{1971{,}778} = ,10$$

$$\eta^2_{\text{Alan}} = \frac{KT_{\text{Alan}}}{KT_{\text{toplam}}}\frac{356{,}044}{1971{,}778} = ,18$$

$$\eta^2_{\text{Gec.xAlan}} = \frac{KT_{\text{Gec.xAlan}}}{KT_{\text{toplam}}}\frac{371{,}956}{1971{,}778} = ,19$$

$$\omega^2_{\text{Gecikme}} = \frac{KT_{\text{Gecikme}} - (\text{gec.} - 1)KO_{\text{hata}}}{KT_{\text{toplam}}\quad KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{188{,}578 - (2)(29{,}311)}{1971{,}778 + 29{,}311}, 06$$

$$\omega^2_{\text{Alan}} = \frac{KT_{\text{Alan}} - (\text{Alan} - 1)KO_{\text{hata}}}{KT_{\text{toplam}}\quad KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{256{,}044 - (2)(29{,}311)}{1971{,}778 + 29{,}311}, 15$$

$$\omega^2_{\text{DelxAlan}} = \frac{KT_{\text{GecxAlan}} - (\text{gec} - 1)(\text{alan} - 1)KO_{\text{hata}}}{KT_{\text{toplam}}\quad KT_{\text{toplam}} + KO_{\text{hata}}} = \frac{371{,}956 - (2)(2)(29{,}311)}{1971{,}778 + 29{,}311} = ,13$$

13.21 Önceki deneme için üç yönlü varyans analizi:

| Kaynak | sd | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Deneyim | 3 | 2931,667 | 977,222 | 3,544* |
| Yoğunluk | 2 | 2326,250 | 1163,125 | 4,218* |
| Koşullu Uyarıcı | 1 | 4563,333 | 4563,333 | 16,550* |
| D x Y | 6 | 67,083 | 11,181 | < 1 |
| D x KU | 3 | 4615,000 | 1538,333 | 5,579* |
| Y x KU | 2 | 55,417 | 27,708 | < 1 |
| D x Y x KU | 6 | 121,250 | 20,208 | < 1 |
| Hata | 96 | 26,471,000 | 275,740 | |
| Toplam | 119 | 41,151,000 | | |

*$p < ,05$ $F_{,05\,(1,96)} = 3{,}94$; $F_{,05(2,96)} = 3{,}09$; $F_{,05(3,96)} = 2{,}70$; $F_{,05(6,96)} = 2{,}19$

Bütün değişkenler ve koşullu uyarıcı etkileşimi için manidar etkiler vardır.

13.23 Epineq.dat için analiz

**Denekler Arası Etkilerin Test Edilmesi**

Bağımsız Değişken: Geri Dönüşüm İçin Denemeler

| Kaynak | Tip III Kareler Toplamı | *sd* | *KO* | *F* | *Mnd.* |
|---|---|---|---|---|---|
| Düzeltilmiş Model | 141,130[a] | 8 | 17,641 | 8,158 | ,000 |
| Kesim | 115,787 | 1 | 153,787 | 533,554 | ,000 |
| DOZ | 133,130 | 2 | 66,565 | 30,782 | ,000 |
| GECİKME | 2,296 | 2 | 1,148 | ,531 | ,590 |
| DOZ* GECİKME | 5,704 | 4 | 1,426 | ,659 | ,622 |
| Hata | 214,083 | 99 | 2,162 | | |
| Toplam | 1509,000 | 108 | | | |
| Düzeltilmiş Toplam | 355,213 | 107 | | | |

[a]R Kare = ,397 (Düzeltilmiş R Kare = ,349)

13.25 Alıştırma 13.25'deki Dozaj verisi üzerinden Tukey:

**Çoklu Karşılaştırmalar**

Bağımlı Değişken: Geri dönüşüme denemeler Tukey HSD

| (I) epinefrine dozu | (J) epinefrine dozu | Ortalama fark (I−J) | Std. Hata | Mnd. |
|---|---|---|---|---|
| 0,0 mg/kg | 0,3 mg/kg | −1,67* | ,35 | ,000 |
| | 1,0 mg/kg | 1,03* | ,35 | ,010 |
| 0,3 mg/kg | 0,0 mg/kg | 1,67* | ,35 | ,000 |
| | 1,0 mg/kg | 2,69* | ,35 | ,000 |
| 1,0 mg/kg | 0,0 mg/kg | −1,03* | ,35 | ,010 |
| | 0,3 mg/kg | −2,69* | ,35 | ,000 |

$p \leq ,05$ olasılıkla bütün gruplar birbirinden farklıdır.

13.27 Alıştırma 13.28'deki veri üzerinde basit etkiler

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Koşul/Durum | 1 | 918,750 | 918,75 | 34,42* |
| Koşul @ Deneyimsiz | 1 | 1014,00 | 1014,00 | 37,99* |
| Koşul @ Deneyimli | 1 | 121,50 | 121,50 | 4,55* |
| Koşul × Deneyim | 1 | 216,750 | 216,75 | 8,12* |
| Diğer Etkiler | 9 | 2631,417 | | |
| Hata | 36 | 961,000 | 26,694 | |
| Toplam | 47 | 4727,917 | | |

*$p < ,05$ $F_{,05(1.36)} = 4,12$

13.29 Kıyafet kodları ve performans:

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Kod | 1 | 494,290 | 494,290 | 2,166 |
| $Hata_1$ | 12 | 2737,280 | 228,107 | |
| Okul (Kodu) | 12 | 2737,280 | 288,107 | 2,784* |
| $Hata_2$ | 126 | 10.323,080 | 81,931 | |
| Toplam | 139 | 13.554,650 | | |

* $p < ,05$

Kod için *F* manidar değildir, fakat okulun iç-içe etkisi manidardır. Fakat iki *F* değeri birbirinden uzakta değillerdir, ama bunların *p* değerleri birbirinden çok farklıdır. Bunun nedeni Test kodu için 12, fakat Okul (Kodu) için 126 serbestlik derecesine sahip olmamızdır.

13.31 Garlett ve Bos'un (2010) analizi

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Ebeveynler | 1 | 832,38 | 832,38 | 44,87* |
| Cinsiyet | 1 | 52,03 | 52,03 | 2,80 |
| E × C | 1 | 110,08 | 110,08 | 5,93* |
| $Hata_2$ | 168 | | 18,55 | |
| Toplam | 171 | | | |

* $p < ,05$

13.35 Bu sorunun sabit bir cevabı yoktur.

## BÖLÜM 14

14.1 (a)

$$X_{ij} = \mu + \pi_j + \tau_j + \pi\tau_{ij} + e_{ij} \; or$$
$$X_{ij} = \mu + \pi_j + \tau_j + e'_{ij}$$

(b)

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Katılımcılar | 7 | 189.666,66 | | |
| Katılımcılar içi | 16 | 5266,67 | | |
| Test oturumu | 2 | 1808,33 | 904,17 | 3,66 Mnd |
| Hata | 14 | 3458,33 | 247,02 | |
| Topam | 23 | 194.933,33 | | |

(c) Oturum ortalamaları arasında manidar fark yoktur - puanlar deneyimin fonksiyonu olarak artmamaktadır.

14.3

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 19 | 106,475 | | |
| Gruplar | 1 | 1,125 | 1,125 | 0,19 |
| *KT* grup içi | 18 | 105,250 | 5,847 | |
| Denekler içi | 20 | 83,500 | | |
| Evre | 1 | 38,025 | 38,025 | 15,26* |
| $E \times G$ | 1 | 0,625 | 0,625 | 0,25 |
| $E \times Kt$ | 18 | 44,850 | 2,492 | |
| Gruplar içi | | | | |
| Toplam | 39 | 189,975 | | |

$p < ,05$ $[F_{,05(1.18)} = 4,41]$

Eğitimin başlangıcında manidar bir değişiklik vardır, fakat iki grup arasındaki fark ayırıcı değildir ve gruplar arasında fark yoktur.

14.5

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Denekler arası | 29 | 159,7333 | | |
| Gruplar | 2 | 11,4333 | 5,7166 | 1,04 |
| *Kt* Gruplar içi | 27 | 148,300 | 5,4926 | |
| Denekler içi | 30 | 95,0000 | | |
| Evre | 1 | 19,2667 | 19,2667 | 9,44* |
| $E \times G$ | 2 | 20,6333 | 10,3165 | 5,06* |
| E × Kt Grp. içi | 27 | 55,1000 | 2,0407 | |
| Toplam | 59 | 254,733 | | |

(b) $KT_{\text{iyi yetişkinlerde maddeler}}$ 4,133; $F = 3,52$*

14.7 Alıştırma 14.6'daki veri için:

(a) Sadece Çoçuklar için olan veriyi kullandığımızdan birleştirilmiş hata terimini kullanmamak uygun olacaktır. Aşağıdaki okuyuculara ait denekler-arası etkiler için ilgili SPSS çıktısıdır.

**Denekler-Arası Etkiler Testleri***

Ölçüm: ÖLÇÜM_1

Dönüştürülmüş Değişken: Ortalama

| Kaynak | Tip III Kareler Toplamı | *sd* | Ortalamalar Toplamı | *F* | Mnd. |
|---|---|---|---|---|---|
| Kesim noktası | 67,500 | 1 | 367,500 | 84,483 | ,000 |
| OKUYUCULAR | 50,700 | 1 | 50,700 | 11,655 | ,009 |
| Hata | 34,800 | 8 | 4,350 | | |

[a]YAŞ = Çocuk

(b) Tekrar, hata terimlerini birleştirmek istemiyoruz. Aşağıdaki, Yetişkenlerde iyi okuyucularla ilgili uygun SPSS çıktısıdır. Fark manidar değildir, eğer düzeltme yaparsak serbestlik derecesinde bir azalma olmayacak ve sonuç aynı olacaktır.

**Denekler-İçi Etkiler Testleri**

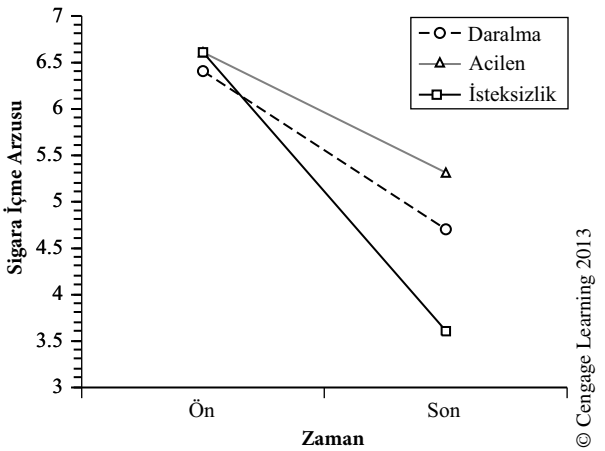
Ölçüm: ÖLÇÜM_1

Küresellik Varsayılarak

| Kaynak | Tip III Kareler Toplamı | *df* | Ortalamalar Toplamı | *F* | Mnd. |
|---|---|---|---|---|---|
| MADDELER | 4,133 | 2 | 2,067 | 3,647 | ,075 |
| Hata(MADDELER) | 4,533 | 8 | ,567 | | |

14.9 Kesin olarak kovaryansı etkiler, çünkü maddeler arasında yüksek düzeyde kovaryansı zorlardı. Bir maddenin bir düzeyine ait sınıflandırılmış cevapların sayısı arttığında, diğer madde aşağıya düşecektir.

14.11 Alıştırma 14.10'un sonuçlarına ait grafik:




14.13 Alıştırma 14.5'teki verinin alalizi:

(b) *F*, ortalama için $H_0 = 0$ hipotezina ait bir testtir.

(c) $KO_{\text{hücre içi}}$ hücre varyanslarının ortalamasıdır.

14.15 Tekrarlı ölçümler A ve B ve bağımsız ölçümler C ve D değişkenlerine ait dört-yönlü ANOVA için özet tablonun kaynak sütunu aşağıdaki gibidir:

| Kaynak |
|---|
| Denekler arası (Denekler arası *Kt*) |
| C |
| D |
| CD |
| *Kt* (Gruplar içi kareler toplamı) |
| Gruplar içi Kt |
| *A* |
| *AC* |
| *AD* |
| *ACD* |
| A X Gruplar içi Kt |
| *B* |
| *BC* |
| *BCD* |
| B X Gruplar içi Kt |
| *AB* |
| *ABC* |
| *ABD* |
| *ABCD* |
| AB X Gruplar içi Kt |
| Toplam |

14.17 Karma modeller yöntemini kullanarak.

Kovaryans matrisi sağa doğru gidildiğinde azalan bir yapı gösterir. Bu yüzden kovaryans matrisini içeren çözüm istedim. Bu AR1'dir. Çözüm aşağıdaki gibidir:

**Sabit Etkiler**

Sabit Etkilerin Tip III Testleri[a]

| Source | Pay *sd* | Payda *sd* | *F* | Mnd. |
|---|---|---|---|---|
| Kesim | 1 | 43,256 | 422,680 | ,000 |
| Grup | 2 | 43,256 | 3,521 | ,038 |
| Zaman | 2 | 81,710 | 71,356 | ,000 |
| Grup × Zaman | 4 | 81,710 | 5,578 | ,001 |

[a]DependentVariable: dv.

14.19 Eşit büyüklükte olmayan alıştırma ile karma model analizi:

Sabit Etkilerin Tip III Testleri[a]

| Kaynak | Pay *sd* | Payda *sd* | *F* | Mnd. |
|---|---|---|---|---|
| Kesim | 1 | 41,724 | 393,118 | ,000 |
| Grup | 2 | 41,724 | 2,877 | ,068 |
| Zaman | 2 | 70,480 | 64,760 | ,000 |
| Grup × Zaman | 4 | 70,459 | 5,266 | ,001 |

[a]DependentVariable: dv.

14.21 Kazanılan puanlara ait SPSS çıktısı:

**Denekler-Arası Etkilere Ait Testler**

Bağımlı Değişken: KAZANÇ

| Kaynak | Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Düzeltilmiş Model | 614,644[a] | 2 | 307,322 | 5,422 | ,006 |
| Kesim | 732,075 | 1 | 732,075 | 12,917 | ,001 |
| İŞLEM | 614,644 | 2 | 307,322 | 5,422 | ,006 |
| Hata | 3910,742 | 69 | 56,677 | | |
| Toplam | 5075,400 | 72 | | | |
| Düzeltilmiş toplam | 4525,386 | 71 | | | |

[a] *R* Kare = ,136 (Düzeltilmiş *R* Kare = ,111)

(b) Ön test ve son teste ait SPSS çıktısı:

**Denekler-Arası Testler**

Ölçüm: ÖLÇÜM_1

Küresellik Varsayıldı

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| ZAMAN | 366,037 | 1 | 366,037 | 12,917 | ,001 |
| ZAMAN × İŞLEM | 307,322 | 2 | 153,661 | 5,422 | ,006 |
| Hata (ZAMAN) | 1955,371 | 69 | 28,339 | | |

(c) Kazanılan puanların üzerinden hesaplanan *F* ile tekrarlı ölçümlerde etkileşim için hesaplanan *F* aynıdır.

(e) $t = ,287$. Bu grup çalışma boyunca verilen kurstan manidar bir kazanç elde etmedi. Bu, diğer gruplarda gördüğümüz herhangi bir kazanç, yaşın bir fonksiyonu olarak görülen normal kazanca atfedilemez.

(f) Kontrol grubu olmadan treapiden ve olgunlaşmadan gelen kazançları ayıramayız.

14.23 (a) $t = ,555$. Zaman 1 puanları arasında fark yoktur ve Zaman 2 puanları arasında fark vardır.

(b) Eğer fark olsaydı, öğrencilerin okulu tesadüfî olarak terk ettiklerinden endişe ederdim.

14.25 Hakimlerden gelen farklılıklar önemli rol oynar.

14.27 Eğer özel olarak denekler arasındaki farklarla ilgilenseydim ve hakimlerin nuhtemelen iyi bir dayanma noktasının olmadığını ve bu eksikliğin manidar olmadığını kabul etseydim bununla ilgilenmezdim.

14.29 Strayer ve arkadaşları (2006)

$t_{1\text{-}2} = 0{,}07$; $t_{1\text{-}3} = 2{,}52^*$; $t_{2\text{-}3} = 2{,}45^*$.

Başlangış ve alkol durumlarının her ikisi de cep telefonu durumuna göre zayıf performans gösterir, fakat birbirinden farklı değillerdir.

**Denekler-İçi Etkilerin Testleri**

Ölçüm: ÖLÇÜM 1

| Kaynak | | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|---|
| Koşul/Durum | Küresellik Varsayılarak | 134696,067 | 2 | 67348,033 | 4,131 | ,020 |
| | Greenhouse-Geisser | 134696,067 | 1,992 | 67619,134 | 4,131 | ,020 |
| | Huynh-Feldt | 134696,067 | 2,000 | 67348,033 | 4,131 | ,020 |
| | Alt-sınır | 134696,067 | 1,000 | 134696,067 | 4,131 | ,049 |
| Hata (Koşul/Durum) | Küresellik Varsayılarak | 1271689,267 | 78 | 16303,709 | | |
| | Greenhouse-Geisser | 1271689,267 | 77,687 | 16369,337 | | |
| | Huynh-Feldt | 1271689,267 | 78,000 | 16303,709 | | |
| | Alt-sınır | 1271689,267 | 39,000 | 32607,417 | | |

## BÖLÜM 15

15.1 Hayat Kalitesinin Kestirimi:

(a) Diğer değişkenler sabit tutularak hava sıcaklığında +1 derecelik artış algılananan Yaşam Kalitesinde -,01 kadar değişikliğe neden olur. Gelirdeki ortanca 1000 $ artış diğer bütün değişkenler sabit tutulduğunda algılananan Yaşam Kalitesinde +,05 ile ilgilidir. Benzer açıklamalar $b_3$ ve $b_4$ için de geçerlidir. Çünkü 0,00 değeri diğer bütün kestiriciler için de geçerli değildir, kesişmeler yorumlama için anlamlı değildir.

(b) $\hat{Y}$ = 5,37 - ,01(55) + ,05(12) + ,003(500) -,01(200) = 4,92.

(c) $\hat{Y}$ = 5,37 - ,01(55) + ,05(12) + ,003(100 ) -,01(200) = 3,92.

15.3 Sıcaklığı silerdim, çünkü en düşük $t$ ($t$ = -1,104) ve böylece bağımlı değişkenle en küçük yarı kısmî korelasyona sahiptir.

15.5 (a) Ölçüt ile en yüksek kısmî korelasyona sahip olan çevredir, çünkü en büyük t değeri çevreye aittir.

(b) Kestirmede bütün kestiricileri kullanarak elde ettiğimiz kazanç ($r$ = ,58'den $R$ = ,697'ye kadar), $p$'nin $N$'ye bağlı olarak büyümesiyle sürdürdüğümüz güç kaybıyla telâfi edilemez.

15.7 İki değişken arasındaki korelasyon azaldıkça üçüncü değişkenle olan varyans azalmayı paylaşır. Böylece, her bir değişkenin ölçütle olan muhtemel yarı-kısmî korelasyonun karesi artacaktır. Onların her biri daha önceden açıklanamamış varyans için açıklama getirebilir.

15.9 Nemsop ve Respon diğer yordayıcılarla oldukça iyi ilişkilidir; YRS onlardan hemen hemen bağımsızdır.

15.11 $KO_{\text{kalan}}$ = 4,232.

15.13 Düzeltilmiş $R^2$ = kestirilen $R^{2*}$ = -,158. Çünkü karesi alınan bir değer negatif olamaz. Bütün bunların hepsi, $H_0$: $R^2$ = 0 hipotezini reddedemeyeceğimiz gerçeğinin ışığında daha mâkûldür.

15.15 $\hat{Y}$ = 0,4067Respon + 0,i1845NumSup + 2,3542.

15.17 İlgi alanımızda olan evrene (10 bölge) ait verimiz için bunun anlamı yoktur.

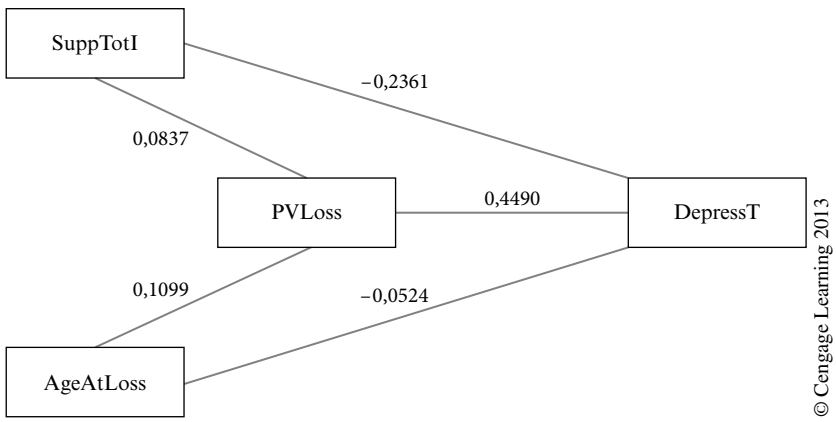
15.19 Onun korelasyonu, onun, diğer değişkenlerin artık bileşenleriyle olan korelasyonları aracılığıyla önemli rol oynar.

15.21 Çoklu-regresyon denklemi bağlamında bir değişkene kendi başına bakamayız. Bir değişkene ait eğim sadece diğer değişkenler sabit tutulduğunda o değişkene aittir.

15.23 Bu sorunun sabit bir cevabı yoktur.

15.25 (b) $R^2$ değeri gerçekte etkisizdir. Bununla beraber, PVLoss için regresyon katsayısının standart hatası 0,105'ten 0,108'e yükseldiğinde. PVLoss için tolerans ,981'den ,345'e düşer, oysa VIF 1,019'dan 2,900'e yükselir. (c) PVTotal modele dahil edilmemelidir çünkü diğer değişkenlerin yanında gereksizdir.

15.27




15.29 104 numaralı durumda Cook D'nin değeri en büyüktür (,137), fakat Student'laştırılmış kalan çok büyük değildir ($t$ = -1,88). Bunu silersek çoklu regresyonun karesi biraz yükselir. Daha önemlisi, regresyonun standart hatası ve kestiricilerden birinin (PVLoss) standart hatası biraz azalır. Bu durum, veri üzerinde büyük bir etki yaratması için yeterli büyüklük değildir.

15.31 Haras.dat üzerinde lojistik regresyon

Manidar olarak katkıda bulunan tek yordayıcı Wald $\chi^2$ = 26,43 değeriyle saldırganlık davranışıdır. Regresyon katsayısının kuvveti 1,649'dur. Bu, saldırganlık davranışındaki 1-puanlık artışın, rapor edilen olasılığın 1,65 ile çarpılacağı anlamına gelir.

15.33 Davranışın frekansı onun kendi saldırganlığıyla bağlantılıdır, o da rapor edilen olasılıkla ilişkilidir.

15.35 BlamPer ve BlamPch orta düzeyde ($r$ = ,52) ilişkilidirler ve bu durumu bir kere denkleme dahil ettiğimizde BlamPch'yi açıklamak için geriye az bir şey kalır.

15.37 Bu problem sizden bir örnek oluşturmanızı gerektiriyor.

15.39 Etkileşimde rol oynayan değişkenlerden birini değiştirmeden etkileşimi değiştirmek mümkün değildir. Bu durumda, diğer bütün değişkenleri sabit tutarak herhangi anlamlı bir yorum getirmek mümkün değildir.

15.41 Feinberg ve Willer'dan (2011) elde edilen sonuçların analizi.

Aşağıdaki çıktı, bu bölümde anılan Preacher ve Leonardelli tarafından yazılan programdan gelir. Regresyon katsayısından $t$ değerlerini ve standart hatalarını hesapladım, daha sonra $t$ değerlerini programa yerleştirdim. Aracı yolun, standart hatanın ne olduğuna bakmadan, manidar olduğunu görebilirsiniz.

a yolu üzerinden = 2,27 b yolu üzerinden = −4,90

| Test | Sobel | Aroian | Goodman |
|---|---|---|---|
| İstatistik | 2,059 | 2,025 | 2,10 |
| $p$ | ,039 | ,043 | ,036 |

## BÖLÜM 16

16.1

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Grup | 2 | 57,733 | 28,867 | 9,312* |
| Hata | 12 | 37,200 | 3,100 | |
| Toplam | 14 | 94,933 | | |

* $p < ,05$ $[F_{.05(2,12)} = 3,89]$

16.3 (a)

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Grup | 2 | 79,0095 | 39,5048 | 14,92* |
| Hata | 18 | 47,6571 | 2,6476 | |
| Toplam | 20 | 126,6666 | | |

*$p < ,05$ $[F_{,05(2,18)} = 3,55]$

16.5

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Gender | 1 | 65,333 | 65,333 | 7,730* |
| SES | 2 | 338,667 | 169,333 | 20,034* |
| $G \times S$ | 2 | 18,667 | 9,333 | 1,104 |
| Error | 42 | 355,000 | 8,452 | |
| Total | 47 | 777,667 | | |

16.7

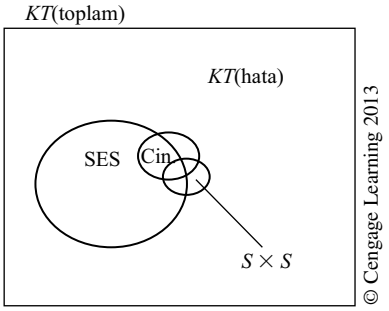
| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Cinsiyet | 1 | 60,015 | 60,015 | 7,21* |
| SES | 2 | 346,389 | 173,195 | 20,80* |
| C × S | 2 | 21,095 | 10,547 | 1,27 |
| Hata | 35 | 291,467 | 8,328 | |
| Toplam | 40 | | | |

* $p < ,05$ [$F$,05(1,35) = 4.12; $F$,05(2,35) = 3.27]

16.9 $\hat{\mu}$ = 13,4167; $\alpha_1$ = 1,167; $\beta_1$ = –3,167
$\beta_2$ = –0,167; $\alpha\beta_{11}$ = 0,833; $\alpha\beta_{12}$ = –0,167

16.11 Eğer gerçekten ağırlıklandırılmamış ortalamalarla ilgileniyorsak, $KT_A$ ve $KT_B$ "0" olacaktır, çünkü bütün satır ve sütunların ortalamalarının ortalaması 7'dir.

16.13




© Cengage Learning 2013

16.15 (a) Her grubun sadece ilk girdisi kullanılarak gösterme amacıyla elde edilen tasarım matrisi:

$$X = \begin{bmatrix} 1 & 0 & 58 & 75 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 60 & 70 \\ \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & 75 & 80 \end{bmatrix}$$

(b)

| Kaynak | *sd* | *KT* | *KO* | *F* |
|---|---|---|---|---|
| Birl. Değ. Kov. | 1 | 1250,6779 | 1250,6779 | 55,81* |
| A (Grup) | 2 | 652,9228 | 326,4614 | 14,57* |
| Hata | 11 | 246,5221 | 22,4111 | |
| Toplam | 14 | 2615,7333 | | |

* $p < ,05$ [$F_{,05(1,11)} = 4,84$; $F_{,05(2,11)} = 3,98$]

16.17 $\hat{Y}$ = -7,9099$A_1$ + 0,84861$A_2$ - 2,4022$B$
+ 0,5667$AB_{11}$ + 0,1311$AB_{21}$ + 0,7260$C$
6,3740

Düzeltilmiş ortalamalar 41,1556, 44,8276, 49,5095, 54,8333 ve 61,0333.

16,19 Klemchuk, Bond ve Howel (1990) (bu sayfanın altında).

16.21 Mireault.dat için GSIT'nin analizi (arka sayfada).

16.23 Bu veri erkekler ve kadınlar arasında Year Coll. (Üniv-Yıl) yönünden manidar sonuç verir. Kadınlar erkeklerin biraz önündedir. Eğer kolejin ilk yılı sonraki yıllardan daha stresli ise Alıştırma 16.21'de bulduğumuz farkın bir kısmı buradan gelebilir.

16.25 Everitt iştahsızlık için iki terapi ve bir kontrol grubunu karşılaştırmıştır. Gruplar ön test ağırlıkları kontrol edilerek karşılaştırıldığında manidar şekilde farklı bulunmuştur ($F$ = 8,71, $p < ,0001$),

Alıştırma 16.19 İçin Sonuçlar

Bağımlı Değişken: DV

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Düzeltilmiş model | 15,7283 | 3 | 5,243 | 8,966 | ,000 |
| Kesim | 2,456 | 1 | 2,456 | 4,201 | ,048 |
| Günlük bakım | 2,640 | 1 | 2,640 | 4,515 | ,041 |
| Yaş | 11,703 | 1 | 11,703 | 20,016 | ,000 |
| Günlük bakım* Yaş | ,037 | 1 | ,037 | ,064 | ,802 |
| Hata | 21,050 | 36 | ,585 | | |
| Toplam | 46,111 | 40 | | | |
| Düzeltilmiş toplam | 36,778 | 39 | | | |

[a]$R$ Kare = ,428 (Düzeltilmiş $R$ Kare = ,380)

Alıştırma 16.19 İçin Sonuçlar

**Tests of Between-Subjects Effects**

Bağımlı Değişken: GSIT

| Kaynak | Tip III Kareler Toplamı | *sd* | Kareler Ortalaması | *F* | Mnd. |
|---|---|---|---|---|---|
| Düzeltilmiş Model | 1216.924[a] | 5 | 243.385 | 2.923 | .013 |
| Kesim | 1094707.516 | 1 | 1094707.516 | 13146.193 | .000 |
| CİNSİYET | 652.727 | 1 | 652.727 | 7.839 | .005 |
| GRUP | 98.343 | 2 | 49.172 | .590 | .555 |
| CİNSİYET × GRUP | 419.722 | 2 | 209.861 | 2.520 | .082 |
| Hata | 30727.305 | 369 | 83.272 | | |
| Toplam | 1475553.000 | 375 | | | |
| Düzeltilmiş Toplam | 31944.229 | 374 | | | |

[a] *R* Kare = ,038 (Düzeltilmiş *R* Kare = ,025)

İki işlem grubunun sadece son testteki ağırlıkları farkını karşılaştırırsak, ortalamaların farkı için etki büyüklüğü (tekrar ön-test ağırlıkları kontrol edilerek) ,62 olmasına rağmen $F$ manidar büyüklüğe ulaşmıyor, $F$ = ,745, $p$ = ,060. Aile Terapi Grubu, Bilişsel Terapi Grubu'ndan yaklaşık altı pount fazladır. Burada manidar olmayan $F$ değerinin yorumlanması oldukça güçtür.

16.27 Eğimin 1,00 olması işlemin, öğrencilerin ön-test puanlarına bir sabit eklemesi anlamına gelir; bu ihtimal dışı gibi görünüyor. Öğrenciler kitaptan herhangi bir ön-test ve son-test puanları kümesi alarak ilişkinin grafiğini çizebilirler.

Fark puanları ve kovaryans analizi sonuçları genel olarak kovaryans analizinin tercih edilebileceğini gösterir. Fark puanın gerçekten ilgi ölçüsü olduğu durumda farklı düşünebilirim.

## BÖLÜM 17

17.1-5 Mazzucchelli ve arkadaşlarının (2010) çalışması

| Yazar | Grup | nl | n2 | £ | sg^2 | Ağırlık | Wg | W × g^2 | W^2 | V(gi-g |
|---|---|---|---|---|---|---|---|---|---|---|
| Barlow86a | E | 12,00 | 12,00 | −0,134 | 0,2740 | 3,6496 | −0,4891 | 0,0655 | 13,3198359 | 1,4609 |
| Besyner79 | E | 14,00 | 16,00 | 0,675 | 0,1790 | 5,5866 | 3,7709 | 2,5454 | 31,21001217 | 0,1737 |
| Lovett88 | E | 33,00 | 27,00 | 0,204 | 0,0930 | 10,7527 | 2,1935 | 0,4475 | 115,6203029 | 0,9338 |
| Stark | E | 10,00 | 9,00 | 0,043 | 0,1930 | 5,1813 | 0,2228 | 0,0096 | 26,84635829 | 1,0759 |
| VanDenHa | E | 15,00 | 14,00 | 0,644 | 0,1380 | 7,2464 | 4,6667 | 3,0053 | 52,5099769 | 0,1530 |
| Weinberg | E | 10,00 | 9,00 | 0,976 | 0,2180 | 4,5872 | 4,4771 | 4,3696 | 21,04199983 | 1,0451 |
| Wilson | E | 9,00 | 11,00 | 1,466 | 0,2750 | 3,6364 | 5,3309 | 7,8151 | 13,2231405 | 3,4025 |
| TOPLAM | | 103,00 | 98,00 | | | 40,6402 | 20,1729 | 18,2580 | 273,7716 | |
| Barlow86a | M | 12,00 | 13,00 | 0,133 | 0,1240 | 8,0645 | 1,0726 | 0,1427 | 65,0364204 | 1,0784 |
| Fordyce77 | M | 50,00 | 60,00 | 0,609 | 0,0380 | 26,3158 | 16,0263 | 9,7600 | 692,5207756 | 0,3202 |
| Fordyce83 | M | 40,00 | 13,00 | 1,41 | 0,2330 | 4,2918 | 6,0515 | 8,5326 | 18,41993774 | 3,5644 |
| Reich81 | M | 49,00 | 49,00 | 0,378 | 0,0320 | 31,2500 | 11,8125 | 4,4651 | 976,5625 | 0,4552 |
| TOPLAM | | 151,00 | 135,00 | | | 69,9222 | 34,9629 | 22,9004 | 1752,5396 | |
| Genel Toplam | | 254,00 | 233,00 | | | 110,5623 | 55,1358 | | 2026,3113 | 13,6631 |

Ortalama g = 0,4987 Q = 13,6631
se(Ortalama) 0,0951

GA-Alt sınır 0,3123 C = 92,2350
GA-Üst sınır 0,6851

Tau (To) = 0,1993

Sabit-Etkiler Modeli (k = 11)

Heterojenlik Testi:
Q(sd = 10) = 13,6678, p-değ = 0,1887)

Model Sonuçları:

| kestirim | std hata | z değ | p değ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| 0,4987 | 0,0951 | 5,2428 | <,0001 | 0,3122 | 0,6851**** |

...

Mnd.kodlar: 0‘***’ 0,001‘**’ 0,01‘*’ 0,05 ‘.’ 0,01 ‘’ 1

**17.6 - 17.12** Bloch ve arkadaşlarının (2009) çalışması

Sabit-Etkiler Modeli (k = 4)

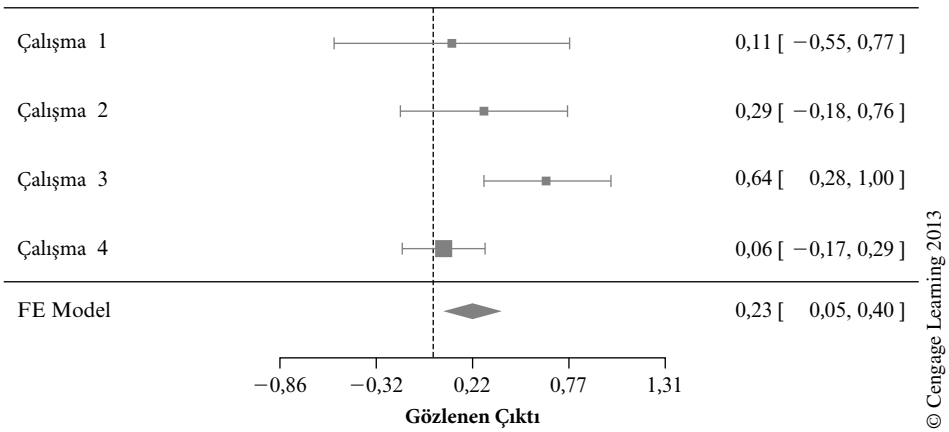
Heterojenlik Testi:
Q(sd = 3) = 7,2655, p-değ =0,0639)

Model Sonuçları:

| kestirim | std hata | z değ | p değ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| 0,2274 | 0,0881 | 2,5813 | 0,0098 | 0,3122 | 0,4001** |

...

Mnd.kodlar: 0‘***’ 0,001‘**’ 0,01‘*’ 0,05 ‘.’ 0,01 ‘’ 1




Tesadüfî bir modele uydurmaya çalışmak mantıklı değildir, çünkü makul ölçüde tesadüfîlik testi yapabileceğimiz çok az sayıda çalışma var.

Güven aralığı "0"ı içine almıyor ve sıfır hipotezini güvenle reddedebiliriz ve methilypnidate'in bu çocukların tiklerinin şiddetini artırdığı kararına varabiliriz.

Sabit-Etkiler Modeli (k = 3)

Heterojenlik Testi:
Q(sd = 2) = 2,1121, p-değ = 0,3478

Model Sonuçları:

| kestirim | std.hata | zdeğ | pdeğ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| 0,7364 | 0,0955 | 7,7109 | <0,0001 | 0,5492 | 0,9236*** |

...

Mnd. kodları: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

**17-13 - 17.14**

Orman Grafiği Atlandı

Sabit-Etkiler Modeli (k = 3)

Heterojenlik Testi:

Q(sd = 8) = 2,1826, p-değ = 0,9749

Model Sonuçları:

| kestirim | std.hata | zdeğ | pdeğ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| 0,5239 | 0,2826 | 1,8542 | 0,0637 | –0,0299 | 1,1077 |

...

Mnd. kodları: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

**17.15 - 17.19 Rajkumar (2010)**

Risk oranları ve log risk oranları aşağıdaki gibidir:

Risk Oranları
4,102326 6,336000 8,212389 1,963636

Log Risk Oranları
1,411554 1,846248 2,105644 0,674798

Ortalama risk oranları ve güven sınırları

**Log Risk Oranı**

| kestirim | std.hata | zdeğ | pdeğ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| 1,5747 | 0,3277 | 4,8055 | <0,0001 | 0,9324 | 2,2170 |

| **Risk Oranı** | **GAalt** | **GAüst** |
|---|---|---|
| 4,8293 | 2,5406 | 9,1798 |

Hatta güven aralığının alt sınırının bitim noktasını ve uyuşturucu başarı şansını, kontrol grubunun başarı şansının 2,5 katına çıkarır.

**17.20 - 17.24 Bisson ve Martin (2007)**

Tesadüfî-Etkiler Modeli (k = 14; tau^2 kestirici: REML)

tau^2 (toplam heterojenlik miktarının kestirimi): 438,6370 (STANDART HATA = 189,2833)

tau (toplam heterojenliğin kestiriminin kare kökü) = 20,9437

1^2 (heterojenlikten gelen toplam değişkenliğin yüzdesi) = %94,80

H^2 (toplam değişkenlik/çalışma-içi varyans): 19,24

Heterojenlik Testi

Q(sd = 13) = 236,1772, p-değ < 0,001

Model Sonuçları:

| kestirim | std.hata | zdeğ | pdeğ | ci.lb | ci.ub |
|---|---|---|---|---|---|
| -28,2812 | 5,8774 | -4,8697 | <0,0001 | -40,1407 | -17,1017*** |

Grafik atlandı.

Bu değişkenliği açıklayabilecek özel bir değişken olmamasına rağmen, heterojenlik testi için sıfır hipotezini reddedebiliriz. Aynı zamanda VBT'nin Kontrol işlemine göre daha etkili bir işlem olduğu kararına varabiliriz.

| | | | | |
|---|---|---|---|---|
| Ortalama(A) | 13,6 | 11 | 22,4 | |
| Std.Sap. | 1,817 | 2 | 2,51 | |
| Ortalama(B) | 5,8 | 5 | 20,2 | |
| Std.Sap. | 5,007 | 3,8 | 2,044 | |
| s(birleşik) | 4,286260671 | 3,350774882 | 2,197933015 | |
| d | 1,819767998 | 1,790630589 | 1,000940422 | |
| s(d) | 0,640613133 | 0,637870361 | 0,577404587 | |
| GA alt | 0,564166258 | 0,540404682 | −0,130772569 | |
| GA üst | 3,075369738 | 3,040856496 | 2,132653414 | |
| Ağırlık | 2,436735134 | 2,457735569 | 2,999435588 | 7,893906291 |
| Widi | 4,434292616 | 4,40089649 | 3,002256323 | 11,83744543 |
| dbar | 1,499567514 | | | |
| s(dbar) | 0,355921331 | | | |
| GA alt(dbar) | 0,801961705 | | | |
| GA üst(dbar) | 2,197173322 | | | |

Üç denekten ikisi manidar gelişme göstermiş (bunlara ait güven aralığı "0"ı içine almıyor ve aynı zamanda genel güven aralığı da "0"ı içine almıyor, bu da manidar genel gelişmeyi gösteriyor).

**17.25 - 17.28** Howard ve Kendal (1996)

Grafikler atlandı.

$\bar{d}$ = -1,0435

$s_{\bar{d}}$ = 0,2114

$-1{,}458 \le \bar{d} \le -0{,}629$

**17.29 – 17.31** Dunlop ve Fox (999)

Bireysel regresyonlar - sadece birinci denek için gösterilmiştir.

**Denek 1**

Katsayılar:

| | Kestirim | Std. Hata | t değeri | Olasılık(>\|t\|) | d | Wt |
|---|---|---|---|---|---|---|
| (Kesim) | 5,000e+01 | 2,182e+01 | 2,291 | 0,0342* | | |
| Evre | −2,482e+01 | 2,272e+01 | −1,092 | 0,2891 | 0,515 | 0,002 |
| Deneme | −7,106e−15 | 1,010e+01 | 0,000 | 1,0000 | 0,000 | 0,010 |
| Aralık1 | −1,447e+00 | 1,012e+01 | −0,143 | 0,8879 | 0,943 | 0,010 |

Kalan standart hata: 18 serbestlik derecesinde 14,29

Çoklu R-kare: 0,5717, Düzeltilmiş R-kare: 0,5003

F istatistiği: 3 ve 18 sd'de 8,008, p-değeri: 0,000

*t* ve *d* sütunlarından bunları ayrı ayrı alarak *d* değerlerinin pek çoğunun bir dereceye kadar yüksek olmasına rağmen manidar farklılığın sadece 3 numaralı Deneğe ait eğimdeki değişimde olduğunu görürüz.

*d*'nin ortalamasını ve standar hatasını yukarıdan hesaplayabiliriz.

### Evre

$$\overline{d}_{\text{Evre}} = \frac{\sum W_i d_1}{\sum W_i} = 0,467$$

$$s_{\overline{d}} = \frac{1}{\sum W_i} = \frac{1}{0,015} = 8,165$$

### Deneme

$$\overline{d}_{\text{Deneme}} = \frac{0,068}{0,149} = 0,046$$

$$s_{\overline{d}} = \sqrt{\frac{1}{0,149}} = 2,59$$

### Etkileşim

$$\overline{d} = \frac{0,070}{0,147} = 0,476$$

$$s_{\overline{d}} = \sqrt{\frac{1}{0,147}} = 2,608$$

Yukarıdaki sonuçlardan görüldüğüne göre, hiçbir etki için *d* manidar değildir. Kitaptaki örneğin aksine, standart hatalar çok büyüktür.

## BÖLÜM 18

18.1 (a) $W_S = 23$; $W_{,025} = 27$.

(b) $H_0$ ret ve yaşlı öğrencilerin özetlerinde daha çok çıkarım yaptıkları kararına varılır.

18.3 $z = -3,15$; $H_0$'ı ret.

18.5 (a) $T_{,025} = 8$; $H_0$'ı reddetme.

(b) Hipotez oluşturmaya ve test etmeye yeterli kanıtımız olduğu kararına varamayız. (Buradaki durumda, kuyrukların kırılmasında alternatif metotlar kullanıldığında farklı kararlara varılabilir.)

18.7 Ödevin her kümesi için istatistikleri hesaplardım. Puan çiftlerini Önceki ve Sonraki sırası içinde tesadüfî olarak atardım ve atamaların her kümesi için bir istatistik hesaplardım. (Bu istatistik fark puanların ortalaması olur veya fark puanlar üzerinden *t* testi olur.) Daha sonra elde ettiğim aşırı sonuçların sayısını hesaplardım ve bunu tekrar örnekleme sayısına bölerdim, bu değer sıfır hipotezi altında olasılığı verirdi.

18.9 $z = -2,20$ $p(z = \geq \pm 2,20) = ,0278$. Tekrar $H_0$ reddederiz, bu sonuç önceki kararımızla uyumludur.

18.11 Grafik, çiftlerin arasındaki farkların yoğun olarak ilk doğana bağlı olduğunu gösterir.

1813 Wilcoxon eşleştirilmiş-çiftler işaretli-sıralar testi, puan çiftleri eş evrenlerden veya eşit ortalamaya sahip (ve ortancaya) eşdeğer evrenlerden çekildiği yokluk hipotezini test eder. Karşılığı olan *t* testi, eşleştirilmiş puanların dağılımlarının normal olduğunu varsayarak ortalamaları aynı olan evrenlerden çekilip çekilmediğini test eder.

18.15 $H_0$'ın *t* testi ile reddedilmesi uygun parametrik olmayan testi kullanarak reddedilmesinde daha belirgin bir ifadedir, çünkü normallik ve varyansın homojenliği varsayımları evren ortalamasına daha belirgin gönderme yapar.

18.17 $H = 6, 757$; $H_0$ ret.

18.19 *N* deneğin tamamı için veriyi al, bunları karıştırarak tesadüfî sıraya koy. Sonra ilk $n_1$ gözlemi al ve bunları İşlem-1 olarak adlandır, ardındaki $n_2$ gözlemi İşlem -2 olarak ata ve böyle devam et. Daha sonra yeniden örneklenmiş veri üzerinde *F* istatistiğini hesapla ve bunu *F* olarak kaydet. Bunu çok sayıda tekrarla (örn: 1000) ve *F*'nin örneklem dağılımına bak. *F* oranı orijinal veri kümesinden hesaplanan *F*'ye eşit veya ondan büyüktür, bu yokluk hipotezi altındaki olasılığı verir.

18.21 Alıştırma 18.18'deki çalışma, Alıştırma 18.17'ye göre bir avantaja sahiptir, bireysel farklılıkların (genel düzeyde bir kişiden diğerine okuldan kaçma farklılıkları) ortadan kaldırır.

18.23 Bu durumda eşdeğer testlerdir.



# Bulduru


Achenbach, T. M., 66, 76, 187, 194
Adams, H. E., 100, 210-215, 220
adımsal regresyon, 540, 541-542
ağırlıklandırılmış ortalama, 210
Agras, S., 201
Agresti, A., 148, 157, 318-319
Aiken, L. S., 508, 527, 531, 532, 536, 545, 554, 575
aile boyu hata oranı (FW), 371
Akaike Bilgi Kritesi (AIC), 538-539
    ANOVA, 339-341, 361-362
    ki-kare, 148, 149-151
    korelasyon, 277-280
    faktöriyel ANOVA, 451-452
    faktöriyel desen, 424-425
    Huynh and Feldt's epsilon (e), 473
    çekirdek yoğunluk grafikleri, 23
    Kolmogorov-Smirnov testi, 78
    baskı, 535
    karma model söz dizimi, 493-498
    iç-içe desen, 436
    tek-örneklem *t*, 197-198
    güç, 246
    iki bağımsız örneklem, 214-215
    iki eşleştirilmiş örneklem, 206
alan,
    ve olasılık, 64, 71
    altındaki eğri, 68-69, 117
Algina, J., 214, 466
Alleyne, E., 478-483
Allison, D. B., 642
Allison, P. D., 559, 565
Allsopp, D. H., 647
alpha (a), 95
alternatif hipotez, (H1), 91
alternatif görüş, 99-101
    ki-kare, 141-142
    korelasyon, 280-288
    bilinen evren ortalaması, 181-182
    tek örneklem, 183-197
    geleneksek yaklaşım, 89-92
    bağımsız iki örneklem, 206-217
    eşleştirilmiş iki örneklem, 197-206
    binom dağılımını kullanma, 129-131
Altman, D. G., 162, 163
Alvarez, W., 563
American Psychological Association, 33, 41, 88
ana etki, 414
Anastasio, E. J., 606
Anderson, N. H., 606, 607
Andrew, M., 653, 753
Aoki, M. F., 187
Appelbaum, M. I., 575
Appleton, D. R., 176, 738
apsis, 67,253
aracılık ilişkisi, 546-551
aralık ölçeği, 7
aralık/ranj, 38-39
araştırma hipotezi, 88
ardışık değerler, 48
Arnholt, A. T., 9
Aronson, J., 215-217, 231, 239-240, 241-242, 243, 246, 248, 249
Attia, E., 146
ayırma/bölme, 458, 469-471
ayrık değişkenler, 4, 114-115
ayrık, 110

bağıl risk, 161
bağımlı değişkenler, 4
bağımsız değişken, 4
bağımsız olaylar, 110
bağımsız ölçümler, 154
Baisden, J., 374
Baisden, K., 374
Baker, F. B., 471
Baker, S. P., 478
Baron, R. M., 546-550
Barrett, P., 235
Barrett, S., 139, 142, 153, 196
basit etkiler, 414, 420-423, 443, 449-451, 474-478, 482-483
basıklık, 29-30
Bauer, M., 652
Bayer-Sandow, T., 153
Bayes teoremi, 109, 120-124
Bayes istatistiği, 124, 194, 212
Beatty, J., 92
Beaujean, A. A., 551
Becker, S., 367
Behrens-Fisher problem, 217, 219
beklenen değer, 46, 331
beklenen frekanslar, 140
Benjamini and Hochberg's Linear Step Up (LSU) yöntemi, 396-397
Benjamini, Y., 396-397
Bent, D. H., 165
Berger, V. I., 130
Bernoulli deneyi, 124
Best, L. A., 214
beta [b], 95, 265
betimsel istatistik, 5
Bhoj, D. S., 205
Bickel, P. J., 158-160, 164-165
bilgisayar yazılımı, 8-9, 10, 50, 641, 686
*ayrıca bkz:* G*Power, Java applets, Minitab, R, S-PLUS, SAS ve SPSS
binom dağılımı, 124-131
bir ve iki kuyruklu tesler, 97-99, 100-101, 156
birleşik olasılık, 113
birlikte değişkenlik, 594
Bisson, J., 653, 753
Blackburn, A. B., 44, 223, 224-225, 277-281, 298
Blanchard, E. B., 462-466
Bloch, M. H., 651, 752
Block, J. J., 293
Boger, J., 111-112, 113, 144, 149-151, 157-158
Boik, R. J., 475, 492
Bond, L. A., 226-227, 229, 454, 620, 621, 750
Boneau, C. A., 344
Bonferroni yöntemi, 384-387, 391
Borenstein, M., 624, 629, 635, 636, 641
Boring, E. G., 201-203
Bos, H., 456, 745
Bouton, M., 485
Box, G. E. P., 344
Bradley, D. R., 152
Bradley, J. V., 28-29, 35, 344, 658
Bradley, M. T., 214
Bradley, T. D., 152
Brand, A., 214
Brasfield, T. L., 478-483
Braun, J., 21
Brennan, J., 178, 186-187
Brooks, L., 571
Brown, D. A., 462-466
Brown, D. R., 459
Brown, J., 215-217
Brown, M. B., 218, 344, 471
Bruce, P., 105
Bryson, S., 201
bu kitabın Web sitesi, 10
Bullen, C., 638
Bulté, I., 645
Burchard, J., 173
Burke, C. J., 153, 154
Bybee, D. I., 224

Cai, L., 219, 376, 473, 536
Camilli, G., 152
Campbell, I., 152
Carlson, J. E., 589
Carmer, S. G., 392
Carter, J. C., 146
Casey, A., 645, 647, 649
Champely, S., 358
Chen, R., 473
Cillessen, A. H. N., 284
Clark, C. R., 195
Clark, K. B., 169
Clark, M. K., 169
Cochran-Mantel-Haenszel (CMH) statistic, 158
Cochran, W. G., 158, 159
Cochrane, A. L., 253, 255
Coffey, P., 148, 162-164, 167
Cohen, J., 8, 166, 167, 203, 230, 235, 238, 308, 355-356, 357, 508, 527, 531, 532, 536, 545, 574-575, 580,589, 596
Cohen, P., 508, 527, 531, 532, 536, 545, 575
Cohen, S., 58
Cohen's *d*, 203, 213, 238, 272
Cohen's kappa (k), 166-167
Collier, R. O., Jr., 471
Compas, B. E., 2, 91, 223, 257, 551-552, 563, 571, 667-668
Conti, L., 346, 348-350, 357-361, 367, 408, 596-597, 603-607, 613, 617, 743
Cook's D, 534

Cooley, W. W., 544
Cooper, H. M., 628, 629
Corman, M. D., 326-330, 332-339, 346, 354-357, 361, 364, 403
Corrigan, B., 545
Cortina, J. M., 606
Crain, B. R., 293
Cramer, E. M., 575, 606
Cramér, H., 165
Cramér's V, 165
Crawford, J. R., 195, 264
Crawley, M. J., 536
Cribbie, R. A., 352, 370, 387
Crockenberg, S., 531, 546-550
Cronkite, R. C., 554
Crouch, D. J., 367, 505
çubuk grafiği, 65
Culhane, S. E., 430
Cumming, G., 213, 242, 309
Cunnick, J. E., 58
Cutcomb, S. D., 152

çapraz deneysel desen/tasarım, 432-434
çapraz tablolar, 144-148
  analiz, 317-319
çapraz-doğrulama, 542-543
çarpıklık, 27-28, 35
çekileni yerine koyarak örnekleme, 109, 112, 659
çekileni yerine koymadan örnekleme, 659
çekirdek yoğunluk grafikleri, 22-24
çeyreğin konumu, 47-48
çeyrekler arası açıklık, 38-39
çeyrekler, 52
çift modlu, 27-28
çok amaçlı sıfır hipotezi, 329, 371
çok değişkenli normal dağılım, 525
çok değişkenli uç değerler, 532-533
çok değişkenli varyans analizi (MANOVA), 462, 471-172
çok düzeyli modelleme, 492
çok terimli dağılım, 131-132
çok terimli eğilim katsayısı, 404, 697
çoklu karşılaştırma yöntemi, 370-407
  seçme, 398
  bilgisayar çözümleri, 398-401
  faktöriyel desende, 426-427
çoklu karşılaştırmalar, 372, 391-393
çoklu korelasyon katsayısı ($R_{0.123...p}$), 525-527
çoklu regresyon, 508-565
  bütün alt kümeler, 540
  varsayımlar, 524-525
  katsayılar, 508, 515
  tanılar, 532-539
  genel doğrusal model, 574-577
  değişkenlerin önemi, 543-545
  yorumlama, 512-513, 516-519, 543-545
  kısmî korelasyon, 527-531
  yeniden örnekleme, 522-524
  artık değişkenler, 525
  örneklem büyüklükleri, 527
  değişkenlerin seçilmesi, 539-540
  yarı kısmî korelasyon, 529-531
  standart hatalar, 521-522
  baskılayıcı değişken, 531-532

*d*-ailesi ölçümleri, 160-166, 203-204, 272, 313, 356-357, 440-442, 605-606
D'Agostino, R. B., 78
Dabbs, J. M., Jr., 173, 559-560, 738
dağılım diyagramı, 253
dağılım grafiği, 253-255
dağılım, 27-30, 64, 351
  çift-modlu, 27-28
  binom, 124-131
  iki değişkenli-normal, 273-274
  ki-kare, 138-139, 141-142
  koşullu, 268, 274
  *F* istatistiği, 230
  sıklık (frekans), 17-19, 28
  marjinal, 274
  çok terimli, 131-132
  çok değişkenli-normal, 525
  merkezî olmayan *F*, 357
  normal, 28, 64-79
  olasılık, 115-117
  örnekleme, 86-88, 89
  ortalamalar arası farkların örneklem dağılımı, 87, 207
  ortalamanın örneklem dağılımı, 178-181, 232
  standart normal, 69-75
  Student's t, 185
  tek düze/bir örnek, 179
dağılımdan bağımsız testler, 658-659
  *k* ilişkili örneklem için Friedman sıra testi, 679-681
  Kruskal-Wallis tek-yönlü varyans analizi, 678-679
  Mann-Whitney U testi, 673
  sıra-tesadüfîleştirme testleri, 673
  Wilcoxon eşleştirilmiş-çiftler işaretli sıralar testi, 673-676
dağılma, 35-36
dal-ve-yaprak grafikleri, 24-27
dal, 24-25
Daly, F., 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
Darley, J. M., 365
Darlington, R. B., 527, 532, 541, 542, 544, 545, 561
Daston, L., 92

Daumann, J., 367
Davey, G. C. L., 365-366, 409, 743
Davies, K. M., 255
Dawes, R. M., 545
dayanıklı, 218-219
Deeks, J. J., 162, 163
değişkenler, 4, 30, 114-117
değişkenlik, 84, 86
Delaney, H. D., 396, 397, 432, 434, 437, 478
delta ($\delta$), 235-236
Delucchi, K. L., 153
DeMoivre, A., 67
denekler-arası etkiler, 469-470
Denison, D. R., 540
Derogatis, L. R., 227, 324
Devlin, M. J., 146
Diener, E., 115
dik karşılaştırmalar, 383-384
Dingli, D., 652
Dispenzieri, A., 652
Dixon, W. J., 352
doğrusal karşılaştırma(c), 377-378
doğrusal kombinasyon, 378
dönüşümler, 52-54, 346-352
  arksinüs, 350-351, 662
  seçme, 352
  doğrusal, 52-54, 71
  log, 347-350
doğrusal fonksiyonlar, 402-403
doğrusal regresyon modeli, 252
doğrusal ilişki, 257, 290
doğrusal dönüşümler, 52-54, 71
Doob, A. N., 85, 103
Dopfmer, S., 652
double blind study, 146
Dracup, C., 194
Draper, N. R., 263, 522, 540, 541, 542
Drews, F. A., 367, 505
Drudge, O. W., 287
Dugard, P., 288-290
Duncan, O. D., 174-175, 223-224, 225, 227
Dunlap, G., 655, 754
Dunlap, W. P., 214, 355, 473
Dunn testi, 384
Dunn-Šidák testi, 385-386
Dunn, O. J., 384, 385
Dunnett testi, 396
Dunning, D., 222
Dunning, T., 292
Duval, R. D., 665
düzeltilmiş ortalamalar, 595, 600-603, 613-615
Eckelberry, E., 653
Edgington, E. S., 665
Edwards, A. L., 472
Efron, B., 665
eğilim analizi, 401-407
eğim (b), 262-265
eğrisel iklişki, 257, 290
Ellis, P. D., 628
elma grafikleri, 64
en küçük manidar rakam, 24
Epping-Jordan, J. E., 563, 571, 667-668
Erfelder, E., 230
Eron, L. D., 293
eşit olmayan aralıklar, 406-407
eşit puanlar, 672-673
eşleştirilmiş örneklemler, 198, 204-205, 492
eta-kare ($\mu^2$), 354-355
etki büyüklüğü, 101-102, 234-235, 272, 313
  ve nokta-çiftserili katsayı ($r_{pb}$), 308-309
  ki-kare testi, 160-166
  karşılaştırmalar için, 389-391
  kovaryans analizinde, 604-606
  ANOVA'da, 353-357
  faktoriyel ANOVA'da, 437-443
  tekrarlı ölçümler deseninde, 466
  ölçümler, 625-627
  iki bağımsız örneklem, 213-214
  iki eşleştirilmiş örneklem, 203-204
etki büyüklüklerinin heterojenliği, 634-635
etkileşim, 412, 419-420
etkinin büyüklüğü, 354, 375
Ettinger, R. H., 374
Evans, S. H., 606
Everitt, B., 23, 54, 81, 198-201, 203, 222, 225, 227, 266, 342, 361, 621-622, 750
evren varyansı ($\sigma^2$), 40, 46-47
evren, 2
evrensel, 110
Ewen, R. B., 230
Eysenck, M. W., 362-363, 412-423, 426, 431-432, 439, 441-443
*F* istatistiği, 281, 336-338, 692
faktöriyel ANOVA, 412-452
  avantajları, 412, 419
  alternatif yaklaşım, 430-437
  matris çözümü olarak, 580-587
  hesaplama, 416-417
  ortalamalar arasında karşılaştırmalar, 426-427
  etki büyüklüğü, 437-443
  yüksek mertebeden desenler, 446-452
  etkileşimler, 419-420
  yorum, 417-418
  sembolleme, 413-414
  güç analizi, 427-480
  basit etkiler, 420-423
  SPSS yaklaşımı, 424-425
  yapısal modeller, 418-419
  eşit olmaya örneklem büyüklükleri, 444-446
  sonuçların yazılması, 443

faktöriyel desen, 412-418
faktoriyel/çarpansal, 118
faktörler, 412
fark puanları, 200, 617
farklı örneklem büyüklükleri, 380, 394-395
    ANOVA, 341-343, 587-593
    faktörüyel varyans analizi, 444-446
farklılıklara karşı ilişkiler, 252
Farmer, J. L., 647
Faul, F., 230
Federer, W. T., 491
Feinberg, M., 572
Feldt, L. S., 472, 473, 492
Ferron, J. M., 647
fi katsayısı (ϕ), 165, 309-313
Fidalgo, A. M., 176
Field, A. P., 365-366, 409, 743
Fimm, B., 367
Finch, S., 213
Finney, J. W., 554
Fischer, M., 287
Fisher en küçük manidar fark (LSD) yöntemi, 392
Fisher korumalı *t*, 392
Fisher, R. A., 88, 90, 92, 145, 147, 219, 230, 284, 684
Fisher'ın koşullu testi, 147-148, 149-151, 152
Fitzhenry-Coor, I., 681
Fleiss, J. L., 167, 490
Foa, E. B., 364, 372, 409, 502
Forsythe, A. B., 218, 344, 471
Fox, J., 508, 532, 590
Fox, L., 655, 754
Franklin, R. D., 642
Freedenthal, S., 154, 156
Freedman, D. A., 292
frekans dağılımı, 17-19, 28
frekans verisi, 4-5
French, J. M., 176, 738
Frigge, M., 50
Frysinger, R., 681
Fuller, J., 253

G*Power, 230, 243-246, 359-361
Gallagher, M., 681
Galton, F., 68, 296
Games ve Howell prosedürü, 395
Games, P. A., 344, 370, 384, 394, 395, 408
Garber, J., 628-631, 633
Garland, C. F., 255
Garland, F. C., 255
Garthwaite, P. H., 195, 264
Gartrell, N., 456, 745
Gauss, C. F., 68
Geffen, G., 195
Geffen, L. B., 195
Geisser, S., 472, 473, 475, 492
genel doğrusal model, 574-577
gerçek alt sınır, 19-20
geriye dönük çalışma, 160
geriye dönük güç, 245-246
Gertz, M. A., 652
Geyer, C. J., 296
Giancola, P. R., 326-330, 332-339, 346, 354-357, 361, 364, 403
Gibson, L., 173, 310-313
Gigerenzer, G., 92
Gilpin, A., 44
Gladen, B. C., 169
Glass, G. V., 203, 308, 311, 440-441
Gleser, L. J., 643
Gold, P. E., 409, 676-677
Goldberg, L. R., 546
Good, C., 215-217
Good, P. I., 147, 194, 665
Gordon, L. R., 148
Gorham, E. D., 255
Gorman, B. S., 642
Gosset, W. S., 185
Gouzoulis-Mayfrank, E., 367
gözlenen frekanslar, 140
grafiği sürekli hale getiriciler, 266
Grant, G., 169
Grant, K., 223
Greenhouse, S. W., 472, 473, 475, 492
Greipp, P. R., 652
Grice, J., 235
Grissom, R. J., 344, 347
Gross, A. E., 85, 103
Gross, J. S., 170
Guber, D. L., 297, 509-510, 512, 572
güç, 97, 230-247, 698
    ve faktoöriyel desen, 427-430
    ve manidarlık, 242-243
    *a*'nın fonksiyonu olarak, 233
    $H_1$'in fonksiyonu olarak, 233
    *n* ve $\sigma^2$'nin fonksiyonu olarak, 233-234
    hesaplama, 234-236, 243, 293-294
    etkileyen faktörler, 232-234
    ANOVA'da, 357-361
    eşleştirilmiş-örneklem *t*, 241-242
    tek-örneklem *t*, 236-238
    iki bağımsız ortalama, 238-240
    G*Power kullanılarak, 243-245
    sonuçların yazılması, 247
güven aralıkları, 192-195, 201, 388-389
güven sınırları, 192, 212-214, 274-277, 286-287, 309
güvenilir (credible) aralık, 194

Haenszel, W., 158

Hall, J. L., 676-677
Hammel, E. A., 158-160, 164-165
Hand, D. J., 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
Harris, C. W., 617
Harris, F. H., 44, 223, 224-225, 277-281, 298
Harris, R. J., 100, 475, 545, 580
Hastalık Kontrolü İçin Merkezler, 116
hata oranları, 370-371
hata varyansı, 267-268, 328
hata, 68
$hata_{arasında}$, 470
Hatada yüzde indirim (PRE), 355
$hata_{içinde}$, 470
Hayes, A. F., 219, 376, 473, 536, 550, 547, 551
Hayes, T. F., 471
Hays, W. L., 316, 321, 462
Hazelton, J., 681
Heaney, R. P., 255
Hedges, L. V., 213, 308, 624, 629, 632, 633, 635, 636, 641
Hedges'g, 213, 238
Heisey, D. M., 246
Henderson, D. A., 540
Henning, K., 148, 162-164, 167
hesaplama formülü, 42-43
heterojen alt örneklemler, 292-293
Higgins, J. P. T., 624, 629, 635, 636, 641
hinges, 48
hipotez testi, 84-104, 220-221
histogram, 18-22, 65
Hoaglin, D. C., 50, 223, 347, 352, 664-666
Hochberg, Y., 370, 396-397
Hodgins, H. S., 174
Hoenig, J. M., 246
Hoffman, J. M., 547, 550, 551
Holick, M. F., 255
Holland, B., 352, 370, 387
Holmes, T. H., 6
Holway, A. H., 201-203
Hopkins, K. D., 152
Horowitz, J., 628-631, 633
Horowitz, M. J., 563
Hosmer, D. W., 561, 562, 563, 565
Hotelling, H., 287
Hothorn, T., 23
Hout, M., 174-175, 223-224, 225, 227
Howard, B. L., 654, 754
Howell, C. T., 187
Howell, D. C., 2, 52, 55, 91, 148, 169, 178, 186-187, 195, 223, 257, 264, 341, 365, 376, 396, 432, 454, 491, 543, 551-552, 563, 571, 575, 590, 620, 667-668, 750
Howell, J. F., 370, 384, 394, 395, 408
Hraba, J., 169
Huberty, C. J., 522, 540, 545
hücre, 144, 414
Huesmann, L. R., 293
Huessy, H. R., 55, 169, 365
Huitema, B. E., 602, 606, 607, 617, 645, 647, 649
Hull, C. H., 165
Hunter, J. E., 101
Huynh ve Feldt epsilonu, 473
Huynh, H., 472, 473, 492

iç çitler, 48
iç-içe desen, 431, 434-437
Iglewicz, B., 50
iki değişkenli normal dağılım, 252, 273-274
iki tepedeğerli (bkz: İki modlu) 27, 28
iki kategorili, 305
iki-kuyruklu test, tanımlı, 98
iki-yönlü faktöriyel desen, 412
ikinci dereceden fonksiyonlar, 402
ileriye dönük çalışma, 160
ilişkilere karşı farklılıklar, 252
ilişkili örneklm, 198
Introini-Collison, I., 366, 408, 620
işaret testi, 130-131, 667-668
istatistik, 5
izleyen basamaklar, 27
ılımlı ilişkiler, 551-555
Jaccard, J., 554
Jankowski, M. K., 148, 162-164, 167
Java uygulaması, 71, 263
Jefferson, K. W., 478-483
Jenkins, J. G., 165
Jo, B., 201
Joiner, B. L., 292
Jones, L.V., 92, 99-102, 106, 142, 182, 203, 207, 211, 220-221, 247
Jose, P., 551, 571
Judd, C. M., 272, 329, 355, 430, 432, 555, 575, 586
Juieng, D., 86-88, 93-94, 102
June, L., 423-429, 440, 608-611

*k* ilişkili örneklem için Friedman sıra testi, 679-681
Kahneman, C., 122-123
kalan varyans, 267-268, 516, 524
kalan/artık, 262, 517
Kane, R. T., 651, 752
Kaplan, A. S., 146
Kaplan, J. R., 58
Kapoor, P., 652
Kapp, B., 681
kappa (κ), 166-167
karar verme, 94-97
kareler ortalaması, 335-336
kareler toplamı(SS), 268-269, 332-335, 378-379, 587-590

karşılaştırmalar, 376
kategorik/sınıflama verisi, 4-5
Katz, S., 44, 223, 224-225, 277-281, 298
Kaufman, J., 191
Kaufman, L., 191-192, 201-202
kayıp veri, 205-206, 341, 543
kazanılan puanlar, 200
Kazdin, A. E., 125
Keen, K. J., 20
Kelley, K., 357
Kendall konkordans katsayısı (*W*), 320-321
Kendall, P. C., 654, 754
Kendall's *t*, 315-317
Kenny, D. A., 329, 546-550
Keough, K., 215-217
Keppel, G., 406
Keselman, H. J., 214, 344, 352, 370, 387, 394, 478
Keselman, J. C., 478
keşfedici veri analizi (EDA), 5, 24, 52, 187
kesim (*a*), 262-264
kestirici, 253
kestirimin hatası, 262, 269
kestirme aralığı, 75, 274
kestirme, 252-253, 266-272
kestirmenin standart hatası, 267-269
ki-kare ($\chi^2$) dağılımı, 138-139, 141-142
  tablolanmış, 141-142
ki-kare ($\chi^2$) testi, 138, 139-141
  varsayımlar, 152-154
  hesaplaması, 145
  bilgisayar analizleri, 149-151
  çapraz tablolar, 144-152
  etki büyüklükleri, 160-166
  değerlendirme, 145-147
  Fisher Tam (Exact) Testi, 147-148
  formül, 140
  uyum iyiliği testi, 139-141
  çoklu kategori hali 142-145
  sembolleme, 138
  sıralama verisi, 152
  problemler, 142, 156
  tekrarlı ölçümler, 154-156
  beklenen küçük frekanslar, 151-152
  sonuçların yazılması, 167-168
  süreklilik için Yates düzeltmesi, 147, 159
King, D. A., 467-470, 476, 622
Kirk, R. E., 481
kırpılmış örneklemler, 35, 351-352
kırpılmış ortalamalar, 35, 39, 387
Klein, M., 565
Kleinbaum, D. G., 565
Klemchuk, H. P., 454, 620, 750
Kline, R. B., 203, 204, 308, 389, 440, 466
$KO'_{\text{hata}}$, 602
Koele, P., 357, 360
$KO_{\text{grup içi}}$, 330
$KO_{\text{hata}}$, 330
Kohr, R. L., 344
$KO_{\text{işlem}}$, 330
Kolmogorov-Smirnov testi, 78
kombinasyonlar, 117-120
kombinetorikler, 118
korelasyon, 252-294
  düzeltilmiş ($r_{\text{düzeltilmiş}}$), 261
  varsayımlar, 272-274, 290
  nedensellik, 265
  katsayılar, 260-261, 265, 304-309, 313-317, 525-527
  bilgisayar çözümü, 277-280
  kovaryans ($\text{cov}_{XY}$ veya $S_{XY}$), 258-260
  etki büyüklüğü, 308-309, 627
  etkileyen faktörler, 291-293
  heterojen alt örneklemler, 292-293
  korelasyonlar arası matris, 543-544
  çoklu ($R_{0.123\ldots p}$), 525-527
  doğrusal olmama (nonlinearity), 290
  Pearson çarpım moment korelasyon katsayısı (*r*), 260-261
  fi($\phi$), 309-313
  nokta çift serili ($r_{pb}$), 304-309
  evren (*p*), 260-261
  güç, 293-294
  ranj sınırlamaları, 291-292
  sıralanmış veri, 313-317
  regresyon doğrusu, 253-254, 261-266
  regresyon ilişkisi, 265
  grafikler, 253-254
  manidarlık testi, 280-288
  sıralanmış veri için Spearman korelasyon
  koerelasyon katsayısı, 314-315
  standartlaştırılmış regresyon (β), 265
  tabloları, 314
korelasyonel ölçümler, 304
koşulsuz olasılık, 114
kovaryans analizi, 594-616
  varsayımlar, 595
  faktöriyel desenler,
  hesaplamalar, 595-604
  sonuçların yazılması, 607
  yorumlama, 606-607
  etki büyüklükleri, 604-606
kovaryans matrisi (Σ), 461-462, 520
kritik değer, 94
Kruger, J., 222
Krüger, L., 92
Kruskal-Wallis tek yönlü varyans analizi, 678-679
$KT_{\text{denekler arası}}$, 459
$KT_{\text{denekler arası}}$, 459
$KT_{\text{hata}}$, 334
$KT_{\text{hücreler}}$, 416

$KT_{işlem}$, 334
$KT_{toplam}$, 334
Kuang, D., 397
Kumar, S., 652
Kunert, H.-J., 367
kutu grafikleri, 47-51
Kyle, R. A., 652

Lacy, M. Q., 652
lambda ($\lambda$), 359
Lancaster, T., 638
Landeros-Weisenberger, A., 651, 752
Landis, J. R., 167
Landwehr, J. M., 12-13
Lane, D. M., 214, 355
Langlois, J. H., 36
Laplace, P.-S., 68
Lappe, J. M., 255
Latané, B., 365
Lautenschlager, G. J., 44, 223, 224-225, 277-281, 298
Le Grange, D., 201
Leckman, J. F., 651, 752
Lecoutre, B., 473
Leerkes, E., 531, 546-550
Lefering, R., 153
Lehman, E. L., 92
Leitenberg, H., 148, 162-164, 167, 173, 310-313
Lemeshow, S., 561, 562, 563, 565
Lenth, R. V., 235
Leroy, A. M., 262, 508
Levene, H., 218, 344, 471
Levine, R., 255-257
Lew, R., 478
Lewis, D., 153, 154
limit, 109
Lipkin, M., 255
Little, R. J. A., 491
Lock, J., 201
Lockwood, C. M., 547, 550, 551
Loew, D. E., 287
Loftus, G. R., 194
log dönüşümleri, 347-350
Logdon, R. G., 368, 506
lojistik regresyon, 556-565
Lohnes, P. R., 544
Lohr, B. A., 100, 210-215, 220
Lord, F. M., 8, 606, 607
Lunn, A. D., 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
Lunneborg, C. E., 561, 665
Lustina, M. J., 215-217
Macauley, C., 660-664
MacDonald, C. B., 365-366, 409, 743
MacKinnon, D. P., 547, 550, 551
Mahon, J., 152, 156, 317, 319
Maindonald, J., 21
Malcarne, V. L., 223, 571
Mandeville, G. K., 471, 472
Mandrekar, S., 652
manidarlık düzeyi,
  korelasyon, 280-288
  doğrusal karşılaştırmalarda, 380-381
  genel *F*, 372-373
  nokta-çift serili korelasyon katsayısının ($r_{pb}$), 308
  Spearman'ın ro'su, 314
  iki bağımsız *r*, 284-285
  iki bağımsız eğim, 282-284
  iki bağımsız olmayan *r*, 287-288
manidarlık düzeyi, 94
Manly, B. F. J., 665
Mann-Jones, J. M., 374
Mann-Whitney U test, 673
Manning, C. A., 676-677
Manolov, R., 645
Mant, D., 638
Mantel-Haenszel istatistiği, 158-160, 165
Mantel, N., 158
Manuck, S. B., 58
Marascuilo, L. A., 156
Maris, E., 607
marjinal dağılım, 274
marjinal toplamlar, 144
Masson, M. E. J., 194
Mauchley test, 472-473
Mauchly, J. W., 464
Maxwell, A. E., 318
Maxwell, S. E., 396-397, 432, 434, 437, 478, 606
Mazzucchelli, T., 651, 752
McCabe, G. P., 298
McClelland, G. H., 234, 240, 272, 355, 361, 420, 430, 432, 555, 575, 586
McConaughy, S. H., 575, 590, 680
McConway, K. J., 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
McCurry, S. M., 368, 506
McGaugh, J. L., 366, 408, 620
McGrath, R. E., 203
McGrath, S. G., 152
McGraw, B., 440-441
McIntyre, S. H., 541
McKean, J. W., 645, 647
McNemar, Q., 155, 171, 528
McNemar's test, 155
Meijs, N., 284
Merikle, P., 125
Merkez, B., 645, 647, 649
merkeze eğilim, 32-35
merkeze eğilme ölçüleri, 32-35
merkezî limit teoremi, 178-181

merkezî olmayan *F* dağılımı, 357
meta-analizi, 634-641
  etki büyüklükleri, 625-628, 634-635
  dosya çekmecesi problemi, 624
  sabit model, 630-632
  orman grafiği, 630-632
  tesadüfî model, 635-636
  alt grup farklılıkları, 635-636
Meyer, G. J., 203
Michels, K. M., 459
Mikhael, J. R., 652
Miller, D. T., 249
Minitab, 9, 292
Mireault, G. C., 173-174, 223, 226-227, 229, 570, 621
Mitchell, R. E., 554
Moar, K. J., 195
Mohr, S. B., 255
Moise-Titus, J., 293
Montgomery, D. B., 541
Mooney, C. Z., 665
Moore, D. S., 298
Moore, F., 253, 255
Moos, R. H., 554
Morris, R., 173, 559-560, 738
Mosteller, F., 223, 347, 352, 664-666
Moye, A., 201
muhtemel sınırlar, 74
Murdock, T. B., 364, 372, 409, 502
Murphy, K. R., 230
Musty, R. E., 346, 348-350, 357-361, 367, 408, 596-597, 603-607, 613, 617, 743
Myers, J. L., 478
Myors, B., 230

neden/sebep olma, 265
Nee, J. C. M., 167
Neugebauer, F. A. M., 153
Newman-Keuls test, 395, 398
Newmark, H., 255
Neyman, J., 88, 91, 92
nicel veri, 4
Nickerson, R. S., 88
Nie, N. H., 165
Nisbett, R. E., 744
nitel veri 4-5
nokta-çiftserili korelasyoın katsayısı ($r_{pb}$), 305-309
Nolen-Hoeksema, S., 451, 456
nominal ölçekler, 6
normal dağılım, 28, 64-79
  tanım, 68
  önemi, 66
  standart, 69-75
  tablosu, 69-72, 714
  tablonun kullanılması, 71-73
Nouri, H., 606
Nurcombe, B., 178, 186-187, 681
O'Bannon, R. E., III., 478-483
O'Brien, R. G., 218, 344, 471
O'Connell, J. W., 158-160, 164-165
O'Hanlon, A. P., 195
O'Neil, R., 370
Oakes, M., 92
olasılığa analitik bakış
olasılığın çarpım kuralı, 111-112
olasılığın frekansçı görünümü, 109
olasılığın toplanması kuralı, 111
olasılık dağılımları,
  sürekli değişkenleri için, 115-117
  sürekli değişkenler için, 115
olasılık oranı, 161-165
olasılık, 161-162
olasılık, 64, 70-71, 108-132
  toplama kuralı, 111
  olasılık hakkında analitik görüş, 108-109
  koşullu olasılık, 113-114
  tanım, 108-110
  frekansçı görüş, 109
  birleşik olasılık, 113
  kanunlar, 110-114
  çarpım kanunu, 111-112
  sonraki olasılık, 121-122
  önsel olasılık, 121-122
  subjektif olasılık, 109, 120-121
  kosulsuz olasılık, 114
olay, 110
Olejnik, S., 466
Olkin, I., 633, 643
Olmsted, M., 146
omega-kare($\omega^2$), 355-356
Onghena, P., 645
orta noktalar, 20
ortalama ($\bar{X}$ or λ), 33-35
  aritmetik, 33
  harmonik($\bar{X}_h$), 239
  kırpılmış, 35, 39, 387
ortalamadan mutlak sapma, 40
ortanca(Ort), 32-33, 34, 47
Osbourne, J. W., 352
Ostrowski, E., 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
Overall, J. E., 152, 498, 574, 575, 586, 588, 589
Owens, C. M., 647
oran ölçeği, 7
ordinat, 67, 253

*Öğrenci Çözüm El Kitabı*, 10, 444
ölçme ölçeği, 6-8
ölçme verisi, 4
ölçüt, 253

önce gelen rakam, 24
önsel olasılık, 121-122
önsel, 372, 376,388
örneklem büyüklükleri,
  etkili, 240
  eşit, 238-239
  kestirim gerekli, 236-237
  eşit olmayan, 239-240, 380, 394-395
örneklem dağılımı, 86-88, 89
  ortalamalar arası farklar, 87, 207
  ortalamanın, 178-181, 232
  örneklem varyansı, 183-184
örneklem hatası, 84, 85
örneklem istatistiği, 93
örneklem varyansı (s2), 40-41, 46-47
örneklem, 2-3
örnekler,
  alkol ve dikkat dağılma (distracting) görevi, 326-330, 332-339, 346, 354-357, 361, 364, 403
  indirimler, 12, 105
  iştahsizlik (anorexia), 54, 81, 198-201, 203, 222, 225, 227, 342, 361, 621-622, 750
  aspirin, 160-162, 165-166, 167
  kaçınma, 563, 667-668
  Berkeley cinsiyet ayrımcılığı, 158-160, 164-165
  beta-endorphin, 664-666
  cancer survival, 556-565
  çocukluk depresyonu, 628-638
  başa çıkma davranışı, 181-183
  ölüm cezası, 111, 113, 144, 149-151, 157-158, 161
  sürüş davranışı, 367, 505
  sürüş davranışı ve durum, 85, 103
  THC'nin etkisi, 346, 348-350, 357-361, 367, 408, 596-607, 613, 617, 743
  Eysenck'in hatırlama çalışması, 362-363, 412-423, 426, 431-432, 439, 441-443
  HIV enfeksiyon riski, 478-483
  homofobi, 210-215, 220
  doğum kilosu düşük bebekler, 178, 186-187
  M&M's, 108, 112
  annelik bakımı, 531, 546-551
  zihinsel durum, 660, 661, 663-664
  migren, 462-466
  ay yanılsaması, 191-192, 201-202
  morfin toleransı, 373-375, 380-382, 388-391, 397, 467
  yaşamın hızı, 255-257
  gezegenler, 288-290
  tecavüz terapisi, 372, 409
  kaya/kâğıt/makas, 143, 153
  SAT ve ACT, 509-531, 532-533, 535-539, 572
  cinsel istismar, 148, 162-164, 167-168
  cinsel tatmin, 174-175, 223-224, 225, 227
  sigara içme ve başarı/performans, 423-429, 440, 608-611
  sigara bırakma, 638-641
  futbol antrenmanı, 646-650
  klişe (stereotype) tehditler, 215-217, 231, 239-240, 241-242, 243, 246, 248, 249
  yüzme takımları, 451-452, 456
  test-alma becerileri, 277-281, 298
özyükleme, 11, 190-191, 658, 659-668
  güven aralığı sınırları, 662-665
  aracılık ilişkilerinde, 550-551
  bağımsız örneklemler, 667-668
  tek örneklem, 661-662
  eşleştirilmiş çiftler, 665-666

Panza, K. E., 651, 752
parametre, 5
parametrik olmayan testler, 658-659
parametrik testler, 658-659
parçalı regresyon, 646-650
Parides, M., 146
Pearson ki-karesi, 138, 148
Pearson momentler çarpımı korelasyon katsayısı (*r*), 260-261
Pearson, E. S., 88, 91, 92, 152
Pearson, K., 138, 145
Péladeau, N., 554
Pelz, S., 367
Penfield, R. D., 214
Perera, R., 638
Peres, B., 12
permütasyon testleri, 119-150, 659
permütasyonlar, 118-119
Perot, A. R., 571
Peterson, W. P., 144
Piazzi, G., 301
Pike, K. M., 146
Podolski, C., 293
Pollin, W., 225-226
Porter, T., 92
Preacher, K. J., 550, 551
Prentice-Dunn, S., 59
Prentice, D. A., 249
Pugh, M. D., 171

Q istatistiği, 634-635
Q-Q grafiği, 75-78
Quetelet, A., 68

R (bilgisayar dili), 9, 196, 266, 665
*r*-ailesi, 160, 165-166, 272, 604-605
Rabin, B. S., 58
Raghunathan, T. E., 287
Ragozzino, M. E., 409
Rahe, R. H., 6
Rajkumar, S. V., 652, 753
Rakow, T., 142
Ramsey, F. L., 175

ranj daralması, 291-292
Rasmussen, J. L., 659
Rauh, V. A., 178, 186-187
Rees, C. S., 651, 752
regresyon katsayıları, 263, 281-282, 283, 508, 515, 545-546
  standartlaştırılmış ($\beta$, beta), 265, 520-521
  regresyon doğruları, 253-254, 261-266
regresyon, 252-260, 261-290
  varsayımlar, 272-274, 290
  katsayılar, 263, 265, 281-282, 283
  bilgisayar örneği, 277-280
  hipotez testi, 280-288
  lojitik, 556-565
  parçalı, 646-650
  standartlaşrıeılmış veri, 263, 265
  adımsal, 540, 541-542
  karşı korelasyon, 252
Reichardt, C. S., 606
Reilly, T. P., 287
Renner, J., 423-429, 440, 608-611
reddetme bölgesi, 94
reddetme düzeyi, 94
Reynolds, C. R., 223
Richmond, B. O., 223
Riggs, D. S., 364, 372, 409, 502
risk farkı, 161
risk oranı, 161, 162-163
risk, 161
Roberto, C. A., 146
Robson, D. S., 407
Rock, I., 191-192, 201-202
Rockert, W., 146
Rogan, J. C., 344, 394
Rogers, R. W., 59
Roggman, L. A., 36
Rosa, E., 139, 142, 153, 196
Rosa, L., 139, 142, 153, 196
Rosen, J. C., 287
Rosenthal, R., 160, 174, 214, 271, 287, 311-313, 323-324, 420, 449, 624
Rosnow, R. L., 420
Rothbaum, B. O., 364, 372, 409, 502
Rothstein, H. R., 624, 629, 635, 636, 641
Rousseeuw, P. J., , 508
Roy, V., 652
Ruback, R. B., 86-88, 93-94, 102
Rubin, D. B., 287, 311-313, 323-324, 449, 491
Rudd, R. J., 409
Ruoff, P., 178, 186-187
Ryan yöntemi (REGWQ test), 395, 400
Ryan, B. F., 292
Ryan, T. A., 292, 370
S-PLUS, 9, 196
Sackett, D. L., 162, 163
sağa doğru çarpık, 20, 27-28
sakal, 49
sapma ölçüleri (değişkenlik), 35-47, 51
sapma puanları, 53, 70
Sarner, L., 139, 142, 153, 196
şartlı olasılık, 113-114
SAS, 9
  Huynh and Feldt epsilon, 473
  iç-içe desen, 436
  örnek çıktı, 398
Sasa, H., 367
SAT, 44-45, 510-511
Satterthwaite, F. E., 217, 219, 477
Sauerland, S., 153
scedasticity, 328-329
Schaefer, M., 173
Schafer, D. W., 175
Scheffé test, 396, 400
Scholte, R. H. J., 284
*sd*', 218
$sd_{\text{hata}}$, 335
$sd_{\text{işlem}}$, 335
$sd_{\text{toplam}}$, 335
Segers, E., 284
Seligman, M. E. P., 170, 451, 456
sembolleme, 30-32, 33, 254, 413-414, 575
serbestlik derecesi(*sd*), 46
  ki-kare, 138, 141, 145
  tekrarlı-ölçümler deseninde, 472-473
  tek örneklem *t*, 188
  tek-yönlü ANOVA, 335, 338
  bağımsız iki örneklem, 210
  eşleştirilmiş iki örneklem, 201
Serlin, R. C., 156, 344
Sgro, J. A., 341
Sheets, V., 547, 550, 551
Shirley, A., 478-483
Shivakumar, C., 498
Shrout, P. E., 490
Šidák, Z., 385
Siegel, S., 373-375, 380-382, 388-391, 397, 467
sigma ($\Sigma$), 30-31
sigma ($\sigma$), 40
sigmoidal curve, 557
Silver, B. V., 462-466
Simon, J. L., 105
Simpson, E. H.., 158
Simpson's paradoksu, 158
sıfır hipotezi ($H_0$), 88-92, 93-94, 99, 371
sıralama ölçeği, 6-7
sıralama, 313-314
sıralı veri için Spearman korelasyon katsayısı ($r_S$), 314

Skiba, R., 645, 647, 649
Smith, H. F., 606
Smith, H., 263, 522, 540, 541, 542
Smith, M. L., 311, 440-441
Sobel, M. E., 174-175, 223-224, 225, 227, 548-550
sola doğru çarpık, 27-28
Solanas, A., 645
sonraki olasılık 121-122
sonuçların yazılması,
  kovaryans analizi, 607
  ANOVA, 338-339
  Benforreni yöntemleri, 391
  ki-kare, 167-168
  bağımlı *t* testi, 206
  faktöriyel ANOVA, 443
  güç, 247
  tekrarlı ölçümler deseni, 466
  *t* test, 214, 216-217
Soper, D., 385
Spiegel, D. K., 574, 586, 588, 589
Spijkerman, R., 284
Spilich, G. J., 423-429, 440, 608-611
SPSS, 9, 10, 51
Srinivasan, V., 541
St. Lawrence, J. S., 478-483
St. Leger, A. S., 253, 255
Stabenau, J. B., 225-226
standart hata, 86, 182, 208, 521-522
standart normal dağılım, 69-75, 714
standart puanlar (z), 53-54, 78-79
standartlaştırma, 53-54
stansart sapma (s, s, or SD), 40-44
  hata ölçüsü olarak, 266-267
Startup, H. M., 365-366, 409, 743
Stead, L. F., 638
Steele, C. M., 215-217
Steiger, J. H., 287, 356, 357
Steinberg, L., 397
Steinbrenner, K., 165
Stephens, M. A., 78
Sterling, T. D., 214
Sternberg, S., 16, 58
Stevens, J., 508
Stevens, S. S., 6
Stigler, S. M., 297
Stoica, G., 214
Stone, W. S., 409
Strayer, D. L., 367, 505
Student *t* dağılımı, 185
Student, 185
Studentlaştırılmış Ranj İstatistiği (q), 392-393
subjektif olasılık, 109, 120-121
Sullivan, C. M., 224
sürekli değişkenler, 4, 114-117
Swanson, M. R., 392
Swartzentruber, D., 485
Swijtink, Z., 92

*T* (Wilcoxon istatistiği), 674-675, 703
*t* istatistiği, 93, 184-186, 188-189, 219, 702
  bağımsız örneklemler, 208-209
  çoklu karşılaştırmalar, 376-377
  ilişkili örneklemler, 200
  sonuçların yazılması, 206
*T* puanları, 79
Tamhane, A. C., 370
tanımsal formül, 42-43
tasarım matrisi, 575-577
tek durum desenleri, 641-650
  A-B desenleri, 641-642
  otokorelasyon, 645
  parçalı korelasyon, 646-650
tek-yönlü ANOVA, 326-341
  matris-çözümü olarak, 577-580
  varsayımlar, 328-329
  hesaplamalar, 332-337
  bilgisayar çözümleri, 339-341
  serbestlik derecesi, 335
  *F* istatistiği, 336-337
  lojik, 329-331
  modeller, 328-329
  sıfır hipotezi, 328
  değişkenlik kaynakları, 335
  özet tablo, 335
  sonuçların yazılması, 338-339
tekrarlı-ölçümler deseni, 412, 458-498
  varsayımlar, 461-462
  Karşılaştırmalar (contrasts), 465-466
  *F* oranları, 460-461
  karma modeller, 492-498
  basit etkiler, 482-483
  yapısal model (structural model), 460
  sonuçların yazılması, 466
tepe (Mod), 32, 34
Teri, L., 368, 506
tesadüfî atama, 3-4
tesadüfî örneklem, 2-3
tesadüfîleştirme testleri, 119, 658
test istatistiği, 93
Teti, D. M., 178, 186-187
Theobald, D. E., 462-466
Thissen, D., 397
Thompson, B., 235
Thornton, K. M., 451, 456
Thornton, N., 451, 456
Tibshirani, R., 665
Timm, N. H., 589
Todman, J., 288-290
tolerans, 520
Tomarken, A. J., 344

Tonidandel, S., 498
Toothaker, L., 370
transfer etkileri, 491
Travers-Gustafson, D., 255
Tuchtenhagen, F., 367
Tukey-Kramer yaklaşımı, 394-395
Tukey, J. W., 5, 24, 47, 48, 52, 92, 99-102, 106, 142, 182, 203, 207, 211, 220-221, 223, 347, 352, 370, 371, 664-666
Tukey's HSD (saf manidar fark) testi, 393-394, 400
Turrisi, R., 554
Tversky, A., 122-123
Tip I hata, 94-97, 101, 103, 230, 232, 370
Tip I kareler toplamı, 589-590
Tip II hata, 92, 94-97, 103, 230
Tip II kareler toplamı, 588
Tip III kareler toplamı, 588

uç değerler, 21, 48, 49, 532-533
uç puanlar, 34, 43-44
Unah, I., 111-112, 113, 144, 149-151, 157-158
Uomoto, J., 368, 506
uyum-iyiliği testi, 140
Valentine, J., 629
Van den Noortgate, W., 645
Vanderpump, M. P. J., 176, 738
varyans analizi, (ANOVA), 326-362
  hesaplamalar, 332-337
  matris çözümü olarak, 576-580
  varsayımlar, 328-329
  bilgisayar çözümleri, 339-341, 359-361, 361-362
  serbestlik derecesi, 335
  $sd_{hata}$, 335
  $sd_{işlem}$, 335
  $sd_{toplam}$, 335
  etki büyüklükleri, 353-357
  F istatistiği, 336-337
  lojik, 329-331
  modeller, 328-329, 353
  sıfır hipotezi, 328
  tek-yönlü, 326-341
  güç, 357-361
  varyans kaynakları, 335
  KThata, 334
  KTtoplam, 334
  KTişlem, 334,
  özet tablo, 335
  dönüşümler, 346-352
  eşit olmayan örneklem büyüklükleri, 341-343
  varysayımların ihlâli, 343-345
  sonuçların yazılması, 338-339
varyans bükülme faktörü (VIF), 520
varyans, 40-41, 42, 52
  hata, 267-268, 328
  kestirme, 331
  uç değerlerin etkisi, 43-44
  birleştirme, 209-210
  artık/kalan, 267-268, 516, 524
  örneklem, 40-41, 46-47
varyansların heterojenliği, 217-220, 473, 329, 344-345
varyansın homojenliği, 217, 273, 328, 472-473
varyansları birleştirme, 209-210
varyansları toplama kanunu, 207
Venables, W. N., 219-220
Venneri, A., 264
veri grafiği, 16-18, 20
Viechtbauer, W., 630
Visintainer, M. A., 170
Volpicelli, J. R., 170

Wagner, B. M., 2, 91, 257, 551-552
Wainer, H. , 545
Wald ki-kare, 562-564
Walsh, T. B., 146
Wan, C. K., 554
Waner, S., 121
Watkins, A. E. , 135
Watkins, A. E., 12-13
Weaver, K. A., 36
Weinberg, C. R., 169
Weinstock, S., 341
Weisberg, H. I., 606
Weitz, B. A., 541
Welch yöntemi, 344-345
Welch,
  B.L., 217, 218, 219, 344, 345, 477
Welch-Satterthwaite yöntemi, 217-218, 219, 344-345, 473
Welkowitz, J., 230
Werner, M., 225-226
West, S. G., 508, 527, 531, 532, 536, 545, 547, 550, 551, 554, 575
Wetherill, G. B., 370
Wilcox, R. R., 35, 205-206, 219, 344, 352, 373, 387
Wilcoxon eşleştirilmiş çiftler işaret testi, 673-676
Wilcoxon sıra-toplamları testi, 668-673
Wilkinson, L., 21, 88
Willer, R., 572
Williams, E. J., 287
Williamson, D. A., 462-466
Williamson, J. A., 181-183
Wilner, N., 563
Winer, B. J., 10, 451, 459, 464, 481, 602, 615
Winsorized örneklemler, 39, 351-352
Woodside, B., 146
Worsham, N. S., 223
Wright, L. W., Jr., 100, 210-215, 220
$W_S$ (Wilcoxon istatistiği), 669-672, 709
WSD (tamamen manidar fark) test, 393
Wuensch, K., 29

Yanlış Buluş Oranı (YBO), 396
yansız kestiriciler, 45-46
yapraklar, 24
yeniden örnekleme, 10-11, 190-191, 352, 658
    ANOVA'da, 357-361
    çoklu regresyonda, 522-524
yoğunluk (f(x)), 67, 117
yönlü test, 98
yönsüz testler, 98, 156
yöntem, 27-28
yoplama kuralları, 30-31
yordamsal istatistik, 5
Yuen, K. K., 352
Yule, U., 158
yuvarlama hatası, 449
yuvarlama metodu, 20
yüzdelikler, 51-52, 78

z puanları, 69, 70-71, 78-79
Zara, A., 365-366, 409, 743
Ziegler, S. G., 646-648
Zimmerman, D. W., 6
Zuckerman, A., 174
Zuckerman, M., 174
Zumbo, B. D., 6



# Semboller Sözlüğü

## Grek Harfi Semboller

| | |
|---|---|
| $a$ | Alfa: manidarlık düzeyi, I. Tip hata olasılığı |
| $\alpha_i$ | *A*'nın *i*'yinci düzeyi için işlem etkisi |
| $\beta$ | Beta: Tip I hata olasılığı |
| $\beta i$ | Standartlaştırılmış regresyon katsayısı; B'nin *i*'nci düzeyi için işlem etkisi |
| $\Gamma$ | Gama fonksiyonu (faktöriyel ile ilişkili) |
| $\delta$ | Delta: merkezî olmama parametresi, etki büyüklüğü |
| $\hat{\varepsilon}, \tilde{\varepsilon}$ | Epsilon: tekrarlı-ölçümler varyans analizinin serbestlik derecesini düzeltme faktörleri için semboller |
| $\eta^2$ | Eta kare: korelasyon oranının karesi |
| $\mu, \mu_X$ | Müyu (mu): evren ortalaması |
| $\mu_{\overline{X}}$ | Örneklem ortalamalarının dağılımının ortalaması |
| $\rho$ | Ro: evren korelasyon katsayısı |
| $\sigma, \sigma_X$ | Sigma (küçük harf): evren standart sapması |
| $\sigma^2$ | Sigma kare: Evren varyansı |
| $\sigma_e^2$ | Hata varyansı |
| $\Sigma$ | Sigma (büyük harf): toplama sembolü; varyans-kovaryans matrisi |
| $\tau$ | Kendall'ın to'su: korelasyonun parametrik olmayan bir ölçüsü |
| $\tau_j$ | j'nci ölçüm için işlem etkisi |
| $\phi$ | Fi katsayısı: 2x2'lik bir çapraz tablo için korelasyon katsayısı |
| $\phi_C$ | Kramer'in f'si: 2x2'likten büyük çapraz tablolar için fi katsaysının genişletilmiş hali |
| $\chi^2$ | Ki-kare istatistiği: çapraz tablolarla ilgili testlerde kullanılır |
| $\chi_F^2$ | Friedman'ın Ki-Karesi |
| $\psi$ | Pisi: doğrusal karşılaştırma |
| $\omega^2$ | Omega-kare: sıklıkla *r*-ailesi etki büyüklüğü ölçüsü |

## İngiliz Harfi Sembolleri

| | |
|---|---|
| $\alpha$ | Kesim noktası; A değişkeninin düzeyleri sayısı |
| $b$, $b_i$ | Regresyon katsayısı – eğim; A değişkeninin düzeyleri sayısı |
| $b_0$ | Çoklu regresyonda kesim noktası |
| $d$ | Etki büyüklüğü kestirimi |
| $sd$ | Serbestlik derecesi |
| $e_{ij}$ | j'yinci işlemin i'yinci işleme ait birim hatası |
| $E(MS)$ | Beklenen kareler ortalaması |
| $F$ | $F$ istatistiği |
| $FW$ | Aile boyu hata oranı |
| $GM$ | Genel ortalama |
| $H$ | Kruskal–Wallis istatistiği |
| $H_0$; $H_1$ | Sıfır hipotezi; alternatif hipotez |
| ln | Doğal logaritma |
| $MS$ | Kareler ortalaması |
| $n$, $n_i$, $N_i$ | Bir örneklemdeki eleman sayısı |
| $N$ | Örneklemdeki toplam eleman sayısı |
| $N(0, 1)$ | "$\mu = 0$ ve $\sigma^2 = 1$ ile normal dağılmış" olarak okunur |
| $p$ | Genel olarak olasılık için sembol; binom olasılığında başarı olasılığı için |
| $q$ | Student'laştırılmış ranj istatistiği; başarısızlık olasılığı |
| $r$, $r_{XY}$ | Korelasyon katsayısı |
| $r'$ | Fisher'ın r transformayonu |
| $r_b$; $r_{pb}$; $r_t$ | Çift serili, nokta-çift serili ve tetrakorik korelasyon katsayısı |
| $r_S$ | Spearman korelasyon katsayısı |
| $R$, $R_{0.1234}$ | Çoklu korelasyon katsayısı |
| $r_{01.234}$ | Kısmî korelasyon katsayısı |

| | |
|---|---|
| $r_{0(1.234)}$ | Yarı kısmî korelasyon katsayısı |
| $R^*, R^*_{0.123}$ | Evren korelasyon katsayısı |
| $R_{a,\beta,a\beta}$ | A, B ve AB etkileşimi ile kestiricilerle ilgili R |
| $s, s_X$ | Standart sapma |
| $s_D$ | Fark puanlarının standart sapması |
| $s^2_{\square}$ | Birleşik örneklem varyansı |
| $s_{\overline{X}}, s_{\overline{X}} - s_{\overline{X}}$ | Ortalamanın standart hatası; ortalamalar arasındaki farkın standart hatası |
| $s_{Y \cdot X}$ | Kestirimin standart hatası |
| $s^2_X$ | X'in varyansı |
| $SS_A$ | A değişkenine ait kareler toplamı |
| $SS_{AB}$ | Etkileşim kareler toplamı |
| $SS_{\text{error}}$ | Hata kareleri toplamı |
| $SS_{\text{regression}}$ | Regresyondan gelen hata kareleri toplamı |
| $SS_{\text{residual}}$ | Kalan kareler toplamı (regresyona atfedilemeyecek) |
| $t$ | Student t istatistiği |
| $t'$ | Heterojen varyanslarla ilgili t istatistiği |
| $t_d$ | Dunnett t istatistiği |
| $T$ | Wilcoxon eşleştirilmiş -çiftler işaretli- sıralar istatistiği |
| $W$ | Kendall konkordans katsayısı |
| $W_r$ | Açıklık (Range) testleri için kritik genişlik |
| $W_S$ | Wilcoxon sıra-toplamları istatistiği |
| $X, \overline{X}_1$ | Örneklem ortalaması |
| $\bar{X}_h$ | Harmonik ortalama |
| $\hat{Y}, \hat{Y}_i$ | Y'nin kestirilen değeri |
| $z$ | Normal sapma (aynı zamanda standart puan) |



# Tercümede Türkçeleştirilen Sembol ve Kısaltmalar

# Tercümede Türkçeleştirilen Sembol ve Kısaltmalar

| İngilizce Kısaltma | İngilizce Açıklama | Tercümede Kullanılan Türkçe Kısaltmalar | Türkçe Açıklama |
|---|---|---|---|
| AIC | Akaike's Informatin Criterion | ABK | Akaike Bilgi Kriteri |
| SIDS | Sudden Infant Death Seyndrome | ABÖS | Ani Bebek Ölümü Sendromu |
| ACT | Standardized Succes Test for University Admissions in ABD | ACT | ABD'de Üniversiteye Girişte Kullanılan Standart Başarı Testi |
| ADD | Attention Deficite Disorder | ADD | Dikkat Eksikliği Bozukluğu |
| ADDSC | Attention Deficite Disorder Puanı | ADDSC | Dikkat Eksikliği Bozukluğu Puanı |
| FW | Familywise Error Rate | AHO | Aile Boyu Hata Oranı |
| PP | Perceived Popularity | AP | Algılanan Popülerlik |
| CBT | Cognitive Behaviour Therapy | BDT | Bilişsel Davranış Terapisi |
| WL | Waiting List | BL | Bekleme Listesi |
| RR | Relative Risk | BR | Bağıl Risk |
| Cook's D | Cook's Distance | Cook U | Cook Uzaklığı |
| CS Covariance | Variance of the Intercepts | CS Kovaryans | Kesim Noktasının Kovaryansı |
| CS Diagonal Offset | Residual Variance with Balanced Desing | CS Köşegen Uzaklığı | Dengeli Desenlerde Kalan Varyans |
| $\delta$ | Noncentrality Parameter | $\delta$ | Merkezi Olmama Parametresi |
| BST | Behavioral Skill Training | DBE | Davranış Becerisi Eğitimi |
| SC | Supportive Counceling | DD | Destekleyici Danışmanlık |
| LBW | Low-Birth-Wight | DDA | Düşük Doğum Ağırlıklı |
| ADD | Attention Deficit Disorder | DEB | Dikkat Eksikliği Bozukluğu |
| CV | Coefficient of Variation | DK | Değişim Katsayısı |
| $\psi$ | Linear Contrast | $\psi$ | Doğrusal Karşılaştırma |
| LBR | Low Birth Rate | DKD | Doğum Kilosu Düşük Bebekler |
| LSU | Linear Set Up | DSU | Doğrusal Kurma |
| DV, dv | Dependent Variable | DV | Bağımlı Değişken |
| IDU | Injected Drug Use | DYU | Damar Yoluyla Uyuşturucu Kullanma |

| | | | |
|---|---|---|---|
| E | Expected | E | Beklenen |
| MSM | Sexual Contact Between Men | EACİ | Erkekler Arası Cinsel İlişki |
| EC | Educational Control | Eğit. Kon. | Eğitimsel Kontrol |
| ML | Maksimum Likelihood | EYO | En Yüksek Olabilirlik |
| CI | Confidence Interval | GA | Güven Aralığı |
| GPE | Great Point | GBO | Genel Başarı Ortalaması |
| DS | Delayed Smoking | GK | Gecikmiş Kullanıcılar |
| GRE | Graduate Record Exam | GRE | Mezuniyet Sınavı |
| Hedge g | Hedge's Correction | Hedge g | Hedge Düzeltmesi |
| $h_i$ | Leverage | $h_i$ | Baskı |
| PC | Error Rate Per Comparison | HKHO | Her Karşılaştırmanın Hata Oranı |
| PRE | Proportional Reduction in Error | HOA | Hatadaki Oransal Azalma |
| HSD | Tukey's Honestly Significant Difference Test | HSD | Tukey Gerçek Manidar Fark Testi |
| PRESS | Predited Residual Sum of Squares | KAKT | Kestirilmiş Artık Kareler Toplamı |
| RMSSE | Root-Mean-Square Standardized Effect | KAOKSE | Kökü-Alınmış-Ortalama-Karenin Standartlaştırılmış-Etkisi |
| covxy | Covariance of X, Y | kovxy | XY'nin Birlikte/Ortak Değişkenliği (kovaryans) |
| SS | Sum of Squares | KT | Kareler Toplamı |
| EDA | Explatory Data Analysis | KVA | Keşfedici Veri Analizi |
| $\delta$ | Noncentralit Parameter | $\delta$ | Merkezî Olmama Parametresi |
| L | Linear Combination | L | Doğrusal (Lineer) Kombinasyon |
| LogPctSAT | Logaritm of Percentage of Taking SAT | LogPcSAT | SAT Sınavına Katılan Öğrenci Yüzdesinin Logaritması |
| P/T | Pupil/Teacher Ratio | P/T | Öğrenci/Öğretmen Yüzdesi |
| LSD | Fishers's Least Significant Difference Test | LSD | Fisher En Küçük Farklar Testi |
| Mallow's Cp | Mallow's Cp Statistic | Mallow's Cp | Mallow Cp İstatistiği |
| MMPI | Minnesota Multiphasic Personality Inventory | MÇYKE | Minnesota Çok Yönlü Kişilik Envanteri |
| GRE | Graduate Record Exam | MS | Mezuniyet Sınavı |

| | | | |
|---|---|---|---|
| $S_w^2$ | Winsorized Sample Variance | $S_w^2$ | Winsorize Edilmiş Örneklem Varyansı |
| FE | Fixed Effect | SA | Sabit Etki |
| SAT | Scolastic Aptitude Test | SAT | Okul Yetenek Testi |
| $\sigma_p^2$ | Pooled Popülation Variance | $\sigma_p^2$ | Birleşik Evren Varyansı |
| $\sigma_p$ | Pooled Population Standart Deviation | $\sigma_p$ | Birleşik Evren Standart Sapması |
| REML | Restiricted Maksimum Likelihood | SEYO | Sınırlandırılmış En Yüksek Olabilirlik |
| df | Degrees of Freedom | sd | Serbestlik Derecesi |
| CSW | Commercial Sex Worker | Sİ | Seks İşçisi |
| CSW-cl | Clients of Commercial Sex Workers | SİM | Seks İşçisi Sektörünün Müşterileri |
| SIT | Stres Inoculation Therapaty | SKAT | Strese Karşı Aşılama Terapisi |
| SP | Sociometric Populatrity | SP | Sosyometrik Popülerlik |
| $\sigma_w$ | Winsorized Population Standart Deviation | $\sigma_w$ | Winsorize Edilmiş Evren Standart Sapması |
| $s_w$ | Winsorized Sample Standart Deviation | $s_w$ | Winsorize Edilmiş Örneklem Standart Sapması |
| TT | Terapeutic Touch | TED | Tedavi Edici Dokunuşlar |
| THC | The Most Psychoactive Ingredient in Marijuana | THC | Esrardaki Psikoaktif Madde |
| RPS | Rock/Paper/Scissors | TKM | Taş/Kâğıt/Makas |
| NHL | National Hokey Leage in the ABD | UHL | ABD'de Ulusal Hokey Ligi |
| PE | Prolonged Exposure | UP | Uzatılmış Pozlama |
| VIF | Variance Inflation Factor | VBF | Varyans Bükülme Faktörü |
| W | Kendall's Coefficient of Concordance | W | Kendal Konkordans Katsayısı |
| DepressT | Depression T Score | DTP | Depresyon T Puanı |
| WRAT | Wide Range Achievement Test | WRAT | Geniş Kapsamlı Başarı Testi |
| Ws | Wilcoxon Statistic | Ws | Wilcoxon İstatistiği |
| WSD | Whooly Significant Difference | WSD | Tümüyle Manidar Fark |
| $\psi$ | Linear Contrast | $\psi$ | Doğrusal (Lineer) Karşılaştırma |
| FDR | False Discovery Rate | YBO | Yanlış Buluş Oranı |

| | | | |
|---|---|---|---|
| PIP | Proportional Improvement in Prediction | YOG | Yordamadaki Oransal Gelişim |
| $\bar{X}_h$ | Harmonik Mean | $\bar{X}_h$ | Harmonik Ortalama |
| Ageatlos | Age at Parent's Death | AgeaLos | Ebeveyn Öldüğünde Öğrencinin Yaşı |
| AnxT | Anxiety T Score | KaygıT | Kaygı T Puanı |
| DepT | Depression T Sccore | DepT | Depresyon T Puanı |
| DROPOUT | Leave School | DROPOUT | Okulu Terk: 1 = Okulu Bitirmeden Terk Etti, 0 = Terk Etmedi. |
| ENGG | Grade in English in ninth grade: 4,5 A; 3,5 B; and so on | ENGG | 9. Sınıf İngilizce Puanı: 4 = A; 3 = B vb. |
| GST | Student's Psychological Symptoms | GST | Öğrencinin Psikolojik Sendromu |
| Host T | Hostility T score | HostT | Düşmanlık T Puanı |
| LBW | Low-Birthweight | DDA | Düşük Doğum Ağırlıklı |
| LostPGren | Gender of Lost Parent | LostPGen | Ölmüş Olan Ebeveynin Cinsiyeti |
| ObsessT | Obsessive-compulsive T score | ObsessT | Takıntı T Puanı |
| ParT | Paranoid Ideation T score | ParT | Paranoit Endişe T Puanı |
| PhobT | Phobic Anxiety T score | PhobT | Paranoit Düşünce T Puanı |
| PsyT | Psychoticism T score | PsyT | Psikotizm T Puanı |
| PVLoss | Perceived Vulnerability to Loss | AGKA | Algılanan Güvenlik Kaybı Puanı |
| PVTotal | Perceived Vulnerability total score | AGATP | Algılanan Güvenlik Kaybı Toplam Puanı |
| REML | Restricted Maximum Likelihood | REML | Sınırlandırılmış En Yüksek Olasılık |
| SensitT | Interpersonal Sensitivity T score | SensitT | Kişiler Arası Duyarlılık T Puanı |
| SexChild | Gender of Child | SexChild | Çocuğun Cinsiyeti |
| SexP | Gender of Patient | SexP | Hastanın Cinsiyeti |
| SexS | Gender of Spouse | SexS | Eşin cinsiyeti |
| SOCPROB | Social Problems in Ninth Grade | SOCPROB | Dokuzuncu Sınıftaki Sosyal Problemleri: 1 = Yok, 2 = Var |
| SomT | Somaticisim T score | SomT | Somatizm T Puanı |
| TotBP | Total Behavior Problems | TotBP | Toplam Davranış Problemleri |
| TotBP T | Total Behavior Problem T Score | TotBP T | Toplam Davranış Problemleri T Puanı |

| | | | |
|---|---|---|---|
| VIF | Variance Inflation Factor | VAF | Varyans Bükülme Faktörü |
| W | Kendall's Coefficient of Concordance | W | Kendal Kondordans Katsayısı |
| CMA | Comprehensive Meta-Analysis | KMA | Kapsamlı Meta-Analiz |
| PTSD | Post-traumatic Stress Disorder | PTSD | Travmatik Stres Bozukluğu Sonrası |
| $\eta^2$ | Eta kare: Squared Correlation Rratio | $\eta^2$ | Eta kare: Korelasyon Oranının Karesi |
| GSIT | Psychological Symptoms | GSIT | Psikolojik Semptomlar |
| GPA | Grade Point Average 4 = A; 3 = B; 2 = C; 1 = D; 0 = F | GBO | ABD'de Okul Öğrenmelerini Ölçmeye Yönelik Başarı Testi |
| WRAT | Wide Range Achievement Test | GKBT | Geniş Kapsamlı Başarı Testi |
| $\overline{X}_h$ | Homofobik Group Average | $\overline{X}_h$ | Homofobik Grubun Ortalaması |
| $\overline{X}_{h_o}$ | Nonhomofobik Group Average | $\overline{X}_{h_o}$ | Homofobik Olmayan Grubun Ortalaması |
| N! | N faktoriyel | N! | N faktöriyel: 1'den N'ye kadar olan doğal sayıların çarpımı |
| NBA | National Basketball Ligeau of ABD | NBA | ABD'nin Ulusal Basketbol Ligi |
| CNr | Number of Combinations Taking r Objects From N objects at a Time | CNr | N Elemanın r'li Kombinasyonlarının Sayısı |
| NBW | Normal-Birght Weight | NDA | Normal Doğum Ağırlıklı |
| $P_r^N$ | Number of Permutations Taking r Objects From N objects at a Time | $P_r^N$ | N Elemanın r'li Permütasyonlarının Sayısı |
| $\omega^2$ | Omega-Squared: Most Often r-family Measure of Effect size | $\omega^2$ | Omega-Kare: Sıklıkla r-ailesi Ölçümlerinde Etki Büyüklüğü Ölçüsü |
| OR | Odds Ratio | OO | Olasılık Oranı |
| P/T | Pupil/Teacher Ratio | P/T | Öğrenci/Öğretmen Oranı |
| PctSAT | Percentage of Number of Students Taking SAT | PctSAT | SAT'a Katılan Öğrenci Yüzdesi |
| PDI | Psycomotor Development Index | PGI | Psikomotor Gelişim İndeksi |
| q | Studentized Range Statistics | q | Student'laştırılmış Ranj İstatistiği |
| Q1 | First Quartile | Q1 | Birinci Çeyrek |
| Q2 | Second Quartile: median | Q2 | İkinci Çeyrek: ortanca |

| | |
|---|---|
| Q3 | Thirty Quartile |
| LogPctSAT | Logaritm of PctSAT |
| Q-Q | Percentile-Percentile Distribution |
| $r_{adj}$ | r adjusted |
| $r_s$ | Sperman's Correlation Coefficient for Ranked Data |
| REGWQ | Ryan Test |
| $r_h$ | Biserial Correlation Coefficient |
| $r_{ph}$ | Point-biserial Correlation Coefficient |
| NBW | Normal-Birthweight |
| RR | Risk Ratio |
| PDI | Psychological Development Index |
| $\rho_{XY}$ | Popülation Covariance of X and Y |
| $s_{XY}$ | Örneklem Covariance of X and Y |

| | |
|---|---|
| Q3 | Üçüncü Çeyrek |
| logPctSAT | PctSAT'ın Logaritması |
| Q-Q | Yüzdelik-Yüzdelik Dağılımı |
| rdüzeltilmiş | Düzeltilmiş r |
| $r_s$ | Sıralı Veri İçin Sperman Korelasyon Katsayısı |
| REGWQ | Ryan Testi |
| $r_ç$ | Çift Serili Korelasyon Katsayısı |
| $r_{nç}$ | Nokta-Çift Serili Korelasyon Katsayısı |
| NDA | Normal Doğum Ağırlıklı |
| RO | Risk Oranı |
| PGİ | Psikomotor Gelişim İndeksi |
| $\rho_{XY}$ | X ile Y'nin Evren Kovaryansı |
| $s_{XY}$ | X ile Y'nin Örneklem Kovaryansı |