TEMEL İSTATİSTİK

Prof. Dr. Münevver TURANLI • Prof. Dr. Selahattin GÜRİŞ

genişletilmiş 3. basım

DER YAYINLARI

TEMEL İSTATİSTİK

Prof. Dr. Münevver TURANLI • Prof. Dr. Selahattin GÜRİŞ

MARMARA ÜNİVERSİTESİ
İktisadi ve İdari Bilimler Fakültesi
Ekonometri Bölümü, Öğretim Üyesi

MARMARA ÜNİVERSİTESİ
İktisadi ve İdari Bilimler Fakültesi
Ekonometri Bölümü, Öğretim Üyesi

Genişletilmiş Üçüncü Basım

DER
YAYINLARI
İSTANBUL - 2010

Kitabın Adı
Temel İstatistik

Yazarlar
Münevver Turanlı – Selahattin Güriş

Yayıncı
Der Yayınları
Molla Fenari Sokak, Der Han 40-42, 34410 Cağaloğlu - İSTANBUL.
Tel: (0212) 527 01 65 - 511 51 90 Belgegeçer: (0212) 511 47 76
www.deryayinevi.com.tr
e-posta:der@deryayinevi.com.tr / info@deryayinevi.com.tr

Basım
Ada Ofset / Topkapı-İSTANBUL

Cilt
Kısmet Ciltevi / Topkapı-İSTANBUL

Yayın No 273
ISBN 978-975-353-199-3

© DER YAYINLARI - Nisan 2010

Copyright © Bu kitabın, Türkiye'de yayın hakları ***Der Yayınevi***'ne aittir. Her hakkı saklıdır. Yayınevimizden yazılı izni alınmadan kısmen veya tamamen alıntı yapılamaz. Hiçbir şekilde kopya edilemez; fotokopi, faksimile veya başka bir şekilde çoğaltılamaz ve yayınlanamaz.

İÇİNDEKİLER

BÖLÜM 1 1-6
GİRİŞ
1.1. İSTATİSTİĞİN TANIMI 1
1.2. İSTATİSTİĞİN KONUSU OLAN OLAYLAR 2
1.3. İSTATİSTİĞİN ÇEŞİTLİ BİLİM DALLARI İLE İLİŞKİSİ 3
1.4. İSTATİSTİĞİN TARİHÇESİ 4
BÖLÜM 2 7-40
VERİLERİN TOPLANMASI ve DÜZENLENMESİ
2.1. VERİ VE VERİ KAYNAKLARI 7
2.2. TAMSAYIM VE ÖRNEKLEME 8
2.3. DOĞRUDAN VERİ TOPLAMA YÖNTEMLERİ 11
2.3.1. Anket 11
2.3.2. Gözlem 14
2.3.3. Deney 14
2.4. VERİLERİN DÜZENLENMESİ 15
2.4.1. Basit Seri 15
2.4.2. Frekans Serisi 15
2.4.3. Sınıflı Seri 17
2.5. BİRİKİMLİ FREKANSLAR 19
2.5.1. Artan Birikimli Frekanslar 20
2.5.2. Azalan Birikimli Frekanslar 21
2.6. ORANSAL FREKANSLAR 22
2.7. İSTATİSTİKSEL SERİLER 25
2.7.1. Zaman Serileri 25
2.7.2. Mekân Serileri 25
2.7.3. Dağılım Serileri 26
2.7.4. Bileşik Seriler 26
2.8. GRAFİKLER 27
2.8.1. Zaman Serilerinin Grafikleri 27
2.8.2. Dağılım Serilerinin Grafikleri 28
2.8.3. Bileşik Serilerin Grafikleri 32
2.9. ÇÖZÜMLÜ ÖRNEKLER 33
BÖLÜM 3 41-90
ORTALAMALAR
3.1. ARİTMETİK ORTALAMA 42
3.1.1. Aritmetik Ortalamanın Özellikleri 44
3.1.2. Tartılı Aritmetik Ortalama 55
3.2. KARELİ ORTALAMA 59
3.3. GEOMETRİK ORTALAMA 61
3.4. HARMONİK ORTALAMA 64
3.5. MOD 66
3.6. MEDYAN 70
3.7. BÖLENLER 75

3.8. UYGUN ORTALAMANIN SEÇİMİ 77
3.9. ÇÖZÜMLÜ ÖRNEKLER 78
BÖLÜM 4 91-146
DEĞİŞKENLİK VE ASİMETRİ ÖLÇÜLERİ
4.1. DEĞİŞKENLİK ÖLÇÜLERİ 91
4.1.1. Değişim Aralığı 93
4.1.2. Standart Sapma 93
4.1.3. Değişim Katsayısı 115
4.1.4. Toplanma Oranı 116
4.2. ASİMETRİ ÖLÇÜLERİ 118
4.2.1. Ortalamalar ve Kartiller Arası Farklar ile Hesaplanan Asimetri Ölçüleri 119
4.2.2. Momentler ile Hesaplanan Asimetri ve Basıklık Ölçüleri 125
4.3. ÇÖZÜMLÜ ÖRNEKLER 131
BÖLÜM 5 147-186
İNDEKSLER
5.1. İNDEKS KAVRAMI 147
5.2. MEKÂN İNDEKSLERİ 148
5.3. ZAMAN İNDEKSLERİ 149
5.3.1. Sabit Esaslı İndeksler 150
5.3.2. Değişken Esaslı İndeksler 153
5.3.3. Sabit Esaslı İndekslerden Değişken Esaslı İndekslerin Hesaplanması 156
5.3.4. Değişken Esaslı İndekslerden Sabit Esaslı İndekslerin Hesaplanması 157
5.3.5. Sabit Esaslı İndekslerde Temel Devrenin Değiştirilmesi 158
5.4. BİLEŞİK İNDEKSLER 160
5.4.1. Tartısız Bileşik İndeksler 160
5.4.2. Tartılı Bileşik İndeksler 169
5.5. CARİ FİYATLAR İLE HESAPLANMIŞ VERİLERİN SABİT FİYATLAR İLE HESAPLANMIŞ VERİLERE İNDİRGENMESİ 178
5.6. GELİR, FİYAT VE MİKTAR İNDEKSLERİ ARASINDAKİ İLİŞKİ 179
5.7. ÇÖZÜMLÜ ÖRNEKLER 180
BÖLÜM 6 187-204
OLASILIK
6.1. ÖRNEK UZAYI, PERMÜTASYON VE KOMBİNASYON KAVRAMLARI 187
6.1.1. Permütasyon 188
6.2.2. Kombinasyon 189
6.2. OLASILIĞIN TANIMI VE TEMEL ÖZELLİKLERİ 191
6.3. BİRDEN FAZLA OLAYIN OLASILIĞI 193
6.3.1. Bağdaşan Olaylar 193
6.3.2. Bağlı Olaylar 193
6.3.3. Şartlı Olasılık 198
6.4. ÇÖZÜMLÜ ÖRNEKLER 199
BÖLÜM 7 205-222
TESADÜFİ DEĞİŞKENLER
7.1. TESADÜFİ DEĞİŞKEN KAVRAMI 205

7.2. SÜREKSİZ TESADÜFİ DEĞİŞKENLER 206
7.3. SÜREKLİ TESADÜFİ DEĞİŞKENLER 212
7.4. ÇÖZÜMLÜ ÖRNEKLER 217
BÖLÜM 8 223-250
ÇOK BOYUTLU TESADÜFİ DEĞİŞKENLER
8.1. SÜREKSİZ TESADÜFİ DEĞİŞKENLER 223
8.1.1. Beklenen Değer 226
8.1.2. Beklenen Değerin Özellikleri 228
8.1.3. Varyans 230
8.1.4. Varyansın Özellikleri231
8.2. SÜREKLİ TESADÜFİ DEĞİŞKENLER 236
8.2.1. Beklenen Değer 237
8.2.2. Beklenen Değerin Özellikleri 239
8.2.3. Varyans 241
8.3. KOVARYANS VE ÖZELLİKLERİ 242
8.4. ÇÖZÜMLÜ ÖRNEKLER 246
BÖLÜM 9
OLASILIK DAĞILIMLARI 251-292
9.1. BİNOM DAĞILIMI 251
9.2. HİPERGEOMETRİK DAĞILIM 255
9.3. POISSON DAĞILIMI256
9.4. NORMAL DAĞILIM 257
9.4.1. Normalliğin İncelenmesi 264
9.4.2. Binom Dağılımının Normal Dağılıma Yaklaştırılması 267
9.5. DİĞER DAĞILIMLAR 271
9.5.1. Ki-kare Dağılımı 271
9.5.2. t- Dağılımı 272
9.5.3. F Dağılımı 273
9.6. ÇÖZÜMLÜ ÖRNEKLER 274
BÖLÜM 10 293-330
TAHMİN TEORİSİ
10.1. TAHMİN VE TAHMİNCİ 293
10.2. ÖRNEKLEME DAĞILIMI VE STANDART HATA 294
10.3. SAPMA, ÖRNEKLEME HATASI VE ORTALAMA HATA KARE 295
10.4. TAHMİNCİNİN ÖZELLİKLERİ 296
10.4.1. Sapmasızlık 296
10.4.2. Etkinlik 297
10.4.3. Tutarlılık 298
10.4.4. Yeterlilik 299
10.5. TAHMİN YÖNTEMLERİ 299
10.5.1. Momentler Yöntemi 299
10.5.2. En Küçük Kareler Yöntemi 300
10.5.3. Maksimum Benzerlik Yöntemi 301
10.5.4. En İyi Doğrusal Sapmasız Tahmin Yöntemi 305

10.6. ARALIK TAHMİNİ 306
10.6.1. Anakütle Ortalamasının Aralık 307
10.6.2. Anakütle Oranının Aralık Tahmini 313
10.6.3. Ortalamalar Arasındaki Farkların Aralık Tahmini 315
10.6.4. Oranlar Arasındaki Farkların Aralık Tahmini 317
10.6.5. Anakütle Varyansmın Aralık Tahmini 319
10.6.6. Varyans Oranlarının Aralık Tahmini 320
10.6.7. Optimum Örnek Büyüklüğünün Belirlenmesi 321
10.7. ÇÖZÜMLÜ ÖRNEKLER 324
BÖLÜM 11 331-356
HİPOTEZ TESTLERİ
11.1. HİPOTEZ VE HİPOTEZ TESTİ KAVRAMLARI 331
11.2. HİPOTEZ TESTİNİN AŞAMALARI 333
11.3. ORTALAMALARIN TESTİ 335
11.4. ORANLARIN TESTİ 338
11.5. ORTALAMALAR ARASINDAKİ FARKLARIN TESTİ 340
11.5.1. Bağımsız Örnekler 340
11.5.2. Bağımlı Örnekler 342
11.6. ORANLAR ARASINDAKİ FARKLARIN TESTİ 344
11.7. ANAKÜTLE VARYANSININ TESTİ 346
11.8. VARYANS ORANLARININ TESTİ 347
11.9. ÇÖZÜMLÜ ÖRNEKLER 349
BÖLÜM 12 357-382
VARYANS ANALİZİ
12.1. İKİDEN FAZLA ORTALAMA ARASINDAKİ FARKLARIN TESTİ 357
12.2. TEK YÖNLÜ VARYANS ANALİZİ 361
12.3. ÇİFT YÖNLÜ VARYANS ANALİZİ 367
12.3.1. Etkileşimsiz Çift Yönlü Varyans Analizi 368
12.3.2. Etkileşimli Çift Yönlü Varyans Analizi 372
12.4. ÇÖZÜMLÜ ÖRNEKLER 378
BÖLÜM 13 383-452
BASİT REGRESYON VE KORELASYON
13.1. BASİT DOĞRUSAL REGRESYON 383
13.1.1. Parametrelerin Tahmini 386
13.1.2. Teorik Değerlerin Hesaplanması 392
13.1.3. Parametre Tahmincilerin Varyansları 394
13.1.4. Belirlilik Katsayısı 399
13.1.5. Katsayıların Testi 403
13.1.6. Regresyon Modelinin Tahmin İçin Kullanılması 408
13.2. KORELASYON KATSAYISI 409
13.3. BASİT DOĞRUSAL KORELASYON VE REGRESYON ARASINDAKİ İLİŞKİ 415
13.4. EĞRİSEL KORELASYON VE REGRESYON 422
13.5. ÇÖZÜMLÜ ÖRNEKLER 438
BÖLÜM 14 453-494

ÇOKLU REGRESYON VE KORELASYON
14.1. ÇOKLU REGRESYON 453
14.1.1. Parametrelerin Tahmini 454
14.1.2. Tahmincilerin Varyansları 463
14.1.3. Belirlilik Katsayısı 466
14.1.4. Katsayıların Testi 469
14.1.5. Çoklu Regresyon Modelinin Tahmininde Kullanılması 470
14.1.6. Eğrisel Çoklu Regresyon 472
14.2. ÇOKLU VE KISMİ KORELASYON 475
14.2.1. Çoklu Korelasyon Katsayısı 475
14.2.2. Kısmi Korelasyon Katsayısı 476
14.3. ÇÖZÜMLÜ ÖRNEKLER 478
BÖLÜM 15 495-536
ZAMAN SERİLERİ ANALİZİ
15.1. ZAMAN SERİLERİNİ ETKİLEYEN FAKTÖRLER 496
15.2. TRENDİN BELİRLENMESİ 497
15.2.1. Yarı Ortalamalar Yöntemi 497
15.2.2. Hareketli Ortalamalar Yöntemi 500
15.2.3. Trendin Matematiksel Modeller İle Belirlenmesi 507
15.3. MEVSİMLİK DALGALANMALARIN BELİRLENMESİ 523
15.4. KONJOKTÜREL DALGALANMALARIN BELİRLENMESİ 527
15.5. ÇÖZÜMLÜ ÖRNEKLER 529
BÖLÜM 16 537-578
PARAMETRİK OLMAYAN TESTLER
16.1. PARAMETRİK OLMAYAN TEST KAVRAMI 537
16.2. İŞARET TESTİ 539
16.3. WILCOXON TESTİ 541
16.3.1. Wilcoxon Sıra Toplamı Testi 541
16.3.2. Wilcoxon İşaretli Sıralar Testi 544
16.4. MANN-WHITNEY U TESTİ 547
16.5. KRUSKAL-WALLIS H TESTİ 551
16.6. SPEARMAN SIRA KORELASYONU KATSAYISI VE TESTİ 553
16.7. Kİ-KARE TESTİ 556
16.7.1. Ki-Kare Bağımsızlık Testi 557
16.7.2. Ki-Kare Homojenlik Testi 560
16.7.3. Ki-Kare Uygunluk Testi 562
16.8. ÇÖZÜMLÜ ÖRNEKLER 564
TABLOLAR 579
YARARLANILAN KAYNAKLAR 603
DİZİN 607

BÖLÜM 1

GİRİŞ

İstatistik kelimesi hemen hemen herkes tarafından bilinirse de, farklı kişilere farklı anlamlar ifade etmektedir. Bu kelime konu ile daha az ilgisi olanlar için genellikle alt alta yazılmış rakam topluluklarını veya grafikleri ifade ederken, konuya daha yakın olanlar için farklı anlam taşımaktadır.

İstatistik kelimesi ilk olarak Almanya'da devlet anlamına gelen status kelimesine dayanılarak kullanılmaya başlanmıştır. İstatistiksel araştırmaların geçmişi, istatistik kelimesinden de eskiye dayanmaktadır. Başlangıçta çok kısıtlı alanlarda başlayan çalışmalar, matematikçilerin geliştirdikleri teorilere paralel olarak istatistiğin gelişmesi ile daha geniş uygulama alanları bulabilmiştir. İstatistik günümüzde pek çok alanda uygulanmakta ve özellikle bilgisayarların yaygın olarak kullanılmaya başlanması ile uygulama alanı giderek genişlemektedir.

1.1. İSTATİSTİĞİN TANIMI

Uygulama alanı çok geniş olan istatistiğin farklı tanımları yapılabilir. İstatistikten genel anlamı ile sayısal analizler anlaşıl-maktadır. Anlamını daha iyi açıklayacak bir tanım şöyle yapılabilir.

Verilerin toplanması, düzenlenmesi ve uygun yöntemlerle analiz edilerek sonuçların yorumlanmasına istatistik denir.

Tanımdan da anlaşılacağı gibi istatistikten söz edebilmek için önce veriye ihtiyaç vardır. Sayısal veya sayısal olarak ifade edilebilen bilgiler, istatistiksel analizlerin verileridir. Toplanan verilerin analize uygun hale getirilmesi için düzenlenmesi gerekebilir. Verilerin düzenlenmesinden sonra analiz için uygun istatistiksel yöntem veya yöntemler seçilecektir.

İstatistikçiler farklı alanlarda karşılaşılan farklı durumlar için çeşitli yöntemler geliştirmişlerdir. Araştırılan konuya uygun yöntemin seçimi araştırmacı veya araş-

tırmacıların görevidir. Bazı durumlarda aynı veri farklı yöntemlerle de incelenebilir. Bu durumda en uygun yöntem seçilecek veya farklı yöntemler ile analiz yapılarak sonuçlar karşılaştırılarak en iyi sonuca ulaşılmaya çalışılacaktır.

Elde edilen verinin analizi ile sonuçlar elde edilecektir. İstatistiksel analiz sayısal sonuçların elde edilmesi ile sona ermez. Sonuçlar genellikle tam bir kesinliği ifade etmediğinden yorumlanmaları gerekecektir. İstatistiksel analizler bir amaç için yapılırlar. Elde edilen sonuçlar araştırılan konu ile ilgili bir kararın alınmasına yardımcı olacaktır. Bu nedenle verilerin toplanmasından sonuçların yorumlanmasına kadar herhangi bir aşamada yapılacak hata, sonuçları etkileyecek, hatalı kararlara neden olacaktır.

Yukarıdaki açıklamalarımızdan da anlaşıldığı gibi istatistik, sayısal değerlerin toplanarak analiz edildiği bilimsel bir metot veya metotlar topluluğudur. Bazı yazarlarca istatistiğin bir bilim olduğu söylenmektedir. Bilimden söz edebilmek için bilimin kendine özgü bir konusu ve bu konuda bir bütünlüğü olması gerekir. İstatistik, matematik temellere dayanır ve kendine özgü bir konusu yoktur. Farklı konularda kullanılabilen yöntemlerin bir araya gelmesi ile ortaya çıkmıştır. Tıp, matematik, iktisat vb. için söylediklerimizi istatistik için söyleyemeyiz. Ayrıca istatistiksel yöntemler arasında tam bir ilişki de söz konusu değildir. İstatistiksel yöntemler çeşitli bilim dallarında araştırmalar yapmak için kullanılırken, bilim dallarının özelliklerine göre istatistiksel yöntemlerin bazıları kullanılmakta, bazıları ise kullanılmamaktadır. Bu açıklamalara göre istatistiğin kendine özgü bir konusu olmadığı ve kullanılan yöntemler arasında tam bir ilişki bulunmadığı sonucuna varılmaktadır. Bu sonuca göre, istatistiğin bir bilim olduğu görüşünü savunmamız zordur.

İstatistiğin bir bilim olduğunu savunanlar ise, istatistiksel yöntemlerin kullanıldığı alanların başka bir şekilde doldurulamadığını, bu nedenle bilim olduğunu ileri sürmektedirler. Sonuç olarak, istatistiğin bilim olsa da, olmasa da yararlı olduğunu ve ihtiyaç duyulduğunu söyleyebiliriz.

1.2. İSTATİSTİĞİN KONUSU OLAN OLAYLAR

Karşılaştığımız olayları tipik ve toplu olaylar olarak iki ana grupta toplayabiliriz. Tipik olaylar az sayıda faktör tarafından etkilenen, fiziksel, kimyasal olaylar gibi olaylardır. Bu olaylar, az sayıda faktör tarafından etkilendiklerinden, bir kaç deneme ile olayların genel karakteri belirlenebilir ve olaylara etki eden faktörler kontrol edilebilirler. Örneğin, kimyada A ve B gibi iki maddenin belirli şartlarda bileşmesi sonucu ne olacağı söylenebilir. Tipik olaylar istatistiğin konusu dışındadır. Ve istatistik bu tür olaylarla ilgilenmez.

Toplu olaylar olarak adlandırılan olaylar ise, tipik olayların aksine çok sayıda faktör tarafından etkilenirler. Bu olaylar çok sayıda faktör tarafından etkilendiklerinden, birkaç deneme ile olayların genel karakteri belirlenemez ve olaylara etki eden faktörler kontrol edilemez. Örneğin, iktisadi olaylar çok sayıda faktör tarafından etkilenen olaylardır. Bu nedenle, olayı etkileyen ana faktörler bilinse de tüm faktörler kontrol edilemez. İstatistik, toplu olaylar olarak adlandırılan bu tür olaylar ile ilgile-

nir. Sonuçları sayısal olan veya sayısal olarak ifade edilebilen olaylar istatistiğin konusudur.

İstatistiği tanımlarken sayısal olarak ifade edilebilen verilerden söz etmiştik. İstatistiksel araştırmalarda kullanılacak veriler değişkenler ile ifade edilirler. Değişkenler sayısal olarak ifade edilebilen özelliklerdir ve X_i, Y_i, Z_i ..... şeklinde gösterilirler.

Değişkenleri ifade ettikleri olayın karakterine göre nicel ve nitel olmak üzere ikiye ayırabiliriz. Nicel değişkenler, sayısal değerler alan değişkenlerdir. Örneğin, boy uzunluğu, vücut ağırlığı birer nicel değişkendir. Nitel değişkenler ise sayısal değerler almayan değişkenlerdir. Örneğin, cinsiyet, medeni hal gibi değişkenler nitel değişkenlerdir. Nitel değişkenlere sayısal değerler verilerek, istatistiksel araştırmalarda kullanılmaları mümkün olmaktadır. Örneğin, medeni hal nitel bir değişkendir, ancak bu değişken, bekarlar için 0, evliler için 1 ve dullar için 2 sayısal değeri verilerek, nicel değişkene çevrilebilir.

Ayrıca değişkenler, sürekli ve süreksiz değişkenler olmak üzere ikiye ayrılırlar. Sürekli değişkenler, tanımlı oldukları aralık içinde bulunan bütün değerleri alabilen değişkenlerdir. Bu değişkenler matematiksel bir fonksiyon ile ifade edilebilirler. Boy uzunluğu, ağırlık, hız gibi değişkenler sürekli değişkenlerdir. Örneğin, boy uzunluğu 173,4 cm. veya hız 70,5 km. olabilir.

Süreksiz değişkenler ise, tanımlı oldukları aralık da bulunan değerlerden bazılarını alan değişkenlerdir. Bir evde yaşayan kişilerin sayısı, bir limana bir günde giren gemi sayısı gibi değişkenler süreksiz değişkenlerdir. Çünkü bu gibi değişkenler ancak tam sayı değerler alırlar. Örneğin, bir evde 2,25 kişi yaşayamayacağı gibi, bir limana bir günde 3,4 gemi giremez.

1.3. İSTATİSTİĞİN ÇEŞİTLİ BİLİM DALLARI İLE İLİŞKİLERİ

İstatistiksel yöntemler sayısal verilerin kollektif olay karakterini taşıyan olayların söz konusu olduğu tüm alanlarda kullanılabilirler. Bu nedenle istatistiğin hemen hemen tüm bilim dalları ile ilişkisi vardır.

İstatistiğin matematik ile olan ilişkisi diğer bilim dallarından çok farklıdır. İstatistik matematikten aldığı yöntemleri işleyerek, toplu olayların incelenmesine yarayacak şekle getirmekte ve diğer bilim dallarının hizmetine sunmaktadır. Bilindiği gibi matematik bir bilim dalıdır ve matematiksel sonuçlar kesindir, yani matematiksel sonuçlarda yorum yoktur. Matematikten gelen yöntemlerden oluşan istatistik de ise sonuçlar yorumlanmaktadır. İstatistiğin tanımında da verilerin çeşitli yöntemlerle analiz edilerek, sonuçların yorumlanmasından söz etmiştik. Görüldüğü gibi matematik ile istatistik arasındaki en büyük fark sonuçların yorumunda karşımıza çıkmaktadır.

İstatistiksel yöntemler iktisadın deneysel gelişimini de sağlamıştır. Makro ve mikro düzeyde teoride kalan pek çok kavramın sayısal olarak ölçülmesi ve teorinin

uygulanması istatistiksel yöntemler ile mümkün olabilmektedir. Örneğin, enflasyon, reel gelir, büyüme gibi olaylar istatistik yardımı ile ölçülebilmektedir.

Gerek devlet, gerekse işletmeler planlama çalışmalarında istatistiksel yöntemleri kullanabilirler. Geçmişe ait verilerin incelenmesi ile geleceğe ait tahminler istatistiksel araştırmalarla yapılabilir. Devletler nüfus, eğitim, işgücü, ücretler, dış ticaret, ulaşım gibi konularda istatistiksel bilgiler düzenlemektedirler. Bu bilgiler hem devlet, hem de araştırmacılar tarafından veri olarak kullanılmaktadırlar. İktisadın bir alt dalı olan ekonometride de istatistiksel yöntemler yardımı ile iktisadi olaylar incelenmekte ve böylece iktisadi değişkenler arasındaki ilişki matematiksel olarak belirlenebilmektedir.

İstatistiksel yöntemler işletmeler tarafından da yaygın olarak kullanılmakta; iş hayatında karşılaşılan belirsizlikler ve buna bağlı problemlerin çözümünde, yol gösterici olmaktadırlar. Bu yöntemlerin başlıcaları; piyasa ve ürün araştırmaları, yatırım politikalarının belirlenmesi, mamulün kalite kontrolü, personel seçimi, iktisadi tahminler ve benzeri diğer konulardır.

İstatistiksel yöntemler sağlık bilimlerinde de biyoistatistik adı altında geniş bir alanda kullanılmaktadır. Bu alanda, sağlık problemlerinin belirlenmesi, tedavi hizmetlerinin planlanması, kliniklerde yapılan araştırmaların sonuçlarının yorumlanması, yeni tedavi yöntemlerinin ve ilaçların değerlendirilmesi gibi konularda istatistiksel yöntemler kullanılmaktadır. İstatistiksel yöntemler, psikoloji, sosyoloji, antropoloji, astronomi, meteoroloji, tarih, linguistik, edebiyat, müzik gibi alanlarda da çok yaygın olmamakla beraber kullanılmaktadır. Örneğin, edebiyat alanında, yazarı bilinmeyen kitapların yazarlarının belirlenmesinde de kullanıldığı gibi.

İstatistiğin pek çok alanda kullanılması, her alanda aynı yöntemlerin, aynı şekilde kullanılacağı anlamına gelmez. Uygulanan alana ve uygulama konusuna göre olaya daha farklı yaklaşmak ve daha uygun yöntemleri seçmek gerekebilir. Araştırmayı yapacak kişi veya kişilerin, istatistik bilgilerinin yeterli düzeyde olması yanında, uygulanan konuyu da yeterince bilmeleri gerekmektedir. Farklı kişi veya kişilerce yapılan araştırmalar, konuya yaklaşım farkından veya farklı yöntemler kullanılmasından farklı sonuçlar verebilir. Bu durumda, ilgili kişilerin hangi sonucun daha geçerli olduğuna karar vermeleri gerekir. İstatistiksel araştırmalar bazen mantıksız sonuçlarda verebilir. Bunun sebebi, araştırma yapıldığı anda verinin tesadüfen ters sonuç verecek şekilde bir araya gelmiş olmasıdır. Örneğin, yapılan bir araştırma sonucunda yüzme bilenlerin matematik dersinde daha başarılı oldukları gibi mantıksız bir sonuca varılabilir. Bu nedenle sonuçların mantık süzgecinden geçirilerek değerlendirilmesi gerekmektedir.

1.4 İSTATİSTİĞİN TARİHÇESİ

İstatistiksel çalışmaların başlangıcı çok eskilere dayanmaktadır. Ancak o dönemlerde yapılan çalışmalar sadece sayım düzeyinde kalmış bugünkü düzeye ulaşamamıştır. İstatistiksel çalışmalarla ilgili ilk bulgulara MÖ. 3000 yıllarında Mısır'da rastlanmıştır. Daha sonraları, Çin'de tarım istatistiklerinin düzenlendiği, Musa'nın 4

kitabında eli silah tutan Musevilerin sayıldığı, İncil'de geçmiş devrelere ait nüfus sayımlarının yapıldığı görülmektedir.

Gerçek anlamda istatistiksel çalışmalar, ancak 16. Yüzyıl da çeşitli matematikçilerin üzerinde çalıştığı olasılık teorisinin ortaya konulmasıyla başlamıştır. Olasılık teorisi ile ilgili çalışmaların çok erken başlamasının nedeni, insanoğlunun şans oyunlarına duyduğu büyük istek olmuştur. Şans oyunlarına meraklı olan kişiler, oyunlarda kazanma oranlarını yükseltmek için matematikçilerden kendilerine şans verecek çeşitli stratejileri bulmalarını istemişlerdi. Bu nedenle Pascal, Leibnitz, Fermat ve James Bernoulli gibi matematikçiler olasılık teorisi ile ilgili yaptıkları çalışmalar sonucunda çeşitli teoriler ortaya koymuşlardır. 18. Yüzyılda da bu çalışmalara devam edilmiş ve 1733 yılında De Moivre birçok istatistik konularının temeli olan Normal dağılım formülünü bulmuştur. Bu formül aynı zamanda "Gaussian" dağılım olarak da bilinmektedir. Çünkü Gauss'da (1777 - 1855) yaptığı çalışmalar sonucunda aynı sayıda ve aynı değerdeki ölçümlerin arasındaki hataları bularak bu formülü ispatlamıştır. Bu tarihlerde Laplace'de istatistiği astronomiye uygulayarak çeşitli çalışmalar yapmıştır. İstatistiğin farklı konulara uygulanması 19. Yüzyılda da devam etmiştir. Bu yüzyıl da Belçikalı istatistikçi olan Adolph Quetelet (1796-1874) istatistiği eğitim ve sosyal konulara, Sir Francis Galton (1822-1911) ise sosyal bilimlere uygulamıştır. Aynı yüzyılda matematikçi olan Karl Pearson (1857-1936) ile Galton birlikte çalışarak istatistiğin önemli konuları olan regresyon ve korelasyonla ilgili teoriler geliştirmişlerdir.

20. Yüzyılın başlarında ise William S. Gosset, küçük örneklerden yararlanılarak karar alma ile ilgili çalışmalarda bulunmuş arasındaki hataları bularak, formülü ispatlamıştır. Ve bu çalışmalar sonucunda elde ettiği bilgileri "Student" adli kitabında yayınlamıştır. Küçük örnekler teorisi ile ilgili bu çalışmalara Sir Ronald Fisher'de (1890-1962) devam etmiş, ve Fisher bu teori yanında, deneysel planlama ile de ilgilenmiştir. 20. Yüzyılda ise birçok değerli istatistikçiler yetişmiştir. Bu istatistikçilerin bazıları hala yeni teoriler geliştirerek ve çeşitli uygulamalar da bulunarak istatistiğin gelişmesine yardımcı olmaktadırlar. İstatistiğin gelişmesinde 1950'den sonra üniversitelerde belirli ve düzenli bir programa dayanılarak okutulmasının ve bilgisayarların çok etkisi olmuştur.

BÖLÜM 2

VERİLERİN TOPLANMASI VE DÜZENLENMESİ

Daha önce belirtildiği gibi istatistiksel araştırmaların yapılabilmesi için verilerin toplanması gerekmektedir. Araştırma yapılmaya karar verildiğinde ilk olarak gerekli olan veriler belirlenerek toplanacak ve toplanan bu veriler üzerinde düzenlemeler yapılarak analize hazır hale getirilecektir. İstatistiksel araştırmalardan sağlıklı sonuçlar alınabilmesi öncelikle sağlıklı ve yeterli verinin elde edilmesine bağlıdır. Bu nedenle, bu bölümde veri kaynakları ile toplanan verilerin düzenlenerek analize hazırlanması üzerinde durulacaktır.

2.1. VERİ VE VERİ KAYNAKLARI

İstatistiksel araştırmalar, araştırma yapılacak konu ile ilgili sayısal olan ve sayısal olarak ifade edilebilen bilgileri gerektirmektedir. Araştırma yapılabilmesi için gerekli sayısal bilgiler veri olarak adlandırılmaktadır. Veriler, araştırma yapılacak konu ile ilgili bilinen veya herhangi kaynaktan elde edilen bilgilerdir. Veriler alındıkları kaynağa göre doğrudan veya dolaylı veriler olmak üzere iki ana grupta toplanabilirler. Doğrudan veriler araştırmayı yapan kişi veya kişiler tarafından kaynağından alınan verilerdir. Doğrudan verilerin toplanmasında araştırmacılar çeşitli yöntemlerle veri kaynağına ulaşarak gerekli verileri elde ederler. Dolaylı veriler ise başka kurum veya kuruluşlar tarafından toplanarak düzenlenen verilerdir. Araştırmacılar bu kurum veya kuruluşların düzenledikleri bu verileri alarak kullanabilirler. Bu verilerin bir kısmı ise basılı olarak belirli dönemlerde yayınlanmaktadır. Örneğin, herhangi bir fakültedeki öğrencilerin başarı durumları ile ilgili bir araştırma yapıldığını varsayalım. Bu araştırma için gerekli olan veriler öğrencilerin kendilerinden alınırsa doğrudan veriler, fakülte yönetimine başvurularak onları daha önce düzenlediği veriler

alınırsa dolaylı veriler söz konusu olur. Diğer taraftan nüfus, tarım, fiyat gibi konularda Devlet İstatistik Enstitüsü'nün yayınlarından alınan verilerde dolaylı verilerdir.

Doğrudan veri toplanması durumunda araştırmacılar veri toplama öncesi ihtiyaçları olan veriyi belirlediklerinden daha kapsamlı veri elde edebilirler ve bunları istedikleri gibi düzenleyebilirler. Dolaylı veri alındığında ise, veriler başka bir kurum tarafından derlenip düzenlendiğinden istenen her verinin elde edilme imkanı olmayabilir. Ayrıca bu verilerin daha sonra tekrar düzenlenmesi mümkün olsa bile, bu aşamada bazı sorunlarla karşılaşılabilir. Örneğin, bir bölgede gelir dağılımı ile ilgili bir araştırma yapıldığını varsayalım. Bu araştırmadaki verileri doğrudan aldığımızda gelir sınıflarını istediğimiz gibi düzenleyebiliriz. Oysa bu veriler dolaylı olarak alınıyorsa, gelir sınıfları daha önce başkaları tarafından düzenlenmiş olacağından bu düzenlemeye uymak zorunluluğu olabilir.

2.2. TAM SAYIM VE ÖRNEKLEME

İstatistiksel araştırmalarda incelemeyi amaçladığımız birimlerin oluşturduğu topluluğa anakütle adı verilir. Birim ise, anakütleyi oluşturan en küçük parçadır. Örneğin, bir bölgede yaşayan ailelerin gelir dağılımları ile ilgili bir araştırma yapıyorsak, bölgede yaşayan tüm aileler anakütleyi, herbir aile ise, anakütlenin birimlerini oluşturur. Aynı şekilde herhangi bir işletmenin muhasebe kayıtları inceleniyorsa, tüm kayıtlar anakütle, herbir kayıt ise birimdir.

Anakütle ile ilgili bilgi toplamak istendiğinde tüm birimlerin teker teker incelenmesi gerekmektedir. Bu işleme tam sayım adı verilir. Anakütle birim sayısı fazla ise, tüm birimlerin incelenmesi fazla zaman alabilir ve masraflı olabilir. Araştırmalar genellikle bir fayda etmek için yapılırlar ve yapılan her araştırma belirli bir sürede bitirilemiyorsa ve araştırmada yapılan masraf faydayı aşıyorsa araştırma yapmanın anlamı kalmayacaktır. Bu nedenlerle örnekleme teorisi geliştirilmiştir.

Bir anakütleden, anakütle birim sayısından daha az sayıda birimin seçilerek, bu birimler yardımı ile anakütle parametrelerinin tahmin edilmesi işlemlerine örnekleme denir. Parametre ise anakütleleri birbirinden ayırmaya yarayan ortalama, varyans, oran gibi ölçülerin genel ifadesidir. Anakütleden seçilen az sayıda birimin oluşturduğu topluluk ise örnek olarak adlandırılır.

Örnek düzenlenmesi ve anakütle parametrelerinin tahmin edilmesi işlemlerini kapsayan örneklemenin tam sayıma göre bazı üstünlükleri vardır. Bu üstünlükleri şöyle sıralayabiliriz.

- Örneklemede daha az sayıda birim incelendiğinden daha kısa sürede ve daha az masrafla sonuç alınacaktır.
- Örnekleme ile az sayıda birim incelendiğinden verilerin toplanması ve düzenlenmesi sırasında yapılabilecek hatalar azalacaktır. Çok sayıda birimlerden oluşan, anakütleler ile ilgili araştırmalarda çok sayıda yetişmiş elemana ihtiyaç duyulmaktadır. Bu elemanların bulunmasında ve yerleştirilmesinde güçlüklerle karşılaşılabilir. Örnekleme ile bu sorunlarda azalmış olacaktır.

- Örnekleme ile daha geniş uygulama alanı sağlanabilir. Örneğin, 500 birim tam sayım ile incelenirse, sadece bu birimlerle ilgili bilgiler elde edilecektir. Oysa 500 birimlik örnek ile örnekleme oranına bağlı olarak, 500 birimden daha çok sayıda birimden oluşan anakütleler incelenerek, gerekli bilgiler elde edilecektir.

Örneklemenin bu yararları yanında başlıca sakıncası, anakütle parametrelerinin tahminleri yapıldığında örneklemeden doğan sapmaların ortaya çıkmasıdır. Ancak örneklemenin sağladığı çeşitli yararlara karşılık, araştırmacılar bu sapmalara katlanmak zorundadırlar.

Örnek birimlerinin seçim şekillerine göre örnekleme ikiye ayrılır. Birimler araştırmacıların isteklerine göre seçiliyorsa iradi örnekleme, etki altında kalınmadan tesadüfi olarak seçiliyorsa istatistiksel örnekleme söz konusudur. İkisi arasındaki en büyük fark örneklemeden kaynaklanan hataların istatistiksel örneklemede kontrol edilebilmesidir. İstatistiksel örneklemede belirlenen bir güven olasılığı ile örneklemeden kaynaklanacak sapmanın sınırları tahmin edilebilmektedir.

Başlıca istatistiksel örnekleme yöntemlerini şöyle sıralayabiliriz.

- **Basit Tesadüfi Örnekleme Yöntemi**

Örneğe seçilecek anakütle birimlerinin her birine eşit seçilme şansı veren örnekleme yöntemine basit tesadüfi örnekleme yöntemi denir. N birimden oluşan bir anakütleden n birimden oluşan bir örnek çekilirken, N>n olduğundan birbirinden farklı çok sayıda örnek düzenlenebilir. Düzenlenebilecek farklı örneklerin sayısı daha sonra verilecektir.

Anakütle birimleri örneğe seçilirken farklı seçim yöntemleri kullanılabilir. Birimlere birer numara verilerek numara yazılı kağıtlar bir torbadan çekilebilir, tesadüfi sayılar tablosundan yararlanılabilir, birimlere numara yerine harf verilebilir. Burada amaçlanan seçimin etki altında kalınmadan tesadüfi olarak yapılmasıdır. Basit tesadüfi örnekleme yönteminde anakütle ile ilgili herhangi bir düzenleme yapılmaz.

- **Tabakalı Örnekleme Yöntemi**

Zümrelere göre örnekleme yöntemi de denilen tabakalı örnekleme yönteminde, anakütle belirlenen kritere göre tabaka adı verilen alt gruplara ayrılır. Anakütlenin herbir birimi birden fazla tabakada yer alamaz ve tabakaların birim sayıları toplamı anakütle toplam birim sayısına eşittir. Tabakalardan örneğe çekilecek birimler tesadüfi çekilmeyebilir. Bu durumda iradi örnekleme sözkonusu olacağından, biz sadece tabakalardan birimlerin tesadüfi olarak çekilmesi durumunu inceleyeceğiz.

Tabakalı örnekleme yöntemi ile elde edilecek bilgiler, genellikle, anakütleyi tabakalara ayırmadan yapılan basit tesadüfi örnekleme yöntemine göre gerçeğe daha yakın sonuçlar verecektir. Tabakalı örnekleme yöntemi örnekleme hatalarının azaltılması bakımından faydalıdır.

Tabakalara ayırma bu yöntemin ilk aşamasıdır. Belirlenen kriterlere göre, değerleri birbirine yakın birimler aynı tabakalara dahil edilir ve tabakalar oluşturulduk-

tan sonra her tabaka için ayrı ayrı basit tesadüfi örnekleme yöntemi uygulanarak bunların sonuçları birleştirilir. Böylece anakütle varyansı küçülmektedir.

Tabakalı örnekleme yöntemine başvurulma nedeni sadece gerçeğe yakın sonuçlar elde edilmesi, varyansın küçültülmesi, örnekleme hatasının azaltılması değildir. Bazı durumlarda anakütle yanında tabakalarla ilgili bilgiler istenebilir. Tabakalı örnekleme yöntemi ile hem anakütle, hem de tabakalarla ilgili bilgiler elde edilmiş olur. Bazı durumlarda da bu yöntemle araştırmanın maliyeti düşebilir veya kolaylık olsun diye tabakalara ayırmaya gidilebilir.

- **Kademeli Örnekleme Yöntemi**

Kademeli örnekleme yönteminde anakütle alt gruplara ayrılır. Bu alt grupların her biri küme adını alır. Daha sonra bu kümelerden bir veya birkaç tanesi tesadüfi olarak seçilerek kümelerin tüm birimleri örneği oluşturur. Anakütlede yer alan farklı birimlerin kümelerde yer alması sağlanırsa kümeler anakütleyi daha iyi temsil edeceklerdir. Kümeler oluşturulurken buna dikkat edilirse kademeli örnekleme yöntemi basit tesadüfi örnekleme yönteminden daha iyi sonuç verebilir.

Bu yöntemde örnekleme maliyeti düşük olacağından, örnekleme maliyeti düşürülmek istendiğinde kademeli örnekleme yöntemi uygulanabilir.

Kademeli örnekleme yöntemi anakütlenin hiyerarşik yapıda olması durumunda önem kazanmaktadır. Hiyerarşik yapı anakütleyi oluşturan birimlerin doğal olarak, yapısal olarak, sıralı olmasını ifade etmektedir. Hiyerarşik yapı doğal kümelenme olarak da anılmaktadır. Üretilen mamüllerin birden fazlasının birlikte ambalajlanması, depoda duran malların raflara dizilmiş olması, muhasebe kayıtlarının deftere sıra ile kaydedilmiş olması durumları hiyerarşik yapıyı ifade etmektedir. Bu gibi durumlarda küme oluşturmaya gerek kalmadan kademeli örnekleme yöntemi uygulanabilir. Her kutu, her raf, defterin her sayfası birer küme olarak kabul edilebilir. Bu gibi durumlarda küme oluşturma maliyeti ve zaman kaybı olmayacaktır.

Kademeli örnekleme yöntemi, tek kademeli ve çok kademeli olarak uygulanabilir. Tek kademeli örnekleme yöntemi küme örneklemesi yöntemi olarak da adlandırılmaktadır.

Küme örneklemesinde, daha önce de belirttiğimiz gibi anakütle belirli bir faktöre göre kümelere ayrılır. Bu kümelerden bir veya birkaç tanesi tesadüfi olarak seçilerek örnek oluşturulur. Kümelerin tüm birimleri örnekte yer almaktadır.

Tek kademeli örnekleme ile örnek belirlendikten sonra bu örnekten tekrar örnek alınabilir. Bu durumda iki kademeli örnekleme söz konusu olacaktır. İki kademeli örneklemede M birim sayılı her bir kümeden m birim sayılı örnek seçilerek örnek oluşturulmaktadır. Aynı işlem üç kademede yapılırsa üç kademeli, daha çok kademede yapılırsa o sayıda kademeli örnekleme söz konusu olacaktır.

Çok kademeli örnekleme, tek kademeli örneklemeye göre daha fazla maliyeti gerektirir. Çok kademeli örnekleme, örnek birimleri hiyerarşik yapıda ise daha uygun bir yöntemdir.

İki kademeli örneklemenin ilk kademesinde kümeler basit tesadüfi örnekleme yöntemi ile seçilecektir. İkinci kademede ise, seçim herhangi bir örnekleme yöntemi ile yapılabilir. Aynı şey ikiden çok kademeli örnekleme için de geçerlidir.

- **Sistematik Örnekleme Yöntemi**

Sistematik örnekleme yöntemi, anakütle ve örnek birim sayıları ile orantılı eşit aralıklarla birimlerin çekilerek örneğin düzenlendiği bir yöntemdir. Anakütlenin birimleri 1'den N'e kadar numaralanır. k=N/n bize örnekleme aralığını verir. İlk k aralığında yer alan birimlerden birinden başlanarak eşit aralıklar ile birimler seçilerek örnek düzenlenmiş olur.

Sistematik örnekleme yöntemini ayrı olarak inceliyorsak da, aslında bu yöntem kademeli örnekleme yöntemidir. Numaralanan ve k aralıkta birimleri belirlenen seçilmesi olası her bir örnek bir küme olarak düşünülürse, sonuç olarak bu kümelerden biri seçilmektedir.

Bu yöntem özellikle anakütle birim sayısı çok olduğunda kullanılmaktadır. Anakütle tesadüfi şekilde düzenlenmişse veya birimler önem derecelerine göre sıralanmış ise, sistematik örnekleme yöntemi ile düzenlenecek örnek daha sağlıklı olacaktır. Fakat bu özellikler olmasa da sistematik örnekleme yöntemi kullanılabilir.

Sistematik örnekleme yöntemi basit tesadüfi örnekleme yöntemine göre bazı üstünlüklere sahiptir.

- Hata yapma olasılığı daha azdır.
- Daha kısa sürede sonuç alınabilir, işlemlerin daha hızlı yapılması mümkündür.
- Bu yöntem anakütleyi k aralıklarla tabakalara bölmekte ve her tabakadan bir birim alınmaktadır. Böylece sonuçlar basit tesadüfi örneklemeye göre daha güvenilir görünmektedir. Bu özelliği ve hızı ile bazen tabakalara göre örnekleme yönteminden bile daha güvenilir olabileceği düşünülebilir.

Sistematik örnekleme yöntemi ile N/n=k sayıda birbirinden farklı örnekler düzenlenebilir.

2.3. DOĞRUDAN VERİ TOPLAMA YÖNTEMLERİ

Önceki açıklamalarda veriler doğrudan ve dolaylı veriler olarak sınıflandırılmıştı. Doğrudan veriler araştırmacılar tarafından toplandıklarından, bu verilerin toplanmasında farklı yöntemler kullanılabilmektedir. Kullanılacak yöntem araştırmanın konusuna ve niteliklerine göre değişecektir. Burada en çok kullanılan üç yöntemden söz edilecektir.

2.3.1. Anket

Söz konusu yöntemler içinde en çok kullanılanıdır. Bu yöntem de anket formu olarak adlandırılan ve bilgi toplayacak kişi veya kişilerce hazırlanan bir form çeşitli şekillerde veri kaynağına ulaştırılır. Formun veri kaynağına ulaştırılmasında anketör

olarak adlandırılan kişiler, posta ve telefon kullanılabilir. Genellikle anketörler bu konuda yetiştirilmiş ve tecrübeli kişilerdir. Yetişmiş ve tecrübeli eleman bulunmadığında araştırmada görev alacak anketörler bir kursa tabi tutulabilirler. Anketörler bu konuda tecrübeli olsalar bile, gerekli olduğunda bir ön çalışma yapılabilir. Bu çalışma anketörlere ön bilgiler verilmesi açısından faydalı olabilir.

2.3.1.1. Anket Formunun Düzenlenmesinde Dikkat Edilmesi Gerekli Noktalar

Anket ile veri toplanmasında aşağıda özetlenen noktalara dikkat edilmesinin, daha sağlıklı veri toplamasına yardımcı olacağı, tecrübeler sonucunda belirlenmiştir. Önemli olarak görülen bu noktaları şöyle özetleyebiliriz.

- Anket formunu cevaplayanlar, verecekleri bilgilerin kendilerine zarar verecek şekilde kullanılmasından çekindikleri için, bilgi vermek istemeyebilirler veya doğru bilgi vermeyebilirler. Bu nedenle bilgilerin gizli kalacağının anket formunun başında belirtilmesi yararlı olabilir. Ayrıca isim, adres gibi anket formunun kimin tarafından cevaplandığını belirtecek bilgilerin formda yer almaması, daha doğru bilgilerin elde edilmesine yardımcı olabilir. Bu noktalara dikkat edildiğinde, anket formunu cevaplamaktan kaçınanların sayısının azaldığı ve daha doğru cevapların elde edildiği görülmüştür.
- Anket formunda yeralan soruların açık, anlaşılır, kısa ve basit olması gerekmektedir. Bu noktalara dikkat edilmemesi, sorular iyi anlaşılamayacağından, yanlış bilgilerin elde edilmesine neden olabilir.
- Anket formunda yer alan sorular mantıklı olmalı ve her sorunun bir cevabı bulunmalıdır. Mantıksız sorulara mantıklı cevapların alınması mümkün olmadığından soruların mantıklı olmasına özen gösterilmelidir. Ayrıca her sorunun bir cevabı olmaması, yanlış bilgi toplanmasına neden olabilir. Örneğin, bir bölgede yaşayan ev kadınlarına, bir malın markasını belirterek o malı kullanıp kullanmadıklarının sorulduğunu düşünelim. Bu durumda sadece evet veya hayır cevaplarından birini seçmeleri isteniyorsa ve soru yöneltilen ev hanımı bu malı kullanıp kullanmadığını hatırlamıyorsa, yanlış cevap verebilir. Bu yanlışlığı önlemek için bir üçüncü seçenek olarak hatırlamıyorum cevabının eklenmesi, yanlış bilgi toplanmasını engelleyecektir.
- Anket ile ilgili bilgi toplamada önemli noktalardan biri de soru sayısıdır. Soru sayısının az olması istenen bilginin elde edilmemesine, soru sayısının fazla olması ise cevap vereni bıktırabileceğinden doğru cevap alınamamasına veya formun sonunda yeralan bazı soruların cevapsız bırakılmasına neden olabilir.

2.3.1.2. Anket Formunun Bilgi Kaynağına Ulaştırılmasında Kullanılan Yöntemlerin Karşılaştırılması

Daha önceki açıklamalarda değinildiği gibi anket formu bilgi kaynağına anketörler posta ve telefon ile ulaştırılabilir. Bu yöntemlerin fayda ve zararlarını şöyle özetleyebiliriz.

- Anketörler İle Yapılan Anket Çalışmalarının Faydaları

- Anketörler ile yapılan çalışmalarda cevap alma oranının diğer yöntemlere göre daha yüksek olduğu görülmektedir. Çünkü, anketör bilgi verecek kişilere, soruları yüz yüze yönelttiğinden, cevap vermek istemeyen kişileri etkileyerek onların cevap vermesini sağlayabilecektir. Bu şekildeki çalışmalarda cevap alma oranının yüksekliği anketörün becerisine göre değişecektir.
- Anket ile bilgi toplamada anketörlerden yararlanmanın en büyük faydası soruların, cevaplayan kişiler tarafından iyi anlaşılmaması durumunda görülmektedir. Çünkü, soruların iyi anlaşılamaması durumunda anketörler cevap verenlere anlayamadıkları soruları açıklayarak daha doğru bilgi toplamasını sağlayabilirler.

- **Anketör İle Yapılan Anket Çalışmalarının Sakıncaları**

- Bir anket formu için anketörlere ödenecek ücret, posta ve telefona ödenecek ücretin çok üzerinden olduğundan, anketörler yardımı ile bilgi toplanması posta ve telefona göre daha masraflı olmaktadır.
- Yeterli sayıda anketör bulunamaması, bilgi kaynağına ulaşmanın zaman kaybına sebep olması gibi nedenlerden ötürü, anketörler kullanarak bilgi toplamak posta ve telefon ile bilgi toplamaktan genellikle daha çok zaman almaktadır. Bu da sonuçların posta ve telefondan daha uzun sürede alınmasına neden olacaktır.
- Anketörlerden yararlanmanın önemli sakıncalarından biri anketörün sonucu etkileyebilmesidir. Anketörler kişilere etkileyerek, gerçek dışı cevaplar vermelerine neden olabilirler. Anketörler tarafından kişilerin etkilenmesi hem olumlu, hem de olumsuz yönde olabilir. Anketörün cevap verecek kişiye kötü veya aşırı derecede iyi davranması, olumlu verilecek cevabın olumsuz veya olumsuz verilecek cevabın olumlu verilmesinde etkili olabilir.
- Anketörler ile bilgi toplamada yanlış sonuçlar alınmasında bir diğer etken, anketörlerin yetişmesidir. Her zaman yeterli sayıda ve yeterince bilgili anketörler bulunamayabilir. Anketörleri yetiştirmek için yapılan çalışmalarda her zaman istenen sonuca ulaşmamış olabilir. Bu durum toplanacak verilerin hatalı olmasına neden olacaktır.
- Anketörler ile çalışmanın en büyük sakıncası, anketörlerin gerçek olmayan sonuçları verebilmeleridir. Anketöre verilen formlar anketör tarafından gerçek dışı bilgilerle doldurulabilir. Bunu önlemek için genellikle çeşitli denetim mekanizmaları oluşturulmuştur. Fakat bu mekanizmalarda tam olarak gerçek dışı sonuçlar verilmesini önleyemeyebilir. Bu nedenle, bu konu üzerinde gerekli hassasiyetin gösterilmesi gerekecektir.

- **Posta ve Telefon İle Yapılan Anket Çalışmalarının Faydaları**

- Posta ve telefon ile yapılan anket çalışmalarında cevaplar anketör ile yapılan çalışmalara göre daha çabuk alınır. Posta ile yapılan çalışmalarda ise uygun bir tarih belirlenerek bu tarihe kadar gelen cevaplar işleme konabilir.
- Posta ve telefon ile yapılan çalışmalarda araştırma maliyeti düşecektir.
- Posta ve telefon ile yapılan çalışmalarda, anketörler kullanılmadığından, anketörlerin cevapları olumlu veya olumsuz yönde etkilemesi sakıncası or-

tadan kalkacaktır. Ancak telefon ile yapılan çalışmalar da kısmen bu sakınca sözkonusu olabilir.

- Posta ile yapılan çalışmaların en büyük faydası, cevaplamanın daha uygun bir zamanda ve daha uzun sürede yapılabilme imkanı yaratmasıdır. Böylece kişi anket formunu uygun bir zamanında cevaplayacağından daha doğru bilgiler elde edilebilir.

- **Posta ve Telefon İle Yapılan Anket Çalışmalarının Sakıncaları**

- Posta ve telefon gibi araçlarla yapılan çalışmalarda cevaplama oranı genellikle daha düşüktür ve anketörün kişiyi ikna ederek cevap alabilmesi imkanı da özellikle postada sözkonusu değildir. Telefon ile cevap istendiğinde kişilerin ikna edilebilme imkanı daha yüksek ise de telefonun yüze kapatılabileceği veya görüşmeden kaçmanın kolaylığı unutulmamalıdır. Posta ile cevap istendiğinde ise unutma, zaman ayıramama gibi nedenlerle cevap verme oranı düşebilmektedir.
- Özellikle posta ile yapılan çalışmalarda anlaşılamayan sorular ile ilgili açıklamalar yapılabilmesi imkanı olmadığından hatalı sonuçlar alınmasını engellemek için soruların anlaşılır olmasına daha çok dikkat edilmesi gerekmektedir.

2.3.2. Gözlem

Bu yöntemde bilgi toplayacak kişi veya kişiler bilgi kaynaklarını gözleyerek sonuçları elde etmektedirler. Gözlemde ankette olduğu gibi veri veya bilgi kaynağının etkilenmesi sözkonusu değildir. Çünkü bu yöntemde veri kaynağı gözlenerek sonuçlar kaydedilir ve böylece istenen bilgiler elde edilmiş olur. Örneğin, trafik ile ilgili olarak yapılan bir araştırmayı düşünelim. Çeşitli yolların belirlenen saatlerdeki trafik yükleri, sözkonusu saatlerde yolun uygun bir yerinde durularak araçların sayılması ile belirlenebilir.

Anket ve gözlem her zaman olmasa bile aynı bilgilerin elde edilebilmesi için kullanılabilirler. Burada yöntemin seçimi önce konunun ve araştırmanın özelliklerine, daha sonra araştırmayı yapan kişi veya kişilere bağlı olacaktır. Örneğin, tüketici tercihlerinin belirlenmesi için hem anket, hem de gözlem ile bilgi toplanabilir. Anket yapılacaksa, anket formunun hazırlanarak tüketicilere ulaştırılması gerekecek, oysa gözlem yapılacaksa tüketicilerin belirlenen alış veriş merkezlerinde davranışları veya tercihleri gözlenecektir.

2.3.3. Deney

Deney ile veri toplanmasında, anket ve gözlemden farklı olarak istenilen olaylar suni olarak oluşturulur ve oluşturulan bu olaylar gözlemlenerek, gerekli olan veriler elde edilir. Olayların yaratılmasında laboratuarlar veya farklı alanlar kullanılabilir. Örneğin, bir ilacın etkisi bir klinikte hastalara ilaç verilerek belirlenebilir. Daha önce söz edildiği gibi istatistik toplu olayları konu aldığından genellikle deneyin yeterli sayıda tekrarlanması gerekebilir. Olayın tekrar sayısının az olmasının hatalı sonuçların alınmasına neden olabileceği de unutulmamalıdır.

2.4. VERİLERİN DÜZENLENMESİ

Buraya kadar ki açıklamalarımızda veri, verilerin toplanması ve bununla ilgili başlıca yöntemlerden sözedildi. Veri hangi yöntemle elde edilirse edilsin, elimize ulaşan veriler ham veri olarak adlandırılırlar ve genellikle istatistiksel analize hazır değillerdir. Bu verilerin analize uygun hale getirilmeleri için düzenlenmeleri gerekmektedir.

Verilerin düzenlenmesi hem doğrudan hem de dolaylı veriler için yapılabilirse de, daha çok doğrudan veriler sözkonusu olduğunda önemlidir. Dolaylı veriler ise genellikle düzenlenmiş olarak elde edilirler, ancak yapılacak araştırmaya göre yeniden düzenlenmeleri gerekebilir.

2.4.1. Basit Seri

Elde edilen ham verinin küçükten büyüğe doğru sıralanması ile elde edilecek serilere basit seriler denir. Basit serilerin incelenecek birim sayısının az olması durumunda kullanılmaları daha uygun olmaktadır. Bir basit seri birden fazla birimden oluşur ve birim sayısı n ile gösterilir. Seride yer alan birimler ise X_i değişkeni ile ifade edilir. Burada i indisi X'in değişken olduğunu ve i'nin değişmesi ile farklı değerler alacağını ifade eder.

Basit serinin toplam değeri X_i'lerin toplamına eşit olacağından,

Serinin Toplam Değeri $= X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$ olacaktır.

ÖRNEK: Bir işletmede çalışan 25 işçiye verilecek çocuk paraları ile ilgili bir araştırma yapılmaktadır. İşçilerin çocuk sayıları aşağıda verilmiştir.

1,3,2,2,3,1,4,5,3,6,0,5,2,3,2,4,8,0,1,2,3,3,1,0,4

Verilen değerleri basit seri şeklinde düzenleyelim.

0,0,0,1,1,1,1,2,2,2,2,2,3,3,3,3,3,3,4,4,4,5,5,6,8

2.4.2. Frekans Serisi

İncelenecek birim sayısının artması ile basit seriler yan yana veya alt alta yazılmış uzun sayı dizileri oluşturulacaklardır. Bu durumda çalışma kolaylığı sağlanması için frekans serileri düzenlenir. Frekans serilerinin, değişkenin çok sayıda farklı değer almadığı, fakat aynı değerlerin çok sayıda tekrarlandığı durumlarda düzenlenmeleri daha uygundur.

Düzenlenmiş bir frekans serisi iki sütundan oluşmaktadır. Birinci sütunda değişkenin aldığı farklı değerler yer almaktadır. Değişkenin aldığı bu farklı değerler genellikle küçükten büyüğe doğru sıralanarak yazılırlar. Ancak gerekli ise, büyükten küçüğe doğruda sıralama yapılabilir.

Frekans serilerinin ikinci sütunu ise f_i ile gösterilir ve bu sütunda değişkenin aldığı farklı değerlerin tekrar sayıları yer almaktadır. Değişkenin aldığı farklı değer-

lerin tekrar sayıları frekans olarak adlandırılır. Bu nedenle f_i sütununa frekanslar sütunu adı verilmektedir. Frekanslar sütununun toplamı basit serinin birim sayısı, yani incelenen birim sayısına (n) eşittir.

Değişkenin k sayıda farklı değer aldığı bir frekans serisi şöyle gösterilebilir.

X_i	f_i
X_1	f_1
X_2	f_2
X_3	f_3
$\vdots$	$\vdots$
X_k	f_k

Frekans serilerinin toplam değeri f_i'ler ile X_i'lerin çarpımlarının toplamlarının alınması ile bulunur. f_1 ile X_1'in çarpımlarının toplamı X_1'ler toplamını; f_2 ile X_2'in çarpımlarının toplamı X_2'ler toplamını; .....; f_k ile X_k'nın çarpımlarının toplamı X_k'lar toplamını vereceğinden,

$$\text{Serinin Toplam Değeri} = f_1X_1 + f_2X_2 + \cdots + f_kX_k = \sum_{i=1}^{k} f_iX_i$$

olacaktır.

ÖRNEK: Daha önce basit seri olarak düzenlenen seriyi frekans serisi düzenleyiniz.

X_i	f_i
0	3
1	4
2	5
3	6
4	3
5	2
6	1
8	1
	25

Önce X_i'ler sütunun oluşturularak bu sütuna değişkenin aldığı farklı değerler, ikinci sütun olan f_i sütununa ise, değerlerin tekrarlanma sayıları yazılmıştır. Yani f_i sütununda 0 üç defa tekrarlandığından sıfırın karşısında 3; 1 dört defa tekrarlandığından 1'in karşısında 4; ..... yer almıştır. Görüldüğü gibi f_i sütunu toplamı basit serinin birim sayısı olan 25'e eşittir.

2.4.3. Sınıflı Seri

İncelenecek birim sayısının çok olması ve değişkenin birbirinden farklı çok sayıda değer alması durumunda frekans serileri de basit seriler gibi uzun sayı dizileri şekline dönüşürler. Bu durumda sınıflı seri, sınıflandırılmış seri, gruplu seri, gruplandırılmış seri gibi isimlerle adlandırılan seriler düzenlenmektedir.

Sınıflı seri adlandıracağımız seriler iki sütundan oluşmaktadır. Birinci sütun sınıflar sütunudur ve bu sütuna oluşturulan belirli sınıf aralıkları yazılır. İkinci sütun ise sınıflarda yer alan değerlerin tekrar sayılarının gösterildiği frekanslar sütunudur.

Sınıfların oluşturulmasında önemli olan nokta sınıf aralığının ne olması veya sınıf sayısının kaç tane olursa en uygun düzenlenmenin yapılmış olacağının belirlenmesidir. Sınıf sayısının ve sınıf aralığının belirlenmesi için bazı yazarlarca belirli ölçüler veya formüller önerilmektedir. Bu formüller ve ölçüler öneri niteliğinde olup kesin birer ölçü değillerdir. Sınıflama işlemi yapılırken dikkat edilmesi gereken, araştırmanın özellikleri ve sınıflama işleminin neden olduğu sapmalardır. Sınıf sayısının az olması sapmayı arttıracağından, sınıf sayısının dörtten az alınmasının doğru olmayacağı söylenebilir. Diğer taraftan sınıflama, değişkenin alacağı çok sayıdaki değerlerin azaltılarak, işlemleri kolaylaştırmaya yönelik olarak yapıldığından, sınıf sayısının sekizden çok olmasının işlemleri arttıracağı da dikkate alınmalıdır.

Sınıf aralığı araştırmanın özellikleri dikkate alınarak araştırmacı tarafından belirlendikten sonra, önce ilk sınıfın alt sınır değeri sınıflar sütununa yazılır. Daha sonra alt sınır değere sınıf aralığı eklenerek üst sınır değeri elde edilir. Her sınıfın üst sınır değeri, bir sonraki sınıfın alt sınır değerini oluşturur. Bu alt sınır değeri yazılıp bu değere sınıf aralığı eklenerek sınıflar düzenlenir. Bu işleme sınıflar değişkenin tüm değerlerini kapsayana kadar devam edilir.

Sınıflarda yer alan frekanslar belirlenirken, sınıfın alt sınır değeri dahil, üst sınır değeri ise hariç tutulur. Bir sınıfın üst sınır değeri ile onu takip eden sınıfın alt sınır değeri birbirine eşit olarak yazılır. Bu sınıfların frekansları yazılırken, değeri bir sınıfın üst sınır değerine eşit olan birim bir sonraki sınıfın frekansına eklenir. Örneğin 2'den başlayarak, 2 eşit aralıklı sınıflar oluşturuluyorsa sınıflar 2-4, 4-6,.... şeklinde oluşacaktır. 2-4 sınıfında 2 dahil 4'e kadar olan değerler bulunmakta 4 ise 4-6 sınıfında yer almaktadır. Yani, her sınıfta alt sınır dahil üst sınıra kadar olan değerler bulunmaktadır.

Sınıflı seriler ile işlemler yapılırken frekansların her sınıfta düzgün dağıldığı düşünülmektedir. Bu nedenle sınıf ortaları X_i olarak belirlenir ve seri ile ilgili hesaplamalar bu değerler dikkate alınarak frekans serilerinde olduğu gibi yapılır. Örneğin, sınıflı serilerin toplam değeri sınıf ortaları bulunduktan sonra frekans serilerinde olduğu gibi hesaplanır.

ÖRNEK: Daha önce frekans serisi olarak düzenlenen örneği, 2 eşit aralıklı sınıflı seri şeklinde düzenleyiniz.

Frekans Serisi	
X_i	f_i
0	3
1	4
2	5
3	6
4	3
5	2
6	1
8	1
	25

Sınıflı Seri	
Sınıflar	f_i
0-2	7
2-4	11
4-6	5
6-8	1
8-10	1
	25

Görüldüğü gibi seride 8'den büyük değer olmamasına rağmen, 6-8 sınıfında 6 dahil 8'e kadar olan değerler yeraldığından, 8 değeri 8-10 sınıfı oluşturularak bu sınıfa yazılmıştır.

Yukarıda verilen örnekte sınıf aralıkları eşit alınmıştır. Ancak uygulamada sınıf aralıklarının eşit olması şart değildir. Araştırmanın özelliklerine göre eşit olmayan aralıklı sınıflı serilerde düzenlenebilir.

Sınıfların oluşturulmasında açık sınıflı seriler de sözkonusu olabilmektedir. İlk sınıfın alt sınır değeri veya son sınıfının üst sınır değeri veya ikisi birlikte belli olmayan serilere açık sınıflı seriler denir. Bu tür düzenlemelere seride az sayıda aşırı derecede büyük veya küçük değer olduğunda başvurulmaktadır. Örneğin verdiğimiz örnekte 18 çocuklu bir işçi olsaydı, ikişer ikişer sınıfların düzenlenmesi yerine son sınıf 8-+ sınıfı şeklinde düzenlenebilirdi.

Sınıflı serilerde sınıfların frekanslarının sınıf aralığında düzgün dağıldığının kabul edilmesi, sınıflı serilerde hesaplanan ölçülerde küçük sapmalara neden olacaktır.

ÖRNEK: Bu sapma miktarını daha önce verilen örnek üzerinde gösteriniz.

Basit seri :

$$\Sigma X_i = 0+0+0+1+1+1+1+2+2+2+2+2+$$
$$3+3+3+3+3+3+4+4+4+5+5+6+8$$
$$=68$$

Frekans serisi:

X_i	f_i	$f_i X_i$
0	3	0
1	4	4
2	5	10
3	6	18
4	3	12
5	2	10
6	1	6
8	1	8
	25	$\sum f_i X_i = 68$

Sınıflı seri:

Sınıflar	f_i	$X_i = \frac{alts - üsts}{2}$	$f_i X_i$
0-2	7	(0+2)/2=1	7
2-4	11	(2+4)/2=3	33
4-6	5	(4+6)/2=5	25
6-8	1	(6+8)/2=7	7
8-10	1	(8+10)/2=9	9
	25		$\sum f_i X_i = 81$

Görüldüğü gibi basit ve frekans serilerinin toplam değeri 68'e eşit iken, sınıflı serinin toplam değeri 81'e eşittir.

Sınıflı serilerde sınıflardan kaynaklanan bu sapmalar, sınıflamanın sağladığı kolaylıklar nedeniyle dikkate alınmamaktadır.

2.5. BİRİKİMLİ FREKANSLAR

Bu bölüme kadar ki açıklamalarımızda frekans serileri ve sınıflı serilerin düzenlenmesine değinilmiştir. Ancak bazı istatistiksel araştırmalarda, bir frekans serisinde veya sınıflı seride belirli bir değerden daha küçük veya daha büyük değerler alan birim sayısının belirlenmesi istenebilir. Bu gibi durumlarda birikimli frekansların hesaplanması gerekir. Birikimli frekanslar seride belirli bir değerden daha küçük veya daha büyük değer alan birim sayısının belirlenmesi için kullanılabileceği gibi, daha sonraki bölümlerde açıklanacak bazı ölçülerin hesaplanmasına da yardımcı olacaklardır.

Birikimli frekanslar, artan birikimli frekanslar ve azalan birikimli frekanslar olmak üzere iki şekilde hesaplanırlar. Hesaplanan frekanslar düzenlenecek ayrı sütunlarda gösterilirler. Bu sütunlar artan birikimli frekanslar ve azalan birikimli frekanslar sütunları olarak adlandırılırlar.

2.5.1. Artan Birikimli Frekanslar

Frekans serilerinde ve sınıflı serilerde belirli bir değere eşit veya daha küçük değer alan birimlerin sayısını göstermek için artan birikimli frekanslar hesaplanır. Artan birikimli frekanslar hesaplanırken frekanslar sütununun ilk frekans değeri aynen alınır ve bu değer artan birikimli frekanslar sütununun ilk değeri olarak yazılır. İkinci frekans değeri ilk frekans değeri ile toplanarak artan birikimli frekanslar sütununun ikinci değeri olarak yazılır. Üçüncü frekans değeri ikinci frekans değeri ile toplanarak artan birikimli frekanslar sütununun üçüncü değeri olarak yazılır. Bu işleme frekanslar bitene kadar devam edilir. Artan birikimli frekanslar sütununun en son değeri frekanslar toplamına eşit olacaktır. Artan birikimli frekansların frekans serilerinde ve sınıflı serilerde ne ifade ettiğini iki ayrı örnek ile açıklayalım.

ÖRNEK: Daha önce frekans serisi olarak düzenlenen örneğin artan birikimli frekanslarını hesaplayınız.

X_i	f_i	Artan Birikimli Frekanslar
0	3	3=3
1	4	3+4=7
2	5	7+5=12
3	6	12+6=18
4	3	18+3=21
5	2	21+2=23
6	1	23+1=24
8	1	24+1=25
	25	

Görüldüğü gibi frekanslar sütunundaki frekansların toplanması ile artan birikimli frekanslar hesaplanmıştır. Artan birikimli frekanslar sütununda yer alan 18 sayısı, değişkenin 0,1,2 ve 3 değerlerinin frekanslarının toplanması ile oluşmuştur. Bu nedenle 18 sayısı seride 3 ve 3'den küçük kaç birim olduğu göstermektedir. Diğer bir ifade ile (X_i) değişkeni işçilerin çocuk sayılarını gösterdiğinden, işyerinde 3 ve 3'den az çocuklu işçilerin sayısının 18 olduğu anlaşılmaktadır. Artan birikimli frekanslar sütununun son değerinin frekanslar toplamı olan 25'e eşit olduğu görülmektedir.

ÖRNEK: Daha önce sınıflı seri olarak düzenlenen örneğin artan birikimli frekanslarını hesaplayınız.

Sınıflar	f_i	Artan Birikimli Frekanslar
0-2	7	7=7
2-4	11	7+11=18
4-6	5	18+5=23
6-8	1	23+1=24
8-10	1	24+1=25
	25	

Bu seride de artan birikimli frekanslar, frekans serilerinde olduğu gibi hesaplanmıştır. Artan birikimli frekanslar sütununda yer alan 23 sayısı 0-2, 2-4, 4-6 sınıflarının frekanslarının toplamından oluşmuştur. 4-6 sınıfında 4 dahil 6'dan küçük değer alan birimler yeraldığından 23, 6'dan küçük değer alan birimlerin sayısını göstermektedir. Artan birikimli frekanslar sütununun son değeri de yine frekanslar toplama eşittir.

2.5.2. Azalan Birikimli Frekanslar

Frekans serilerinde ve sınıflı serilerde, belirli bir değere eşit veya daha büyük değer alan birimlerin sayısını göstermek için azalan birikimli frekanslar hesaplanır. Azalan birikimli frekanslar hesaplanırken, artan birikimli frekansların hesaplanmasında ilk frekanstan başlanarak yukarıdan aşağıya doğru yapılan işlem, son frekans değerinden başlanarak aşağıdan yukarıya doğru yapılmaktadır.

Azalan birikimli frekansların hesaplanmasının bir diğer yolu, frekanslar toplamından başlanarak çıkartma işlemi yapılmasıdır. Bu durumda azalan frekanslar sütununa ilk olarak frekanslar toplamı yazılır. Daha sonra bu ilk frekans, frekanslar toplamından çıkartılarak bulunan değer azalan birikimli frekanslar sütununun ikinci değerini verir. Bu değerden ikinci frekans çıkarılarak bulunan değer, üçüncü azalan frekans değeridir. Bu işleme aynı şekilde serinin son değerine kadar devam edilir.

Artan birikimli frekanslar sütunun son değeri frekanslar toplamına eşit iken, azalan birikimli frekanslar sütununun ilk değeri frekanslar toplamına eşit olmaktadır.

Azalan birikimli frekansların frekans serilerinde ve sınıflı serilerden hesaplanmasını iki ayrı örnek ile açıklayalım.

ÖRNEK: Daha önce artan birikimli frekansları hesaplanan frekans serisinin azalan birikimli frekanslarını, açıklandığı gibi iki şekilde de hesaplayınız.

X_i	f_i	1.Yol Azalan Birikimli Frekanslar	2.Yol Azalan Birikimli Frekanslar
0	3	22+3=25	25
1	4	18+4=22	25-3=22
2	5	13+5=18	22-4=18
3	6	7+6=13	18-5=13
4	3	4+3=7	13-6=7
5	2	2+2=4	7-3=4
6	1	1+1=2	4-2=2
8	1	1	2-1=1
	25		

Azalan birikimli frekanslar sütununda yer alan 13 sayısı, değişkenin 3,4,5,6 ve 8 değerlerinin frekanslarının toplamından oluşmuştur. Bu nedenle seride 3 ve 3'den büyük değer olan birim sayısının 13 olduğu sonucuna varılmaktadır.

ÖRNEK: Daha önce artan birikimli frekansları hesaplanan frekans serisinin azalan birikimli frekanslarını hesaplayınız.

Sınıflar	f_i	Azalan Birikimli Frekanslar
0-2	7	25
2-4	11	18
4-6	5	7
6-8	1	2
8-10	1	1
	25	

Azalan birikimli frekans sütununda yer alan 7 sayısı, 4-6, 6-8 ve 8-10 sınıflarının frekanslarının toplamından oluşmuştur. 4-6 sınıfında 4 dahil 6'ya kadar değer alan birimler yer aldığından 7 sayısı serinin en az 4 veya 4'ten daha büyük değer alan birimlerin sayısını göstermektedir.

Artan ve azalan birikimli frekanslar ile ilgili olarak yaptığımız açıklamalara dikkat edilirse, artan birikimli frekanslar seride en az bir değere eşit veya o değerden daha küçük değer alan birimlerin toplam sayılarını verirken neden artan birikimli frekanslar olarak adlandırıldığı düşünülebilir. Aynı şey azalan birikimli frekanslar için de sözkonusu olacaktır. Çelişki gibi düşünülebilecek bu durum, birikimli frekanslara, artan veya azalan birikimli frekans sütunlarında yer alan frekansların sayılarının artması veya azalmasına göre ad verilmesinden kaynaklanmaktadır.

2.6. ORANSAL FREKANSLAR

Frekanslar, frekans serilerinde değişkenin aldıkları değerlerin tekrar sayılarını, diğer bir ifade ile seride herhangi bir değere eşit değer alan kaç birim olduğunu göstermektedirler. Eğer seri sınıflı seri ise, frekans sınıf aralığında yer alan birimlerin sayısını ifade etmektedir. Bazı araştırmalarda birimlerin sayıları yerine oranlarının hesaplanması gerekebilir. Bu gibi durumlarda oransal frekanslar hesaplanmaktadır.

Değişkenin aldığı değerlerin veya sınıfların frekanslarının, toplam frekansa oranlanması ile elde edilen frekanslara oransal frekanslar denir. Diğer bir ifade ile oransal frekanslar hesaplanırken frekanslar toplam frekansa bölünmektedir. Oransal frekanslar frekans serilerinde, değişkenin aldığı herhangi bir değere eşit değer alan birimlerin, toplam birimlere oranını göstermektedir. Sınıflı serilerde ise, herhangi bir sınıfta yer alan birimlerin oranını ifade etmektedir.

Sınıflı frekans serilerinde oransal frekansları örnek üzerinde açıklayalım.

ÖRNEK: Daha önce örnek olarak verdiğimiz frekans serisi ve sınıflı serinin oransal frekanslarını hesaplayalım.

X_i	f_i	Oransal Frekanslar
0	3	3/25=0,12
1	4	4/25=0,16
2	5	5/25=0,20
3	6	6/25=0,24
4	3	3/25=0,12
5	2	2/25=0,08
6	1	1/25=0,04
8	1	1/25=0,04
	25	25/25=1,00

Yukarıda oransal frekanslar sütununda yer alan 0,24 değeri serideki %24'ün değerinin 3'e eşit olduğunu, diğer bir ifade ile işyerinde çalışan işçilerin %24'ünün 3 çocuklu olduğunu göstermektedir.

Sınıflar	f_i	Oransal Frekanslar
0-2	7	7/25=0,28
2-4	11	11/25=0,44
4-6	5	5/25=0,20
6-8	1	1/25=0,04
8-10	1	1/25=0,04
	25	25/25=1,00

Yukarıda oransal frekanslar sütununda yer alan 0,20 değeri serideki birimlerin %20'sinin değerinin 4 dahil 6'dan az olduğunu, diğer bir ifade ile işyerinde çalışan işçilerin %20'sinin çocuk sayısının 4 (dahil) ile 6 (hariç) arasında olduğunu göstermektedir.

Oransal frekanslar artan ve azalan birikimli oransal frekanslar olarak da düzenlenebilirler. Artan birikimli frekanslar seride bir değere eşit veya daha küçük değer alan birimlerin toplam sayısını gösterirken, artan oransal birikimli frekanslar seride bir değere eşit veya daha küçük değer alan birimlerin oranını göstermektedir. Azalan birikimli frekanslar ise seride bir değere eşit veya daha büyük değer alan birimlerin toplam sayısını gösterdiğinden, azalan oransal birikimli frekanslarda seride bir değere eşit veya daha büyük değer alan birimlerin oranını gösterecektir.

Artan ve azalan oransal birikimli frekanslar, oransal frekansların artan veya azalan birikimli frekans şeklinde düzenlenmesiyle elde edilebileceği gibi bu birikimli frekansların toplam frekansa bölünmesi ile de elde edilebilirler. Artan oransal birikimli frekanslar sütununun son değeri ve azalan oransal birikimli frekanslar sütunun ilk değeri 1'e eşittir.

ÖRNEK: Daha önce verilen örnekler üzerinde artan ve azalan birikimli frekansları hesaplayınız.

X_i	f_i	Oransal Frekanslar	Artan Birikimli F.	Azalan Birikimli F.	Artan Oransal Birikimli F.	Azalan Oransal Birikimli F.
0	3	0,12	3	25	0,12	1,00
1	4	0,16	7	22	0,28	0,88
2	5	0,20	12	18	0,48	0,72
3	6	0,24	18	13	0,72	0,52
4	3	0,12	21	7	0,84	0,28
5	2	0,08	23	4	0,92	0,16
6	1	0,04	24	2	0,96	0,08
8	1	0,04	25	1	1,00	0,04
	25	1,00				

Yukarıda verilen frekanslı serinin artan oransal birikimli frekanslar sütununun da yer alan 0,72 değeri serideki birimlerin %72'sinin değerinin 3 ve 3'ten küçük olduğunu, diğer bir ifade ile bu işyerinde çalışan işçilerin %72'sinin çocuk sayısının 3 veya daha az olduğunu göstermektedir.

Bu serinin azalan oransal frekansları ise, serideki birimlerin 0,52'sinin değerinin 3 ve 3'den büyük olduğunu, yani işçilerin %52'sinin çocuk sayısının 3 ve 3'den fazla olduğunu göstermektedir.

ÖRNEK: Daha önce verilen sınıflı serinin artan ve azalan oransal birikimli frekanslarını hesaplayınız.

Sınıflar	f_i	Oransal Frekanslar	Artan Birikimli Frekanslar	Azalan Birikimli Frekanslar	Artan Oransal Birikimli F.	Azalan Oransal Birikimli F.
0-2	7	0,28	7	25	0,28	1,00
2-4	11	0,44	18	18	0,72	0,72
4-6	5	0,20	23	7	0,92	0,28
6-8	1	0,04	24	2	0,96	0,08
8-10	1	0,04	25	1	1,00	0,04
	25	1,00				

Yukarıda verilen sınıflı serinin artan oransal birikimli frekanslar sütununda yer alan 0,92 değeri serideki birimlerin %92'sinin değerinin 6'dan küçük olduğunu, diğer bir ifade ile işyerinde çalışan işçilerin %92'sinin çocuk sayısının 6'dan az olduğunu göstermektedir.

Bu serinin azalan oransal birikimli frekanslar sütununda yer alan 0,28 değeri ise serideki birikimlerin %28'in değerinin 4 ve 4'den büyük olduğunu yani işçilerin %28'inin çocuk sayısının 4 veya daha fazla olduğunu göstermektedir.

2.7. İSTATİSTİKSEL SERİLER

Toplanan verilerin özelliklerine göre sıralanması ile elde edilen sayı dizilerine istatistiksel seri adı verilmektedir. İstatistiksel seriler zaman, mekan, dağılım ve bileşik seriler olmak üzere dört ana grupta toplanabilirler.

2.7.1. Zaman Serileri

Toplanan verilerin gün, ay, yıl gibi zaman birimlerine göre düzenlenmesi ile oluşturulan serilere zaman serileri adı verilir. Örneğin, yıllara göre tütün üretim miktarlarını, aylara göre doğan çocuk sayılarını gösteren seriler birer zaman serisidir. Zaman serileri daha sonra zaman serileri analizi bölümünde açıklanacak, trend, mevsimlik dalgalanmalar, konjoktürel dalgalanmalar ve tesadüfi dalgalanmalar olarak adlandırılan dört faktör tarafından etkilenmektedirler. Bir zaman serisi iki sütundan oluşmakta ve birinci sütunda zaman birimleri, ikinci sütunda ise incelenen değişkene ait sayısal değerler yer almaktadır.

ÖRNEK:

Yıllar	Toplam İstihdam (1=bin kişi)
2005	21540,385
2006	21536,15
2007	21473,225
2008	21295,123
2009	21658,783

2.7.2. Mekan Serileri

Mekan serileri yersel seriler olarak da adlandırılır. Toplanan verilerin köy, il, bölge gibi mekan birimlerine göre düzenlenmesi ile elde edilen serilere mekan serileri denmektedir. Mekan serileri de zaman serileri gibi iki sütundan oluşmakta ve birinci sütunda mekan birimleri, ikinci sütunda ise mekan birimlerine ait sayısal değerler yer almaktadır. Örneğin, bölgelere göre pamuk üretim miktarlarını, illere göre ekmek fiyatlarını gösteren seriler mekan serileridir.

ÖRNEK: Kamu kesimi toplam gelir ve giderlerinin G.S.Y.İ.H. içindeki payları

	Gelir		Gider	
Ülkeler	2008	2009	2008	2009
Belçika	49,7	49,2	51,0	50,4
Danimarka	58,8	59,1	57,6	56,5
Almanya	44,8	44,8	47,5	47,2
Yunanistan	39,0	39,0	41,5	41,1
İspanya	40,8	40,8	42,9	42,4
Fransa	51,3	50,9	54,1	53,2
İngiltere	41,4	42,2	41,3	41,8
Türkiye	25,3	24.1	34,5	38,4

2.7.3. Dağılım Serileri

Dağılım serileri, değişkenlerin aldığı değerlerin dağılımını gösteren serilerdir. Bu serilerde, zaman ve mekan gibi bir karakter bulunmamakta sadece değişkenin aldığı değerler yer almaktadır. Daha önce açıkladığımız, basit seri, frekans serisi ve sınıflı seri dağılım serileridir.

Dağılım serilerini nicel ve nitel dağılım serileri olmak üzere ikiye ayırabiliriz. Nicel dağılım serileri niceliklere göre, nitel dağılım serileri ise niteliklere göre dağılımı gösterirler. Örneğin, nüfusun medeni hale göre dağılımını gösteren seriler nitel dağılım serileri, nüfusun yaşlara göre dağılımını gösteren seriler ise nicel dağılım serileridir.

Dağılım serileri birikimli seriler şeklinde de düzenlenebilirler. Daha önce birikimli frekanslar anlatılırken oluşturulan seriler birikimli serilerdir.

2.7.4. Bileşik Seriler

Birimlerin iki veya daha çok karaktere göre dağılımını gösteren serilere bileşik seriler denir. Örneğin, bir sınıftaki öğrencilerin matematik ve istatistik derslerinden aldıkları notların dağılımını veya ailelerin gelirlerinin ve tasarruflarının dağılımını gösteren seriler bileşik serilerdir.

ÖRNEK:

Aileler	Yıllık Gelir (1000 TL)	Yıllık Tasarruf (1000 TL)
A	12	1
B	32	3
C	56	18
D	24	10
E	18	5
F	27	8

Bileşik serilerde, dağılım serileri gibi frekans serisi ve sınıflı seri şeklinde düzenlenebilirler. Yukarıda verilen örnek sınıflı seri şeklinde şöyle düzenlenebilir.

Tasarruf (1000 TL) / Gelir (1000 TL)	0-5	5-10	10-15	15-20
0-15	1			
15-30		2	1	
30-45	1			
45-60				1

2.8. GRAFİKLER

Araştırma sonucunda elde edilen ve düzenlenen verilerin daha kolay anlaşılabilmesi için gösterildiği şekillere grafik denir. Grafikler göze hitap ettiğinden toplanan verilerin daha açık bir şekilde görülmesine ve daha kolay yorumlanmasına yardımcı olurlar. Konu ile ilgisi olmayan kişiler de grafikler yardımı ile konuyla ilgili fikir sahibi olabilirler. Ancak burada önemli olan nokta grafiklerin açık ve anlaşılır biçimde çizilmeleridir.

Farklı araştırma konuları, farklı veriler ve farklı amaçlar için çeşitli grafikler kullanılabilir. Diyagramlar, kartogramlar, stereogramlar, kutupsal grafikler, istatistiksel haritalar gibi çok çeşitli grafiklerden söz edilebilirse de, burada sadece daha sonraki analizlerde gerekli olacak grafiklere yer verilecektir.

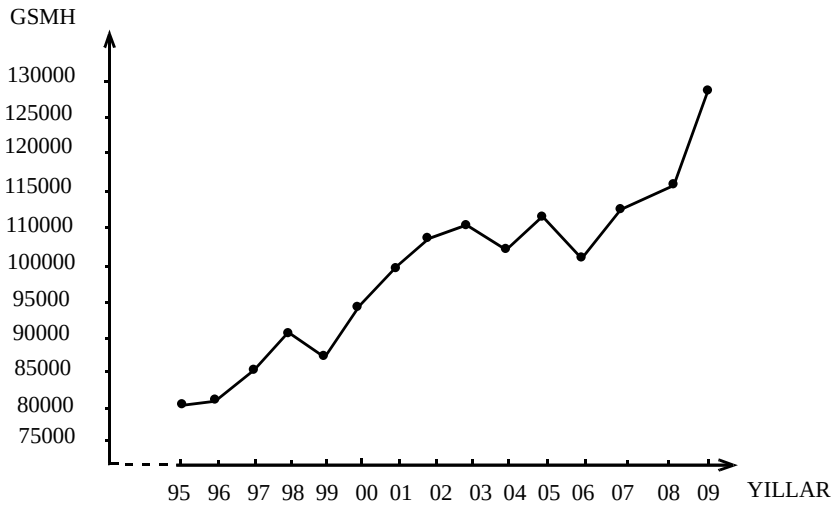
2.8.1. Zaman Serilerinin Grafikleri

Zaman serileri, değişkenin aldığı değerlerin zaman birimlerine göre dağılımını gösterdiklerinden, bunların grafikleri bir koordinat sistemi üzerine çizilebilir. Bu koordinat sistemin de yatay eksende zaman birimleri, düşey eksende ise değişkenin aldığı değerler yer almakta ve eksenler zaman birimleri ve değişkenin aldığı değerler dikkate alınarak ölçeklendirilmektedir. Bu grafikler zaman serilerinin analizi bölümünde kullanılacaktır.

ÖRNEK: Aşağıda verilen serinin grafiğini çiziniz.

Yıllar	Sabit Fiyatlar ile GSMH (1987) (1000 TL)
1995	80142,7
1996	80776,8
1997	85418,9
1998	91301,8
1999	87892,1
2000	94066,4
2001	100108,3
2002	107145,1
2003	110684,0
2004	105526,4
2005	112231,4
2006	104971,0
2007	112485,1
2008	117975,7
2009	127219,4

Verilen serinin grafiği Şekil.2.1'de görülmektedir.



Şekil 2.1

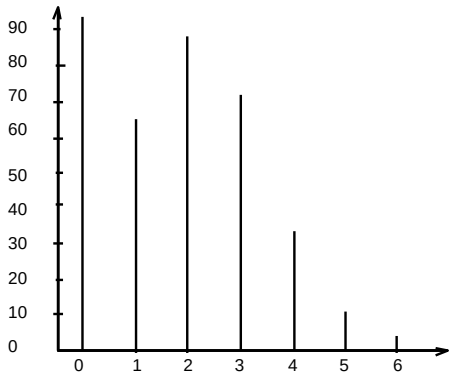
2.8.2. Dağılım Serilerinin Grafikleri

Değişkenlerin aldıkları değerlerin dağılımını gösteren ve basit seri, frekans serisi ve sınıflı seri şeklinde düzenlenen bu dağılım serilerinin grafikleri de serilerin türlerine göre çizilmektedir.

Basit serilerde birim sayısı az olduğundan ve frekanslar bulunmadığından bu serilerin grafikleri, değişkenin aldığı değerlerin büyüklüklerini gösterecek şekilde çizilmektedir. Frekans serilerinin grafikleri ise değişkenin aldığı değerlere göre frekansların dağılımını göstermektedir. Frekans serilerinin grafikleri bir koordinat sistemi üzerine çizilebilir. Grafik çiziminde yatay eksende değişkenin aldığı değerler, düşey eksende ise frekanslar yer almaktadır. Yatay ve düşey eksenlerin ölçeklendirilmesi, değişkenin aldığı değerlerin ve frekansların büyüklükleri dikkate alınarak yapılmakta ve bu ölçeklendirme işleminden sonra, değişkenin aldığı değerlerin karşılığı olan frekanslar eksen sistemi üzerinde işaretlenmektedir. Bu serilerde değişken süreksiz değişken olduğundan noktaların birleştirilmesi doğru olmayacaktır. Ancak elde edilen noktalar birer dikey çizgi ile yatay eksenle birleştirilebilir.

ÖRNEK: Yapılan bir anket sonucunda sahip olunan çocuk sayısına göre aile sayısının dağılımı aşağıda verilmiştir. Bu serinin grafiğini çiziniz.

Çocuk Sayısı(X_i)	Aile Sayısı (f_i)
0	94
1	65
2	87
3	71
4	34
5	12
6	5
	368

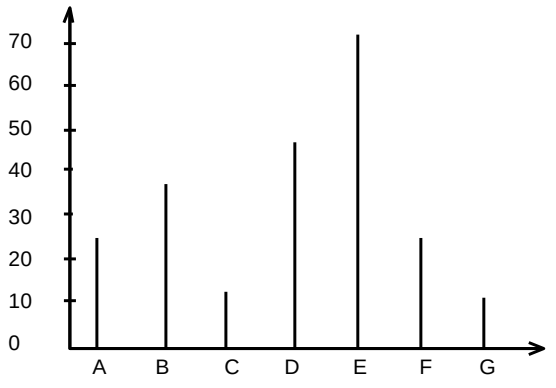


Şekil 2.2

Verilen bu serinin grafiği şekil 2.2.'de görüldüğü gibidir. Şekil 2.2.'de nicel dağılım serisinin grafiği görülmektedir. Serinin nitel dağılım serisi olması durumunda grafikteki yatay eksende nitelikler yer alacaktır.

ÖRNEK: Bölgelere göre televizyon sayılarının dağılımı aşağıda verilmiştir. Bu dağılımın grafiğini çiziniz.

Bölgeler	Televizyon Sayısı (1=10.000)
A	25
B	36
C	12
D	49
E	71
F	28
G	12
	233



Şekil 2.3

Serinin grafiği şekil 2.3.'de görülmektedir.

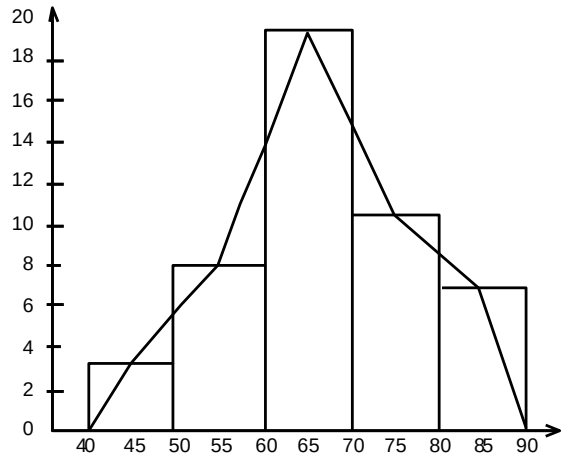
Frekans serilerinde olduğu gibi sınıflı serilerin grafikleri de bir koordinat sistemi üzerine çizilebilir ve çizilen bu frekans serilerinin grafiklerine histogram adı verilir. Histogramın çiziminde yatay eksende sınıf aralıkları, düşey eksende ise frekanslar yer almaktadır. Sınıf aralıkları ve frekansların değerleri bu eksenlerde belirlendikten sonra, sınıf aralıklarının alt ve üst sınırlarından frekans değerlerine kadar birer dikme çizilir ve sınıflı serilerde sınıfların frekanslarının sınıf aralığında düzgün dağıldığı kabul edildiğinden, çizilen dikmeler yatay eksene paralel bir çizgi ile birleştirilerek dikdörtgenler elde edilir. Bu dikdörtgenlerin tümü bize histogramı verir.

Histogramı oluşturan bu dikdörtgenlerin üst orta noktalarının birleştirilmesi ile elde edilen grafiğe de frekans poligonu adı verilir. Frekans poligonu sınıf aralıklarının mümkün olduğunca küçültülmesi durumunda bir eğriye dönüşür, bu eğriye de frekans eğrisi adı verilir.

ÖRNEK: Bir sınıftaki 50 öğrencinin ağırlıklarına göre dağılımı aşağıda verilmiştir. Bu serinin histogramı ve frekans poligonunu çiziniz.

Sınıflar	f_i
40-50	5
50-60	8
60-70	19
70-80	11
80-90	7
	50

Serinin histogramı ve frekans poligonu Şekil 2.4.'de görülmektedir.



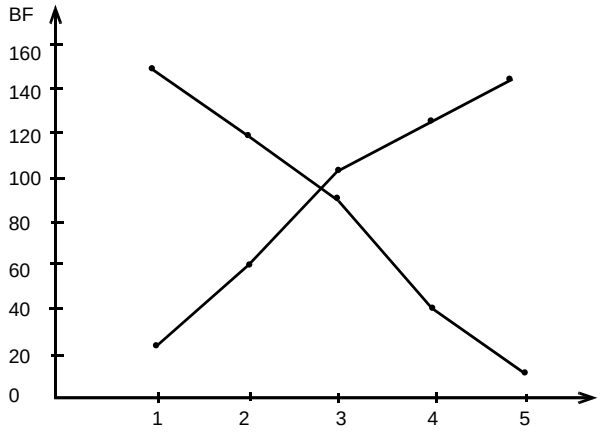
Şekil 2.4

Daha önce anlatılan artan ve azalan birikimli frekans serilerinin grafikleri de koordinat sistemi üzerine çizilebilir. Grafiklerin çiziminde yatay eksende değişkenin değerleri, düşey eksende ise birikimli frekanslar yeralır. Değişkenin aldığı değerler ile birikimli frekansların kesiştiği noktaların birleştirilmesiyle birikimli serilerin grafiği çizilmiş olur.

ÖRNEK: Bir bölgedeki binaların kat sayısına göre dağılımı aşağıda verilmiştir.

X_i	f_i	Artan Birikimli Frekanslar	Azalan Birikimli Frekanslar
1	26	26	145
2	34	60	119
3	43	103	85
4	29	132	42
5	13	145	13

Yukarıda verilen serinin artan ve azalan birikimli frekanslarının grafiğini çiziniz.

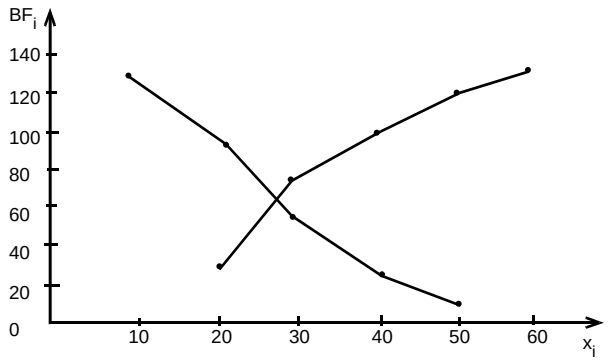


Şekil 2.5

Verilen serinin artan ve azalan birikimli frekansları Şekil 2.5.'de görülmektedir.

ÖRNEK: Dayanıklı tüketim malları satan bir mağazadaki bir aylık kredili satış miktarlarının kişilere göre dağılımı aşağıda verilmiştir. Serinin artan ve azalan birikimli frekanslarının grafiğini çiziniz.

Sınıflar (1=1000 TL)	f_i	Artan Birikimli Frekanslar	Azalan Birikimli Frekanslar
10-20	32	32	128
20-30	44	76	96
30-40	29	105	52
40-50	15	120	23
50-60	8	128	8



Şekil 2.6

Verilen serinin artan ve azalan birikimli frekanslarının grafiği Şekil 2.6.'da görüldüğü gibidir.

Örneklerde görüldüğü gibi frekans serileri için birikimli frekansların grafikleri çizilirken birikimli frekanslar ve bu frekansların ait olduğu X değerleri alınmaktadır. Sınıflı serilerde ise durum biraz farklıdır. Bu serilerin artan birikimli frekansların grafikleri çizilirken sınıfın üst sınır değeri, azalan birikimli frekanslar çizilirken ise alt sınır değeri alınmaktadır.

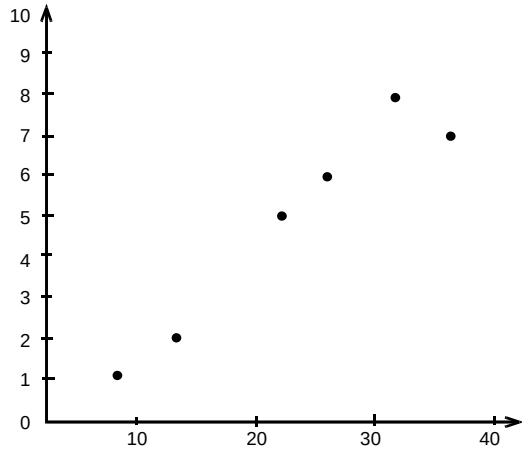
2.8.3. Bileşik Serilerin Grafikleri

Birimlerin birden fazla karaktere göre dağılmalarını gösteren serilere bileşik seriler denilmekteydi. Bileşik serilerin grafikleri, iki karakter sözkonusu olduğunda, bir karakter yatay eksende, diğer karakter düşey eksende gösterilerek kolaylıkla çizilebilir. Ancak karakter sayısı ikiden fazla ise çizim güçleşmektedir.

ÖRNEK: Aşağıdaki aynı iş kolunda faaliyet gösteren 6 işletmenin yıllık satış miktarları ve karları verilmiştir. Verilen serinin grafiğini çiziniz.

İşletmeler	Satış Miktarı (1=10000 TL)	Kar Miktarı (1=10000 TL)
A	22	5
B	38	7
C	14	2
D	8	1
E	33	8
F	27	6

Yukarıda verilen serinin grafiği Şekil 2.7.'de görüldüğü gibidir.



Şekil 2.7

2.9. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir işletmenin pazarlama bölümünde görev almak için işletmeye başvuran 50 kişinin yaşları aşağıda verilmiştir.

26,25,29,24,27,20,23,26,23,30,27,26,26,24,26,24,23,24,26,25,18,24,21,24,27,
25,24,24,26,25,23,25,25,27,27,35,26,25,25,24,23,26,26,33,27,25,24,24,21,29.

Verilen yaşları,

a- Basit seri

b- Frekans serisi

c- En küçük değerden başlayarak üçer eşit sınıf aralıklı sınıflı seri şeklinde düzenleyiniz.

d- Frekans ve sınıflı serilerin toplam değerlerini hesaplayınız.

Çözüm :

a- 18,20,21,21,23,23,23,23,23,24,24,24,24,24,24,24,24,24, 24,
24, 25,25,25,25,25,25,25,25,25,26,26,26,26,26,26,26,26,26,
26,27,27,27,27,27,27,29,29,30,33,35.

b-

Xi	fi
18	1
20	1
21	2
23	5
24	11
25	9
26	10

27	6
29	2
30	1
33	1
35	1
	50

c-

Sınıflar	fi
18-21	2
21-24	7
24-27	30
27-30	8
30-33	1
33-36	2
	50

d-

Xi	fi	fiXi
18	1	18
20	1	20
21	2	42
23	5	115
24	11	264
25	9	225
26	10	260
27	6	162
29	2	58
30	1	30
33	1	33
35	1	35
	50	1262

e-

Sınıflar	fi	Xi	fiXi
18-21	2	19,5	39
21-24	7	22,5	157,5
24-27	30	25,5	765
27-30	8	28,5	228
30-33	1	31,5	31,5
33-36	2	34,5	69
	50		1290

ÖRNEK 2: Bir bölgedeki 180 ev dolaşılarak anket yapılmıştır. Ankete katılanların "Evde kaç kişi yaşıyorsunuz?" sorusuna verdiği cevapların dağılımı aşağıda verilmiştir.

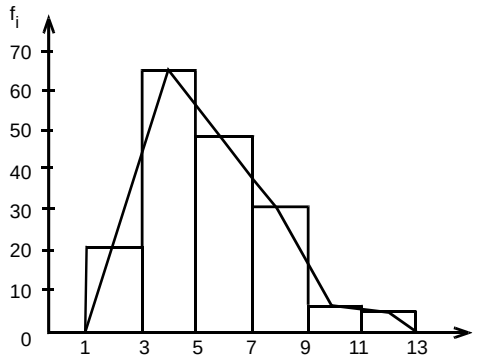
Evde yaşayan kişi Sayısı(Xi)	fi
1	6
2	18
3	24
4	41
5	33
6	15
7	19
8	11
10	7
11	5
12	1
	180

a- Verilen değerleri iki eşit aralıklı sınıflı seri şeklinde düzenleyiniz.
b- Düzenlediğiniz sınıflı serinin histogramını ve frekans eğrisini çiziniz.

Çözüm:

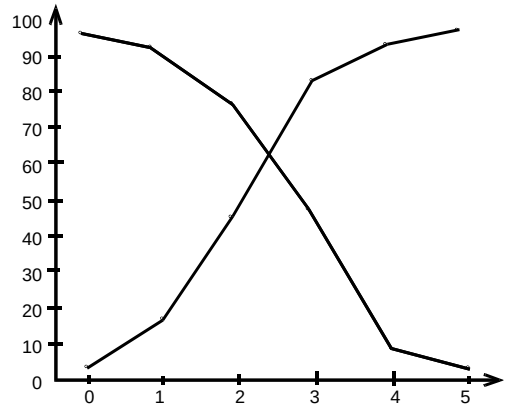
a-

Sınıflar	fi
1-3	24
3-5	65
5-7	48
7-9	30
9-11	7
11-13	6
	180



ÖRNEK 3: İşe alınmak için bir işletmeye 95 kişi müracaat etmiş ve adaylara 5 hukuk sorusu sorulmuştur. Adayların verilen doğru cevap sayılarına göre dağılımı aşağıda verilmiştir. Dağılımın artan ve azalan birikimli frekanslarını hesaplayarak, grafiklerini çiziniz.

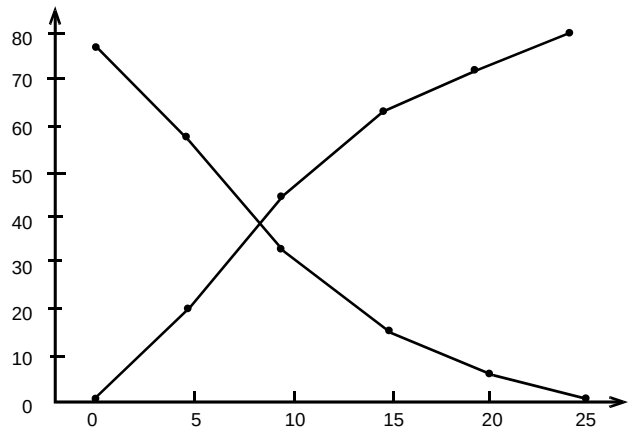
Xi	fi	Artan Birikimli Frekanslar	Azalan Birikimli Frekanslar
0	2	2	95
1	14	16	93
2	30	46	79
3	37	83	49
4	8	91	12
5	4	95	4
	95		



ÖRNEK 4: Aynı iş kolunda faaliyet gösteren ve ihracat yapan 78 işletmenin yıllık ihracat miktarlarına göre dağılımı aşağıda verilmiştir. Dağılımın artan ve azalan birikimli frekanslarını hesaplayarak grafiklerini çiziniz. (1=1 milyon dolar)

Sınıflar	fi	Artan Birikimli Frekanslar	Azalan Birikimli Frekanslar
0-5	19	19	78
5-10	27	46	59
10-15	15	61	32
15-20	10	71	17
20-25	7	78	7
	78		

Çözüm:



ÖRNEK 5: Bir işletmenin 60 günlük otomobil satışlarının dağılımı aşağıda verilmiştir.

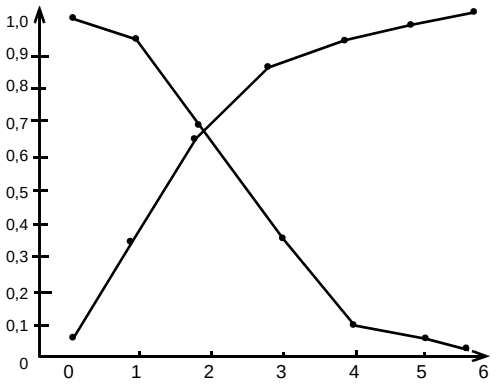
Xi	fi	Oransal Frekanslar	Artan Oransal Birikimli F.	Azalan Oransal Birikimli F.
0	3	0,050	0,050	1,000
1	16	0,267	0,317	0,950
2	21	0,350	0,667	0,683
3	12	0,200	0,867	0,333
4	5	0,083	0,950	0,133
5	2	0,033	0,983	0,050
6	1	0,017	1,000	0,017
	60	1,000		

Günlük otomobil satışlarının dağılımının,

a- Oransal frekanslarını

b- Artan ve azalan oransal birikimli frekanslarını hesaplayınız.

c- Artan ve azalan oransal birikimli frekans grafiğini çiziniz.



ÖRNEK 6: Bir bölgede yaşayan ailelerin yıllık gelirlerine göre dağılımı aşağıda verilmiştir. (1=1 milyon TL)

Sınıflar	fi	Oransal Frekanslar	Artan Oransal Birikimli F.	Azalan Oransal Birikimli F.
5-15	171	0,122	0,122	1,000
15-25	422	0,301	0,423	0,878
25-35	384	0,274	0,697	0,577
35-45	210	0,150	0,847	0,303
45-55	121	0,086	0,933	0,153
55-65	92	0,067	1,000	0,067
	1400	1,000		

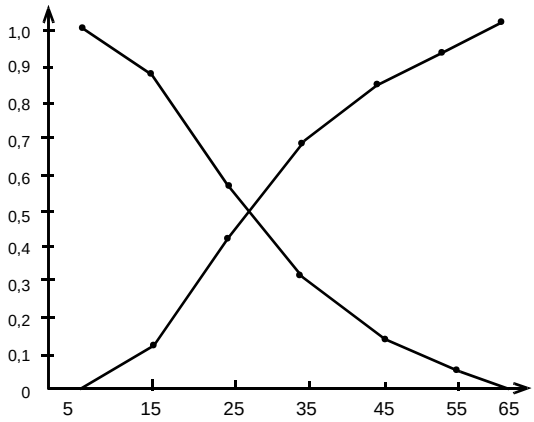
Gelir dağılımının,

a-Oransal frekanslarını

b-Artan ve azalan oransal birikimli frekanslarını hesaplayınız.

c-Artan ve azalan oransal birikimli frekans grafiğini çiziniz.

Çözüm:



ÖRNEK 7: Bir işletmede çalışan 24 kişinin kıdemleri ve aylık net ücretleri aşağıda verilmiştir. Kıdem belirlenirken tamamlanan yıl esas alınmıştır.

Aylık Net Ücret (1=10 TL)	Kıdem (Yıl)
36	1
52	3
81	7
12	0
15	1
23	0
28	1
29	2
35	2
91	5
97	6
63	8
61	4
55	5
53	4
58	6
42	3
41	3
40	0
73	11
34	3
35	2
51	4
58	5

Bu verileri kıdemi 3 yıl eşit aralıklı, ücreti 20 Türk lirası eşit aralıklı olarak, sınıflı bileşik seri şeklinde düzenleyiniz.

Çözüm:

	Kıdem Sınıfları				
Ücret Sınıfları	0-3	3-6	6-9	9-12	TOPLAM
0-20	2				2
20-40	6	1			7
40-60	1	7	1		9
60-80		1	1	1	3
80-100		1	2		3
TOPLAM	9	10	4	1	24

BÖLÜM 3

ORTALAMALAR

Elde edilen ham verinin seri şeklinde düzenlenmesi, düzenlenen verinin grafiklerle gösterilmesi, konu ile ilgili bir fikir verebilir. Fakat bu serilerle ilgili daha kesin değerlendirmeler yapılabilmesi, farklı serilerin birbirleri ile karşılaştırılabilmesi için bazı ölçülerin hesaplanması gerekmektedir. Bu bölümde açıklanacak olan ve merkezi eğilim ölçüleri olarak da adlandırılan ortalamalar, istatistiksel analizlerin temel ölçüleri olarak kabul edilebilirler.

Ortalamalar, genel olarak iki temel amaç için hesaplanan ölçülerdir. Ortalamaların hesaplanmasının birinci amacı, seriyi temsil etme özelliğini taşıyan tek bir sayısal değerin belirlenmesidir. Ortalama olarak öyle bir sayısal değer hesaplanır ki, bu sayısal değer serinin bir nevi özetidir ve bu değere bakılarak serinin genel karakteri hakkında fikir sahibi olunur. Ortalamaların hesaplanmasının diğer amacı, farklı serilerin birbirleri ile karşılaştırılmalarını sağlamaktır. Seriler ortalamaları hesaplanarak karşılaştırılabilirler. Örneğin, bir sınıftaki öğrencilerin istatistik dersinden aldıkları notların ortalaması hesaplanarak öğrencilerin istatistik dersindeki başarıları hakkında fikir sahibi olunabilir. Aynı ortalama matematik, iktisat gibi farklı dersler için de hesaplanırsa, hesaplanan bu ortalamalar karşılaştırılarak, öğrencilerin hangi derslerde daha başarılı veya hangi derslerde daha başarısız oldukları belirlenebilir.

Ortalamaları, hesaplanmalarında kullanılan birimlere göre iki ana gruba ayırabiliriz. İlk grupta, aritmetik ortalama, kareli ortalama, geometrik ortalama ve harmonik ortalama yer almakta ve bu ortalamaların hesaplanmasında serinin bütün birimleri kullanılmaktadır. İkinci grupta ise mod ve medyan yer almakta ve bunların hesaplanmasında serinin bütün birimleri kullanılmamaktadır. İki grup arasındaki en büyük fark, seri birimlerindeki değişmelere karşı olan duyarlılıklarıdır. Hesaplanmasında serinin bütün birimleri kullanılan ortalamaların değeri, serinin her hangi bir biriminin değeri değiştiğinde değişecektir. Oysa diğer gruptaki ortalamaların değerinin değişmesi için bu ortalamaların hesabında kullanılan birimlerin değerinin değişmesi gerekmektedir. Hesaplanmasında

kullanılan birimlerin değeri değişmedikçe bu ortalamaların değerleri de değişmez. Ancak hangi ortalama hesaplanırsa hesaplansın, bu ortalamanın değeri serinin maksimum ve minimum değerleri arasında yer alacaktır.

Bu bölümde sırasıyla aritmetik ortalama, kareli ortalama, geometrik ortalama, harmonik ortalama, mod ve medyan anlatılacaktır.

3.1. ARİTMETİK ORTALAMA

Serinin toplam değerinin seri birim sayısına bölümüne aritmetik ortalama denilmektedir. Aritmetik ortalama, seride değişkeni ifade eden harfin üzerine çizgi (-) konularak $\overline{X}, \overline{Y}, \overline{Z}, \cdots$ gibi sembolize edilmektedir.

Daha önce açıklandığı gibi basit serilerin toplam değeri, değişkenin aldığı değerlerin toplamına eşittir. Bu toplamın seri birim sayısı olan n'e bölümü basit seriler için aritmetik ortalamayı verecektir. Aritmetik ortalama,

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

formülü ile hesaplanır.

ÖRNEK: X_i : 2; 3; 4; 5; 8; 11

serisinin aritmetik ortalamasını hesaplayınız.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{2+3+4+5+8+11}{6} = \frac{33}{6} = 5{,}5$$

Frekans serilerinde ise, serinin toplam değeri değişkenin değerleri ile bu değerlerin frekanslarının çarpımlarının toplanması ile bulunacaktır. Serinin toplam değeri genellikle açılan $f_i X_i$ sütununun toplamının alınması ile hesaplanmaktadır. Bu toplamın seri birim sayısı n'e eşit olan frekanslar toplamına bölümü ise serinin aritmetik ortalamasına eşit olacaktır. Frekans serilerinde aritmetik ortalama

$$\overline{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \cdots + f_k X_k}{f_1 + f_2 + f_3 + \cdots + f_k} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}$$

formülü ile hesaplanır.

ÖRNEK: Aşağıdaki verilen serinin aritmetik ortalamasını hesaplayınız.

X_i	f_i	f_iX_i
1	3	3
2	8	16
3	15	45
4	10	40
5	4	20
	40	124

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{124}{40} = 3{,}1$$

Aritmetik ortalamanın hesabında sınıflı serilerin kullanıldığı durumlarda, serilerin toplam değeri hesaplanırken, frekans serilerindeki X_i'ler yerine sınıf ortaları kullanılmaktadır. Bu nedenle sınıflı serilerde önce sınıf ortaları bulunmakta daha sonra serinin aritmetik ortalaması frekans serilerinin aritmetik ortalamasının hesaplandığı

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}$$

formülü ile hesaplamaktadır.

ÖRNEK: Aşağıdaki serinin aritmetik ortalamasını hesaplayınız.

Sınıflar	X_i	f_i	f_iX_i
0-4	2	5	10
4-8	6	9	54
8-12	10	16	160
12-16	14	13	182
16-20	18	7	126
		50	532

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{532}{50} = 10{,}64$$

3.1.1. Aritmetik Ortalamanın Özellikleri

Basit, frekanslı ve sınıflı serilerde hesaplanan aritmetik ortalamanın birçok özellikleri vardır. Bu özellikleri örnekler ile açıklayalım.

- Bir serinin aritmetik ortalamasının seri birim sayısı ile çarpımı, serinin toplam değerini verir.

Basit serilerde aritmetik ortalama,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

formülü ile hesaplanmaktadır. İçler dışlar çarpımı yapılırsa,

$$\sum_{i=1}^{k} X_i = n\overline{X}$$

elde edilir. Frekans serileri ve sınıflı serilerde ise aritmetik ortalama;

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}$$

formülü ile hesaplanıyordu. İçler dışlar çarpımı yapılırsa,

$$\sum_{i=1}^{k} f_i X_i = \overline{X} \sum_{i=1}^{k} f_i$$

bulunur.

ÖRNEK: Aritmetik ortalaması 8 olan 25 birimli bir basit serinin toplam değerini hesaplayınız.

$$\text{Toplam değer} = \sum_{i=1}^{k} X_i = n\overline{X}$$

$$= 8(25) = 200$$

- Seri değerlerinin aritmetik ortalamadan farklarının toplamı sıfırdır.

Basit seriler için,

$$\sum_{i=1}^{n} (X_i - \overline{X}) = 0$$

olarak gösterilir. Bu ifadeyi açarsak,

$$\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \overline{X} = \sum X_i - n\overline{X}$$

ve $\sum X_i = n\overline{X}$ olduğundan

$$\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - \sum X_i = 0$$

olarak elde edilir.

Frekans serileri ve sınıflı seriler için

$$\sum_{i=1}^{k} f_i(X_i - \overline{X}) = 0$$

olarak gösterilir.

$$\sum f_i X_i = \overline{X} \sum f_i$$

olduğundan parantezi açarsak

$$\sum_{i=1}^{n} f_i(X_i - \overline{X}) = \sum f_i X_i - \sum f_i \overline{X}$$

$$= \sum f_i X_i - \overline{X} \sum f_i$$

$$= \sum f_i X_i - \sum f_i X_i = 0$$

olarak elde edilir.

ÖRNEK: X_i : 4; 6; 7; 10;I3

serisinde $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$ olduğunu gösteriniz.

$$\overline{X} = \frac{4+6+7+10+13}{5} = \frac{40}{5} = 8$$

X_i	$X_i - \overline{X}$
4	4-8=-4
6	6-8=-2
7	7-8=-1
10	10-8=2
13	13-8=5
	$\sum(X_i - \overline{X}) = 0$

ÖRNEK: Aşağıdaki seride $\sum_{i=1}^{k} f_i(X_i - \overline{X}) = 0$ olduğunu gösteriniz.

X_i	f_i	$f_i X_i$	$(X_i - \overline{X})$	$f_i(X_i - \overline{X})$
0	2	0	-1,9	-3,8
1	5	5	-0,9	-4,5
2	8	16	0,1	0,8
3	3	9	1,1	3,3
4	2	8	2,1	4,2
	20	38		$\sum f_i(X_i - \overline{X}) = 0$

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{38}{20} = 1{,}9$$

- Bir serinin değerlerinin aritmetik ortalamadan farklarının kareleri toplamı minimumdur. Basit seriler için,

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 < \sum_{i=1}^{n}(X_i - A)^2 \qquad A \neq \overline{X}$$

frekans serileri ve sınıflı seriler için

$$\sum_{i=1}^{k} f_i(X_i - \overline{X})^2 < \sum_{i=1}^{k} f_i(X_i - A)^2 \qquad A \neq \overline{X}$$

olarak gösterilir.

Basit serilerde;

$$\sum(X_i - \overline{X})^2 < \sum(X_i - A)^2$$

olarak gösterilmişti. Eşitsizliğin sağ tarafına $\overline{X}$ ekler ve çıkarırsak, eşitsizlik;

$$\sum(X_i - \overline{X})^2 < \sum(X_i - \overline{X} + \overline{X} - A)^2$$

$$\sum(X_i - \overline{X})^2 < \sum[(X_i - \overline{X}) + (\overline{X} - A)]^2$$

olacaktır. Sağ tarafın karesini alarak kısaltalım.

$$\sum(X_i - \overline{X})^2 < \sum[(X_i - \overline{X}) + 2(X_i - \overline{X})(\overline{X} - A) + (\overline{X} - A)]^2$$

$$\sum(X_i - \overline{X})^2 < \sum(X_i - \overline{X})^2 + 2\sum(X_i - \overline{X})(\overline{X} - A) + \sum(\overline{X} - A)^2$$

X ve A sabit sayılar ve $\sum(X_i - \overline{X}) = 0$ olduğundan;

$$\sum(X_i - \overline{X})^2 < \sum(X_i - \overline{X})^2 + 2(\overline{X} - A)\sum(X_i - \overline{X}) + n(\overline{X} - A)^2$$

$$\sum(X_i - \overline{X})^2 < \sum(X_i - \overline{X})^2 + n(\overline{X} - A)^2$$

olarak gösterilir. A sayısı $\overline{X}$'dan hem büyük hem de küçük olması durumunda, $(\overline{X} - A)^2$ pozitif bir sayı olacağından Eşitsizliğin doğruluğu ispatlanmaktadır.

Frekans serisi veya sınıflı seride ise;

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i(X_i - A)^2$$

olacağı daha önce gösterilmişti. Yine eşitsizliğin sağ tarafına $\overline{X}$ ekleyelim ve çıkaralım.

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i(X_i - \overline{X} + \overline{X} - A)^2$$

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i[(X_i - \overline{X}) + (\overline{X} - A)]^2$$

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i[(X_i - \overline{X}) + 2(X_i - \overline{X})(\overline{X} - A) + (\overline{X} - A)]^2$$

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i(X_i - \overline{X}) + 2(\overline{X} - A)\sum f_i(X_i - \overline{X}) + \sum f_i(\overline{X} - A)^2$$

$\overline{X}$ ve A sabit sayılar ve $\sum f_i(X_i - \overline{X}) = 0$ olduğundan;

$$\sum f_i(X_i - \overline{X})^2 < \sum f_i(X_i - \overline{X})^2 + n(\overline{X} - A)^2$$

olarak elde edilir, böylece eşitsizliğin doğruluğu ispatlanmış olur.

ÖRNEK: X_i : 4; 6; 7; 13; 15

serisinde $\sum(X_i - \overline{X})^2 < \sum(X_i - 6)^2$ olduğunu gösteriniz.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{4+6+7+13+15}{5} = \frac{45}{5} = 9$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$X_i - 6$	$(X_i - 6)^2$
4	-5	25	-2	4
6	-3	9	0	0
7	-2	4	1	1
13	4	16	7	49
15	6	36	9	81
		90		135

90<135

ÖRNEK: Aşağıdaki seride

$\sum f_i (X_i - \overline{X})^2 < \sum f_i (X_i - 7)^2$ olduğunu gösteriniz.

Sınıflar	f_i	X_i	$f_i X_i$	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$	$X_i - 7$	$(X_i - 7)^2$	$f_i(X_i - 7)^2$
1-3	10	2	20	-3,4	11,56	115,60	-5	25	250
3-5	15	4	60	-1,4	1,96	1,96	-3	9	135
5-7	11	6	66	0,6	0,36	0,36	-1	1	11
7-9	8	8	64	2,6	6,76	6,76	1	1	8
9-11	6	10	60	4,6	21,16	21,16	3	9	54
	50		270			330,00			458

330<458

- Bir serinin bütün değerlerine aynı sayının eklenmesi veya çıkarılması ile oluşturulacak yeni serinin aritmetik ortalaması, ilk serinin aritmetik ortalamasına aynı sayının eklenmesi veya çıkarılmasına eşittir.

Aritmetik ortalaması $\overline{X}$ olan X_i serisinin bütün değerlerine A gibi sabit bir sayı eklenerek veya çıkarılarak bir yi serisi oluşturuluyorsa, Yi serisinin aritmetik ortalaması,

$Y_i = X_i - A$ ise $\overline{Y} = \overline{X} - A$

$Y_i = X_i + A$ ise $\overline{Y} = \overline{X} + A$

olacaktır.

Bir serinin bütün değerlerine aynı sabit A sayısı eklenirse yeni serinin toplam değeri

$$\sum_{i=1}^{n} Y_i = (X_1 + A) + (X_2 + A) + \cdots + (X_n + A)$$

$$\sum_{i=1}^{n} (X_i + A) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} A = \sum_{i=1}^{n} X_i + nA$$

olarak elde edilir.

Yeni serinin aritmetik ortalaması ise,

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{\sum X_i + nA}{n} = \overline{X} + A$$

olacaktır.

Seri frekans serisi veya sınıflı seri ise,

$$\sum_{i=1}^{k} f_i Y_i = f_1(X_1 + A) + f_2(X_2 + A) + \cdots + f_k(X_k + A)$$

$$= \sum_{i=1}^{k} f_i (X_i + A) = \sum_{i=1}^{k} f_i X_i + \sum_{i=1}^{k} f_i A$$

$$= \sum_{i=1}^{k} f_i X_i + A \sum_{i=1}^{k} f_i$$

olarak elde edilir. Yeni serinin aritmetik ortalaması ise,

$$\overline{Y} = \frac{\sum_{i=1}^{n} f_i Y_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i X_i + A \sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} f_i} = \overline{X} + A$$

olacaktır. Aynı işlemler çıkartma için de yapılabilir.

ÖRNEK: Aşağıdaki serinin bütün değerlerinden 7 çıkartarak bir Y_i serisi oluşturunuz ve $\overline{Y} = \overline{X} - 7$ olduğunu gösteriniz.

X_i	$Y_i = X_i - 7$
15	15-7=8
18	18-7=11
22	22-7=15
30	30-7=23
33	33-7=26
38	38-7=31
156	114

$$\overline{X} = \frac{\sum X_i}{n} = \frac{156}{6} = 26$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{114}{6} = 19$$

$$\overline{Y} = \overline{X} - 7 = 26 - 7 = 19$$

Görüldüğü gibi iki ayrı hesaplamada aynı sonucu vermektedir.

ÖRNEK: Aşağıdaki serinin bütün değerlerine 2 ekleyerek bir Z_i serisi oluşturunuz ve $\overline{Z} = \overline{X} + 2$ olduğunu gösteriniz.

X_i	f_i	f_iX_i	$Z=X_i+2$	f_iZ_i
1	14	14	1+2=3	42
2	20	40	2+4=4	80
3	23	69	3+2=5	115
4	16	64	4+2=6	96
5	7	35	5+2=7	49
	80	222		382

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{222}{80} = 2{,}775$$

$$\overline{Z} = \frac{\sum f_i Z_i}{\sum f_i} = \frac{382}{80} = 4{,}775$$

$$\overline{Z} = \overline{X} + 2 = 2{,}775 + 2 = 4{,}775$$

Görüldüğü gibi iki ayrı hesaplamada aynı sonucu vermektedir.

- Bir serinin bütün değerlerinin aynı sayı ile çarpılması veya bölünmesi ile oluşturulacak yeni serinin aritmetik ortalaması, ilk serinin aritmetik ortalamasının aynı sayı ile çarpımına veya bölümüne eşittir.

Aritmetik ortalaması $\overline{X}$ olan serisinin bütün değerlerinin A gibi sabit bir sayı ile çarpılması veya bölünmesi ile bir Y_i serisi oluşturuluyorsa, Y_i serisinin aritmetik ortalaması,

$Y_i = A.X_i$ ise $\overline{Y} = A.\overline{X}$

$Y_i = \frac{X_i}{A}$ ise $\overline{Y} = \frac{\overline{X}}{A}$

olacaktır.

Bir serinin bütün değerleri A gibi sabit bir sayı ile çarpıldığında yeni serinin toplam değeri,

$$\sum Y_i = AX_1 + AX_2 + \cdots + AX_n$$

$$= A\sum X_i$$

olacaktır. Bu durumda Y ortalama,

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{A\sum X_i}{n} = A\overline{X}$$

olarak bulunur.

Seri, frekans serisi veya sınıflı seri ise yeni serinin toplam değeri,

$$\sum f_i X_i = Af_1X_1 + Af_2X_2 + \cdots + Af_kX_k$$

$$= \sum Af_iX_i$$

$$= A\sum f_iX_i$$

olacaktır. Y serisinin aritmetik ortalaması ile,

$$\overline{Y} = \frac{\sum f_i Y_i}{\sum f_i} = \frac{A\sum f_i X_i}{\sum f_i} = A\overline{X}$$

olarak bulunur. Aynı işlemler bölme işlemi için de yapılabilir.

ÖRNEK: Aşağıda verilen X_i serisinin bütün değerlerini 3 ile çarparak bir Y_i serisi oluşturunuz. Oluşturduğunuz seride $\overline{Y} = 3\overline{X}$ olduğunu gösteriniz.

X_i : 4; 5; 8; 8; 10

$$\overline{X} = \frac{\sum X_i}{n} = \frac{4+5+8+8+10}{5} = \frac{35}{5} = 7$$

Y_i
3.4=12
3.5=15
3.8=24
3.8=24
3.10=30
105

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{105}{5} = 21$$

$$\overline{Y} = 3\overline{X} = 3.7 = 21$$

Görüldüğü gibi iki ayrı hesaplama da aynı sonucu vermektedir.

ÖRNEK: Aşağıda verilen serinin bütün değerlerini 2'ye bölerek bir Y_i serisi oluşturunuz ve $\overline{Y} = \overline{X}/2$ olduğunu gösteriniz.

Sınıflar	X_i	f_i	f_iX_i
0-4	2	8	16
4-8	6	16	96
8-12	10	24	240
12-16	14	12	168
16-20	18	10	180
		70	700

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{700}{70} = 10$$

Sınıflar	Sınıflar	Y_i	f_i	f_iY_i
0/2-4/2	0-2	1	8	8
4/2-8/2	2-4	3	16	48
8/2-12/2	4-6	5	24	120
12/2-16/2	6-8	7	12	84
16/2-20/2	8-10	9	10	90
			70	350

$$\overline{Y} = \frac{\sum f_i Y_i}{\sum f_i} = \frac{350}{70} = 5$$

$$\overline{Y} = \frac{\overline{X}}{2} = \frac{10}{2} = 5$$

Görüldüğü gibi iki ayrı hesaplamada aynı sonucu vermektedir.

- İki veya daha fazla serinin değişkenlerinin aldıkları değerlerin toplamları veya farkları, yeni bir seri oluşturuyorsa, yeni serinin aritmetik ortalaması, kendisini oluşturan serilerin aritmetik ortalamalarının farklarına veya toplamlarına eşittir.

$Y_i = X_i + Z_i - T_i$ ise $\overline{Y} = \overline{X} + \overline{Z} - \overline{T}$

olacaktır.

$$Y = \frac{\sum Y_i}{n}$$

olduğundan, formülde Y_i yerine eşiti olan $(X_i + Z_i - T_i)$ ifadesi konulursa

$$Y = \frac{\sum Y_i}{n} = \frac{\sum (X_i + Z_i - T_i)}{n} = \frac{\sum X_i + \sum Z_i - \sum T_i}{n}$$

$$= \frac{\sum X_i}{n} + \frac{\sum Z_i}{n} - \frac{\sum T_i}{n}$$

$$\overline{X} = \frac{\sum X_i}{n}; \qquad \overline{Z} = \frac{\sum Z_i}{n}; \qquad \overline{T} = \frac{\sum T_i}{n}$$

olduğundan

$$\overline{Y} = \overline{X} + \overline{Z} - \overline{T}$$

eşitliğinin, elde edildiği görülmektedir.

ÖRNEK: Aşağıda verilen seri için $\overline{Y} = \overline{X} + \overline{Z}$ ve $\overline{V} = \overline{X} - \overline{Z}$ olduğunu gösteriniz.

X_i	Z_i	$Y_i=X_i+Z_i$	$V_i=X_i-Z_i$
2	1	3	+1
3	4	7	-1
6	7	13	-1
9	10	19	-1
20	13	33	+7
40	35	75	+5

$$\overline{X} = \frac{\sum X_i}{n} = \frac{40}{5} = 8$$

$$\overline{Z} = \frac{\sum Z_i}{n} = \frac{35}{5} = 7$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{75}{5} = 15$$

$$\overline{V} = \frac{\sum V_i}{n} = \frac{5}{5} = 1$$

olarak hesaplanır.

$Y_i = X_i + Z_i$ olduğundan $\overline{Y} = \overline{X} + \overline{Z} = 8 + 7 = 15$

$V_i = X_i - Z_i$ olduğundan $\overline{V} = \overline{X} - \overline{Z} = 8 - 7 = 1$

olarak bulunur.

Görüldüğü gibi iki ayrı hesaplamada aynı sonuçları vermektedir.

- Bir seri farklı serilerin birleştirilmesi ile oluşturuluyorsa, yeni serinin aritmetik ortalaması kendini oluşturan serilerin aritmetik ortalamaların tartılı aritmetik ortalamasına eşittir. Burada tartı olarak serilerin birim sayıları alınmaktadır.

m birimli X_l serisi ile , r birimli Y_j serisinin birleştirilerek
n = m + r birimli Z_i serisi oluşturduğunu varsayalım.

Bu seriler;

$X_l = X_1 + X_2 , ... , X_m$
$Y_j = Y_1 + Y_2 , ... , Y_r$ ise

$Z_i = X_1 + X_2 , ... , X_m , Y_1, Y_2 , ..., Y_r$
olacaktır. Bu durumda,

$$\overline{Z} = \frac{m\overline{X} + r\overline{Y}}{m + r}$$

olarak elde edilir.

Z serisinin toplam birim sayısı n = m + r dir. $m\overline{X}$, X serisinin toplam değerini; $r\overline{Y}$, Y serisinin toplam değerini vereceğinden,

$$\sum_{i=1}^{n} Z_i = \sum_{l=1}^{m} X_i + \sum_{j=1}^{r} Y_j$$

$$= m\overline{X} + r\overline{Y}$$

olur ve Z serisinin aritmetik ortalaması

$$\overline{Z} = \frac{\sum_{i=1}^{n} Z_i}{n} = \frac{m\overline{X} + r\overline{Y}}{m + r}$$

olarak elde edilir.

ÖRNEK: Aşağıda verilen X ve Y serilerini birleştirerek bir Z serisi oluşturunuz ve ortalamaları arasındaki ilişkiyi gösteriniz.

X_l	Y_j
4	1
5	2
7	4
9	6
25	7
	20

Yeni oluşturulacak Z serisi,

Z_i= 4; 5; 7; 9; 1; 2; 4; 6; 7

olacaktır.

$$\bar{Z} = \frac{\sum_{i=1}^{n} Z_i}{n} = \frac{4+5+7+9+1+2+4+6+7}{9}$$

$$= \frac{45}{9} = 5$$

$$\bar{X} = \frac{\sum_{l=1}^{m} X_l}{m} = \frac{25}{4} = 6{,}25$$

$$\bar{Y} = \frac{\sum_{j=1}^{r} Y_j}{r} = \frac{20}{5} = 4$$

$$\bar{Z} = \frac{m\bar{X} + r\bar{Y}}{m+r} = \frac{4(6,25)+5(4)}{4+5} = \frac{25+20}{9}$$

$$= \frac{45}{9} = 5$$

Görüldüğü gibi iki ayrı hesaplamada aynı sonucu vermektedir.

- Bir anakütlenin aritmetik ortalaması (μ) örnekten alınarak tahmin ediliyorsa, örnek ortalaması ana kütle ortalamasının sapmasız tahmincisi olduğundan $\mu = \bar{X}$ olarak tahmin edilir.

3.1.2. Tartılı Aritmetik Ortalama

Buraya kadar yapılan açıklamalardan anlaşılacağı gibi, aritmetik ortalama ile ilgili hesaplamalarda serilerin bütün birimlerinin aynı öneme sahip olduğu kabul edilmektedir. Uygulamada bazı durumlar da birimlerin önemleri birbirinden farklı olabilir. Bu gibi durumlar da, tartısız aritmetik ortalama yerine tartılı aritmetik ortalamanın hesaplanması gerekmektedir. Örneğin, öğrencilerin herhangi bir ders için başarı notlarının hesaplandığını ve başarı notu hesabında aritmetik ortalama kullanıldığını düşünelim. Ara sınavlar ile final notlarının başarı notunu ayni oranda etkilemesi isteniyorsa tartısız aritmetik ortalama hesaplanacaktır. Final sınavının, başarı notunu ara sınavlardan daha çok etkilemesi isteniyorsa, tartılar belirlenerek, tartılı aritmetik ortalamanın hesaplanması gerekecektir.

Uygulamada tartılar farklı şekillerde belirlenebilir. Örneğin, yukarıda verdiğimiz başarı notu örneğinde, tartıları kurumların yöneticileri kendi görüşlerine göre belirlemektedirler. Bazı durumlarda tartılar araştırmalar yapılarak belirlenmekte, Bazı durumlarda ise incelenen konu ile ilgili Bazı değerler tartı olarak alınabilmektedir. Örneğin, 3 bölgede ekmek fiyatları farklı ise ve bunların tartılı aritmetik ortalaması hesaplanacaksa, bölgedeki ekmek tüketim miktarları tartı olarak alınabilir. Ancak bu 3 bölgenin ekmek tüketim miktarları ayni ise tartılar birbirine eşit olacağından tartılı ve tartısız aritmetik ortalamalarında birbirine eşit olacağı görülecektir.

Basit serilerde tartılar söz konusu olduğu zaman, serideki değişken ile birlikte, ikinci bir değişken olarak tartılar da yer almaktadır. Tartıların yer aldığı basit seri genel olarak şöyle gösterilebilir.

X_i	t_i
X_1	t_1
X_2	t_2
X_3	t_3
$\vdots$	$\vdots$
X_n	t_n

Bu serinin tartılı aritmetik ortalaması,

$$\overline{X}_t = \frac{\sum_{i=1}^{n} t_i X_i}{\sum_{i=1}^{n} t_i}$$

formülü ile hesaplanmaktadır. Burada (t) indisi aritmetik ortalamanın tartılı olduğunu göstermek için kullanılmıştır.

Tartıların a gibi sabit bir sayıya eşit olduğu durumlarda tartılı aritmetik ortalama ile tartısız aritmetik ortalama birbirine eşit olmaktadır. Bu eşitliği aşağıdaki gibi gösterebiliriz.

$$X_t = \frac{\sum_{i=1}^{n} a_i X_i}{\sum_{i=1}^{n} a} = \frac{a\sum_{i=1}^{n} X_i}{na} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

Görüldüğü gibi,

$$\overline{X}_t = \overline{X}$$

olarak elde edilmektedir.

ÖRNEK: Aşağıda verilen basit serinin tartılı ve tartısız aritmetik ortalamasını hesaplayınız.

X_i	t_i	$X_i t_i$
3	4	12
4	5	20
5	1	5
6	2	12
8	3	24
26	15	73

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{26}{5} = 5{,}2$$

$$\overline{X}_t = \frac{\sum_{i=1}^{n} t_i X_i}{\sum_{i=1}^{n} t_i} = \frac{73}{15} = 4{,}86$$

Tartıların hesaba katılması ile sonucun değiştiği görülmektedir.

Aynı örnekte tartıların birbirine eşit ve 3 olduğunu varsayalım.

X_i	t_i	$X_i t_i$
3	3	9
4	3	12
5	3	15
6	3	18
8	3	24
	15	78

$$\overline{X}_t = \frac{\sum t_i X_i}{\sum t_i} = \frac{78}{15} = 5{,}2$$

Hesaplamalardan görüleceği gibi, basit serilerde tartılar birbirine eşit olduğunda, tartılı ve tartısız aritmetik ortalamalar da birbirine eşit olmaktadır.

Basit serilerde olduğu gibi, frekans serilerinde de tartılı aritmetik ortalamalar hesaplanabilir. Bu hesaplamalarda tartıların, üçüncü sütun olarak frekans serilerine eklenmeleri gerekecektir. Tartıların yer aldığı frekans serileri genel olarak şöyle gösterilebilir.

X_i	f_i	t_i
X_1	f_1	t_1
X_2	f_2	t_2

X_3	f_3	t_3
⋮	⋮	⋮
X_k	f_k	t_k

Bu frekans serinin tartılı aritmetik ortalaması

$$\overline{X}_t = \frac{\sum_{i=1}^{k} t_i f_i X_i}{\sum_{i=1}^{k} t_i f_i}$$

formülü ile hesaplanmaktadır. Formülde görülen tartıların a gibi sabit bir sayıya eşit olduğu durumlarda tartılı aritmetik ortalama ile tartısız aritmetik ortalama birbirine eşit olmaktadır. Bu eşitliği aşağıdaki gibi gösterebiliriz.

$$\overline{X}_t = \frac{\sum t_i f_i X_i}{\sum t_i f_i} = \frac{\sum a f_i X_i}{\sum a f_i} = \frac{a\sum f_i X_i}{a\sum f_i}$$

$$= \frac{\sum f_i X_i}{\sum f_i} = \overline{X}$$

Görüldüğü gibi,

$$\overline{X}_t = \overline{X}$$

olarak elde edilmektedir.

Tartılı aritmetik ortalama, sınıflı serilerde de, frekans serilerinde kullanılan formül ile hesaplanmaktadır. Ancak seri sınıflı olduğundan, formüldeki X_i yerine sınıf ortaları alınacaktır.

Frekans serilerinde ve sınıflı serilerde aritmetik ortalamanın hesaplanmasını iki örnek ile gösterelim.

ÖRNEK: Aşağıdaki frekans serisinin tartılı ve tartısız aritmetik ortalamalarını hesaplayınız.

X_i	f_i	t_i	f_iX_i	$t_if_iX_i$	t_if_i
1	8	1	8	8	8
2	12	2	24	48	24
3	15	4	45	180	60
4	5	3	20	60	15
	40		97	296	107

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{97}{40} = 2{,}425$$

$$\overline{X}_t = \frac{\sum t_i f_i X_i}{\sum t_i f_i} = \frac{296}{107} = 2{,}766$$

Tartılı ve tartısız aritmetik ortalamaların farklı sonuçlar verdiği görülmektedir.

ÖRNEK: Aşağıdaki sınıflı serinin tartılı ve tartısız aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	t_i	X_i	f_iX_i	$t_if_iX_i$	t_if_i
5-15	4	5	10	40	200	20
15-25	7	2	20	140	280	14
25-35	6	2	30	180	360	12
35-45	3	1	40	120	120	3
	20			480	960	49

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{480}{20} = 24$$

$$\overline{X}_t = \frac{\sum t_i f_i X_i}{\sum t_i f_i} = \frac{960}{49} = 19{,}59$$

Sınıflı serilerde de tartıların hesaba katılması ile aritmetik ortalamanın değerinin değiştiği görülmektedir.

3.2. KARELİ ORTALAMA

Seri değerlerinin karelerinin aritmetik ortalamasının kare köküne kareli ortalama denmektedir. Kareli ortalama K harfi ile gösterilecektir.

Kareli ortalamanın hesaplanması için, seri değerlerinin karelerinin toplam değerinin bulunması gerekecektir. Bu toplam değerin seri birim sayısına bölümü ile seri birimlerinin karelerinin aritmetik ortalaması elde edilecek ve elde edilen bu sonucun karekökü ise kareli ortalamayı verecektir. Kareli ortalama hesaplanırken, seri değerlerinin kareleri alındığından karekök ile düzeltme yapılmaktadır.

Basit serilerde seri değerlerinin kareleri toplamı,

$$X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2 = \sum X_i^2$$

olacağından, kareli ortalama;

$$K = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$$

formülü ile hesaplanacaktır.

ÖRNEK: 3; 4; 4; 7; 8; 9 serisinin kareli ortalamasını hesaplayınız.

$$K = \sqrt{\frac{3^2 + 4^2 + 4^2 + 7^2 + 8^2 + 9^2}{6}}$$

$$= \sqrt{\frac{9 + 16 + 16 + 49 + 64 + 81}{6}}$$

$$= \sqrt{\frac{235}{6}} = \sqrt{39{,}1666} = 6{,}25$$

Frekans serilerinde de kareli ortalamanın hesaplanması için seri değerlerinin karelerinin toplam değerinin bulunması gerekmektedir.

Bu serilerde $f_1 X_1^2$, X_1^2'lerin toplam değerini; $f_2 X_2^2$, X_2^2'lerin toplam değerini; .....; $f_k X_k^2$, X_k^2'lerin toplamını vereceğinden, seri değerlerin karelerinin toplam değeri;

$$f_1 X_1^2 + f_2 X_2^2 + f_3 X_3^2 + \cdots + f_k X_k^2 = \sum f_i X_i^2$$

olacaktır. Bu nedenle frekans serilerinde kareli ortalama,

$$K = \sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{\sum_{i=1}^{k} f_i}}$$

formülü ile hesaplanmaktadır.

ÖRNEK: Aşağıda verilen serinin kareli ortalamasını hesaplayınız.

X_i	f_i	X_i^2	$f_i X_i^2$
2	5	4	20
3	8	9	72
4	14	16	224

5	17	25	425
6	6	36	216
	50		957

$$K = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{957}{50}} = \sqrt{19,14} = 4,37$$

Sınıflı serilerde kareli ortalama frekans serilerinde kullanılan formül ile hesaplanmaktadır. Ancak seri sınıflı olduğundan formüldeki X_i yerine sınıf ortaları alınacaktır.

ÖRNEK: Aşağıda verilen serinin kareli ortalamasını hesaplayınız.

Sınıflar	f_i	X_i	X_i^2	$f_i X_i^2$
1-5	3	3	9	27
5-9	4	7	49	196
9-13	8	11	121	968
13-17	15	15	225	3375
17-21	10	19	361	3610
	40			8176

$$K = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{8176}{40}} = \sqrt{204,4} = 14,29$$

3.3. GEOMETRİK ORTALAMA

Seri değerlerinin çarpımlarının, seri birim sayısı derecesinden köküne geometrik ortalama denmektedir. Geometrik ortalama G harfi ile gösterilecektir.

Basit serilerde seri birimlerinin çarpımları,

$$X_1.X_2.X_3.\cdots.X_n = \prod_{i=1}^{n} X_i$$

olacağından, geometrik ortalama;

$$G = \sqrt[n]{X_1.X_2.X_3.\cdots.X_n} = \sqrt[n]{\prod_{i=1}^{n} X_i}$$

formülü ile hesaplanmaktadır.

ÖRNEK: 3;5;5;6;8;11 serisinin geometrik ortalamasını hesaplayınız.

$$G = \sqrt[n]{\prod_{i=1}^{n} X_i} = \sqrt[6]{3.5.5.6.8.11} = \sqrt[6]{39600} = 5{,}83$$

Verilen örnekte 6. dereceden kök üstel fonksiyonlu bir hesap makinesi ile bulunmuştur. Böyle bir makine olmadığında n. dereceden kök logaritma yardımı ile alınabilir. Bu durumda,

$$\log G = \log \sqrt[n]{X_1.X_2.X_3.\cdots.X_n}$$

$$\log G = \frac{1}{n}(\log X_1 + \log X_2 + \log X_3 + \cdots + \log X_n)$$

$$\log G = \frac{1}{n}\sum_{i=1}^{n} \log X_i$$

olacaktır. log G'nin değeri hesaplandıktan sonra antilogaritması alınarak geometrik ortalama bulunmaktadır.

ÖRNEK: Daha önce verilen basit serinin geometrik ortalamasını logaritma yardımı ile hesaplayınız.

$$\log G = \log \sqrt[n]{X_1.X_2.X_3.\cdots.X_n}$$

$$\log G = \frac{1}{6}(\log 3 + \log 5 + \log 5 + \log 6 + \log 8 + \log 11)$$

$$= \frac{1}{6}(0{,}47712 + 0{,}69897 + 0{,}69897 + 0{,}77815 + 0{,}90308 + 1{,}04139)$$

$$= \frac{1}{6}.4{,}59768$$

log G = 0,76628 $\rightarrow$ G = 5,83

Her iki hesaplamada da geometrik ortalamanın aynı değerde olduğu görülmektedir.

Frekans serilerinde geometrik ortalama hesaplanırken de seri birimlerinin çarpımlarının bulunması gerekmektedir. Bu değerlerin çarpımı,

$$X_1.X_1.\cdots.X_1 \qquad X_2.X_2.\cdots.X_2 \cdots.X_k.X_k.\cdots.X_k$$

olacaktır.

X_1'in tekrar sayısı f_1; X_2'nin tekrar sayısı f_2; ....; X_k'nın tekrar sayısı f_k olduğundan,

$$X_1^{f_1} X_2^{f_2} X_3^{f_3} \cdots X_k^{f_k}$$

olarak yazılabilir. Bu durumda geometrik ortalama;

$$G = \sqrt[\sum_{i=1}^{k} f_i]{X_1^{f_1} X_2^{f_2} X_3^{f_3} \cdots X_k^{f_k}} = \sqrt[\sum_{i=1}^{k} f_i]{\prod_{i=1}^{k} X_i^{f_i}}$$

formülü ile hesaplanmaktadır.

Geometrik ortalama logaritma yardımıyla hesaplanacaksa;

$$\log G = \frac{1}{\sum_{i=1}^{k} f_i} . \sum_{i=1}^{k} f_i \log X_i$$

formülü kullanılacaktır.

ÖRNEK: Aşağıdaki frekans serisinin geometrik ortalamasını hesaplayınız.

X_i	f_i	$\log X_i$	$f_i \log X_i$
1	2	0	0
2	4	0,30103	1,20412
3	7	0,47712	3,33984
4	4	0,60206	2,40824
5	3	0,69897	2,09691
	20		9,04911

$$\log G = \frac{1}{\sum f_i} \sum f_i \log X_i$$

$$= \frac{1}{20}(9{,}04911)$$

$$= 0{,}45245 \rightarrow G = 2{,}83$$

Sınıflı serilerde geometrik ortalama frekans serilerinde kullanılan,

$$G = \sqrt[\sum_{i=1}^{k} f_i]{\prod_{i=1}^{k} X_i^{f_i}}$$

formülü ile hesaplanmakta, ancak formülde yer olan X_i değerleri yerine sınıf orta değerleri kullanılmaktadır.

ÖRNEK: Aşağıdaki sınıflı serinin geometrik ortalamasını hesaplayınız.

Sınıflar	f_i	X_i	$\log X_i$	$f_i \log X_i$
0-5	4	2,5	0,39794	1,59176

5-10	9	7,5	0,87506	7,87554
10-15	15	12,5	1,09691	16,45365
15-20	2	17,5	1,24304	2,48608
	30			28,40703

$$\log G = \frac{1}{\sum f_i} \sum f_i \log X_i$$

$$= \frac{1}{30}(28{,}40703)$$

$$= 0{,}94690 \rightarrow G = 8{,}84$$

3.4. HARMONİK ORTALAMA

Seri değerlerinin terslerinin aritmetik ortalamasının tersine harmonik ortalama denmektedir. Harmonik ortalama H harfi ile ifade edilmektedir.

Basit serilerde seri birim sayısı, seri birimlerinin terslerinin toplamına bölünerek harmonik ortalama hesaplanmış olur. Bu nedenle basit serilerde harmonik ortalama,

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

$$= \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

formülü ile hesaplanacaktır.

ÖRNEK: 3, 5, 5, 6, 8, 11 serisinin harmonik ortalamasını hesaplayınız.

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$= \frac{6}{\frac{1}{3} + \frac{1}{5} + \frac{1}{5} + \frac{1}{6} + \frac{1}{8} + \frac{1}{11}} = \frac{6}{1{,}1159} = 5{,}37$$

Frekans serilerinde harmonik ortalama hesaplanırken diğer ortalamalarda olduğu gibi frekanslar, yani birimlerin tekrar sayıları dikkate alınır. Bu durumda birimlerin terslerinin toplamı,

$$\underbrace{\frac{1}{x_1}+\frac{1}{x_1}+\ldots+\frac{1}{x_1}}_{f_1}+\underbrace{\frac{1}{\lambda_2}+\frac{1}{x_2}+\ldots+\frac{1}{x_2}}_{f_2}+\ldots+\underbrace{\frac{1}{x_4}+\frac{1}{x_k}+\ldots+\frac{1}{\lambda_4}}_{f_k}$$

$$=\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_k}{x_k}=\sum_{i=1}^{k}\frac{f_i}{x_i}$$

olacaktır. Bu durumda harmonik ortalama,

$$H=\frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

olacaktır.

ÖRNEK: Aşağıdaki frekans serisinin harmonik ortalamasını hesaplayınız.

X_i	f_i	f_i/X_i
1	3	3,0000
2	5	2,5000
3	8	2,666
4	5	1,2500
5	4	0,8000
	25	10,216

$$H=\frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

$$=\frac{25}{10,2166}=2,44$$

Sınıflı serilerde harmonik ortalama frekans serilerinde kullanılan,

$$H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

formülü ile hesaplanmakta, ancak formülde yer alan xi değerleri yerine sınıf orta değerleri kullanılmaktadır.

ÖRNEK: Aşağıdaki sınıflı serinin harmonik ortalamasını hesaplayınız.

Sınıflar	f_i	X_i	f_i/X_i
0-2	3	1	3,0000
2-4	8	3	2,6666
4-6	12	5	2,4000
6-8	5	7	0,7142
8-10	2	9	0,2222
	30		9,0030

$$H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

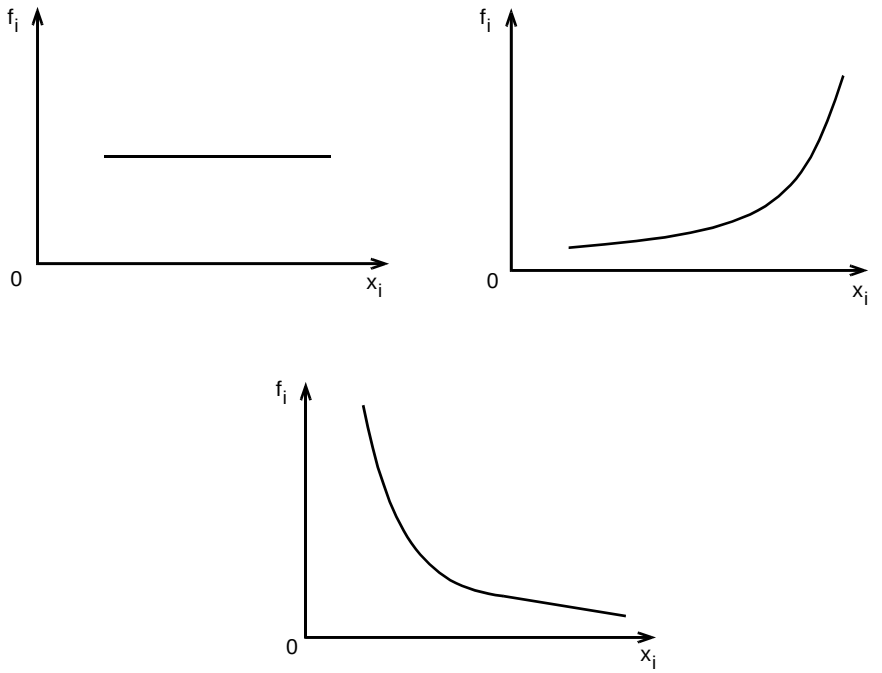
$$= \frac{30}{9,0030} = 3,33$$

Seri birimlerinin bir veya birkaçının değeri sıfırsa eşit olduğunda harmonik ortalama hesaplanamaz. Ayrıca seride ters işaretli birimler bulunduğunda harmonik ortalama anlamsız sonuç verebilir.

3.5. MOD

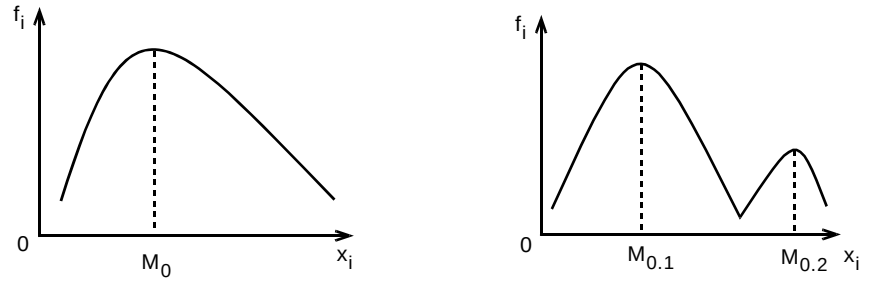
Serinin en çok tekrarlanan değerine mod denmekte ve Mo sembolü ile gösterilmektedir. Modun uygun ölçü olabilmesi için, serinin tek modlu olması gerekmektedir.

Bir seride değişkenin aldığı farklı değerler aynı sayıda tekrarlanıyorsa bu seri modsuz seridir. Aynı şekilde, tekrar sayıları sürekli artan ve azalan serilerde modsuz seriler olarak adlandırılmaktadır. Örneğin, bir seride değişkenin aldığı farklı değerler Üçer kere tekrarlanıyorsa bu seri modsuz seridir. Ayni şekilde değişkenin aldığı farklı değerlerin tekrar sayıları 2,5,6,9,.... Gibi artıyor veya 21, 18, 12, 10,..., gibi azalıyorsa bu serilerde modsuz serilerdir. Bu serilerin frekans dağılımlarının grafikleri Şekil 3.1.'de görülmektedir.



Şekil 3.1

Serinin modlu bir seri olabilmesi için, tekrar sayılarının, (frekanslarının) maksimum değerinin veya değerlerinin olması gerekmektedir. Eğer frekans dağılımlarının tek maksimum değeri varsa seri tek modlu, birden fazla maksimum değeri varsa, seri çok modludur. Tek ve çok modlu serilerin frekans dağılımlarının grafikleri Şekil 3.2.'de görülmektedir.



Şekil 3.2

Basit serilerde, özellikle örnek birim sayısının az olduğu durumlarda mod seriyi yeterince temsil etmeyebilir. Bu gibi durumlarda mod, serinin en çok tekrarlanan biriminin değeri Olacaktır. Ancak burada serinin tek maksimumlu yani tek modlu olmasına dikkat edilmesi gerekecektir.

ÖRNEK: Aşağıda verilen serinin modunu hesaplayınız.

2; 3; 3; 4; 4; 4; 6; 6; 6; 6; 6; 6; 7; 7; 7; 7; 8; 8; 9

Yukarıda verilen 20 birimli basit serinin en çok tekrarlanan değeri olan 6 serinin modudur. Modun uygun ölçü olabilmesi için serinin tek maksimumlu olup olmadığını araştırmak gerecektir. Bunun için seri değerlerinin tekrar sayılarına bakılması yeterli olmaktadır. Burada, 2, 1 defa; 3, 2 defa; 4, 3 defa; 6, 7 defa; 7, 4 defa; 8, 2 defa ve 9, 1 defa tekrarlandığından tekrar sayıları, önce artarak 6'nin 7 defa tekrarlanmasında maksimuma ulaşmakta, daha sonra düşmektedir. Seri tek modlu olduğundan bu seri için mod, uygun ölçü olarak kabul edilmektedir.

Frekans serilerinde ise frekansı en yüksek olan değer serinin modudur. Bu serilerde de modun uygun ölçü olup olmayacağının belirlenmesi için, serinin tek modlu olup olmadığına bakılması gerekmektedir.

ÖRNEK: Aşağıdaki frekans serisinin modunu hesaplayınız.

X_i	f_i
2	15
3	21
4	36
6	27
7	18
8	13
	130

Seride en yüksek frekans 36 olduğundan bu frekansın değeri olan 4 serinin modudur. Frekansların dağılımı incelendiğinde de serinin tek modlu olduğu görülecektir.

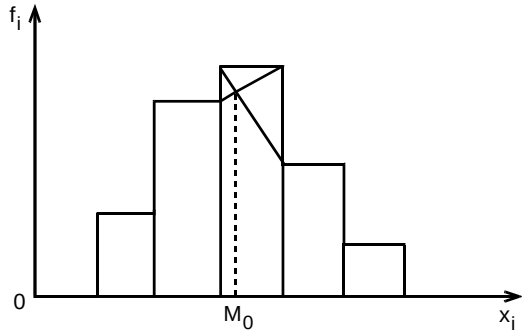
ÖRNEK: Aşağıdaki serinin modunu hesaplayınız.

X_i	f_i
1	8
2	13
3	20
4	14
5	16
6	19
7	26
8	14
9	10
	140

Serinin frekansları artarak 20 olduğunda ilk maksimumu, daha sonra azalıp tekrar artarak 26 olduğunda ikinci maksimumu yapmaktadır. Bu nedenle seri iki modludur. Serinin modları 3 ve 7 dir.

Burada seri iki modlu olduğundan mod uygun ölçü olma özelliğini kaybetmektedir.

Sınıflı serilerde ise mod, frekansı en yüksek olan sınıfta yer almaktadır. Bu nedenle frekansı en yüksek olan sınıf mod sınıfıdır. Modun hesaplanmasında, mod sınıfı ile birlikte, mod sınıfından bir önceki ve bir sonraki sınıfların frekansları alınarak enterpolasyon işlemi yapılır. Enterpolasyon işlemine, mod sınıfında yer alan birimlerin dağılımının, mod sınıfından bir önceki ve bir sonraki sınıfların frekanslarının etkisinde kaldığının varsayılması nedeniyle başvurulmaktadır. Bu varsayım nedeniyle mod bu sınıflardan frekansı yüksek olana daha yakın olacaktır. Şekil 3,3'de bir Sınıflı serinin modu görülmektedir.



Şekil 3.3

Sınıflı serilerde modun hesaplanmasında yapılacak enterpolasyon işlemi formül haline getirilebilir. Hesaplamada kullanılacak formül aşağıda verilmiştir.

$$M_o = I_a + \frac{\Delta_1}{\Delta_1 + \Delta_2} c$$

Formülde;
I_a : mod sınıfın alt sınırını

Δ_1 : mod sınıfın frekans ile bir önceki sınıf frekansı arasındaki farkı

Δ_2 : mod sınıfın frekans ile bir sonraki sınıf frekansı arasındaki farkı

c= sınıf aralığını

göstermektedir.

Bu formülün uygulanabilmesi için mod sınıfının, mod sınıfından bir önceki ve bir sonraki sınıfların sınıf aralıklarının eşit olması gerekmektedir. Mod sınıfından bir önceki ve bir sonraki sınıfların frekansları birbirine eşit olması durumunda mod, mod sınıfının orta değerine eşit olacaktır.

ÖRNEK: Aşağıdaki sınıflı serinin modunu hesaplayınız.

Sınıflar	F_i
0-3	4
3-6	9
6-9	13
9-12	6
12-15	3
	35

$$M_o = I_a + \frac{\Delta_1}{\Delta_1 + \Delta_2} c$$

Mod sınıfı = 6-9

$I_a = 6$

Δ_1 =13-9= 4

Δ_2 =13-6=7

c= 3

$$M_o = 6 + \frac{4}{4+7}.3 = 6 + \frac{12}{11} = 7{,}09$$

3.6. MEDYAN

Seriyi iki eşit parçaya bölen değere medyan denmektedir. Medyan seriyi iki eşit parçaya böldüğünden birimlerin % 50'sinin değeri medyandan küçük, % 50'sinin değeri medyan büyük olacaktır.

Toplanan verilerin küçükten büyüğe doğru sıralanmasıyla elde edilen basit serilerde medyanın hesaplanmasında seri birim sayısının tek veya çift sayıda olması önemlidir. Seri birimlerinin tek sayıda olması durumunda serinin $\frac{n+1}{2}$'inci biriminin değeri medyan olacaktır.

Seri birimlerinin çift sayıda olması durumunda ise, serinin $\frac{n+1}{2}$'inci birimi olmayacağından, medyan serinin $\frac{n}{2}$'inci ve $\frac{n}{2}$+1'inci birimlerinin değerlerinin ortalaması olarak kabul edilmektedir.

ÖRNEK: Aşağıda verilen serilerin medyanlarını hesaplayınız.

X_i : 2; 4; 7; 11; 13; 13; 17
Y_i: 3; 9; 13; 17; 19; 19; 23; 24

X_i serisi, 7 birimli bir seri olduğundan,

$$\frac{n+1}{2} = \frac{7+1}{2} = 4$$

olarak elde edileceğinden serinin medyanı 4. biriminin değeri olan 11 olacaktır.

Y_i serisi ise 8 birimli bir seri olduğundan,

$$\frac{n}{2} = \frac{8}{2} = 4$$

ve

$$\frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$$

olarak elde edileceğinden serinin medyanı 4'üncü ve 5'inci birimlerin değerlerinin ortalamasına eşit yani,

$$M_e = \frac{17+19}{2} = \frac{36}{2} = 18$$

olacaktır.

Frekans serilerinde medyanın bulunabilmesi için artan birikimli frekansların hesaplanması gerekir. Bu frekanslar hesaplandıktan sonra seri birim sayısının yarısı $\left(\frac{n}{2}\right)$ bulunarak, bu değerlerin artan birikimli frekanslar sütununda hangi aralıkta yer aldığı belirlenir. Bu şekilde belirlenen aralığın büyük değeri karşısındaki X değeri serinin medyan değerini verecektir.

ÖRNEK: Aşağıda verilen frekans serisinin medyanını hesaplayınız.

X_i	f_i	Artan Birikimli Frekanslar
10	8	8
11	13	21
12	19	40
13	15	55
14	12	67
16	3	70
	70	

$$\frac{n}{2} = \frac{70}{2} = 35$$

35 artan birikimli frekanslar sütununda 21-40 değerleri arasında yer aldığından, 40'ın karşısındaki 12 değeri serinin medyanıdır.

Frekans serilerinde medyanın bu şekilde hesaplanma nedenini, herhangi bir frekans serisi, basit seri şeklinde gösterilerek açıklanabilir.

ÖRNEK: Yukarıda verilen frekanslı seriyi basit seri şeklinde göstererek medyanı hesaplayınız.

$$\underbrace{10,\cdots,10}_{8},\underbrace{11,\cdots,11}_{13},\underbrace{12,\cdots,12}_{19},\underbrace{13,\cdots,13}_{15},\underbrace{14,\cdots,14}_{12},\underbrace{16,\cdots,16}_{3}$$

Bu basit serinin birim sayısı çift olduğundan,

$$\frac{n}{2}=\frac{70}{2}=35$$

$$\frac{n}{2}+1=\frac{70}{2}+1=36$$

35'inci ve 36'inci birimlerin değerlerinin ortalaması medyandır. 35'inci birim ve 36'inci birim 12 olduğundan, ortalamaları da 12 Olacaktır. Her iki şekilde hesaplanan medyan değerinin birbirine eşit olduğu görülmektedir.

Artan birikimli frekanslar sütununda seri birimlerinin yarısına yani $\frac{n}{2}$ değerine eşit olan değer yer alıyorsa bu birikimli frekansın X değeri ile bir sonraki x değerinin ortalamasının medyan olduğu kabul edilecektir.

ÖRNEK: Aşağıdaki serinin medyanını hesaplayınız.

X_i	f_i	Birikimli Frekanslar
10	6	6
18	9	15
21	15	30
37	18	48
42	12	60
	60	

30 artan birikimli frekanslar sütununda yer aldığından serinin medyanı

$$M_e=\frac{21+37}{2}=\frac{58}{2}=29$$

olacaktır. Önceki örnekte görüldüğü gibi frekans serisi basit seriye dönüştürüldüğünde, basit serinin $\frac{n}{2}=\frac{60}{2}=30$ 'uncu biriminin değeri 21 ve $\frac{n}{2}+1=30+1=31$ biriminin

değerinin 37 olarak elde edilecektir. Burada medyan değeri, iki değerin ortalaması alınarak 29 olarak bulunacaktır. Böylece her iki şekilde hesaplanan medyan değerinin birbirine eşit olduğu görülmektedir.

Sınıflı serilerde ise medyan enterpolasyon işlemi ile hesaplanmaktadır. Bu serilerde medyanın hesaplanması için önce medyan sınıfının belirlenmesi gerekir. Medyan sınıfını belirlemek için, önce $\frac{n}{2}$ değerinin artan birikimli frekanslar sütununda yer aldığı aralık bulunur ve bulunan bu aralığın büyük değeri karşısındaki sınıf serinin medyan sınıfı olarak kabul edilir. Görüldüğü gibi frekans serilerinde medyan bulunması için yapılan işlemler ile, sınıflı serilerde medyan sınıfının bulunması için yapılan işlemler aynıdır.

Sınıflı serilerde medyanın hesaplanmasında yapılacak enterpolasyon işlemi formül haline getirilebilir. Hesaplamada kullanılacak formül aşağıda verilmiştir.

$$M_e = I_a + \frac{\frac{n}{2} - \sum_{i=1}^{k-1} f_i}{f_m}.c$$

Formülde;

I_a: Medyan sınıfının alt sınırını

$\sum_{i=1}^{k-1} f_i$: Medyan sınıfından bir önceki sınıfın artan birikimli frekansını

f_m: medyan sınıfının frekansını

c : sınıf aralığını

göstermektedir.

ÖRNEK: Aşağıdaki sınıflı serinin medyanını hesaplayınız.

Sınıflar	f_i	Birikimli Frekanslar
0-5	16	16
5-10	24	40
10-15	33	73
15-20	28	101
20-25	19	120
	120	

$$M_e = I_a + \frac{\frac{n}{2} - \sum_{i=1}^{k-1} f_i}{f_m}.c$$

$$\frac{n}{2} = \frac{120}{2} = 60$$

Medyan Sınıfı = 10-15

$I_a = 10$

$$\sum_{i=1}^{2} f_i = 40$$

$f_m = 33$

$c = 5$

$$M_e = 10 + \frac{60 - 40}{33}.5 = 10 + \frac{100}{33} = 13{,}03$$

Örnekte de görüldüğü gibi sınıflı serilerde medyan hesaplanırken enterpolasyon işlemi medyan sınıfı için yapılmaktadır. Bu nedenle sınıf aralıklarının eşit olması zorunluluğu yoktur. Medyanın bu özelliği yanında diğer bir özelliğide, seri birimlerinin değerlerinin medyandan farklarının mutlak değerleri toplamının minimum olmasıdır. Bu özellik;

Basit seriler için

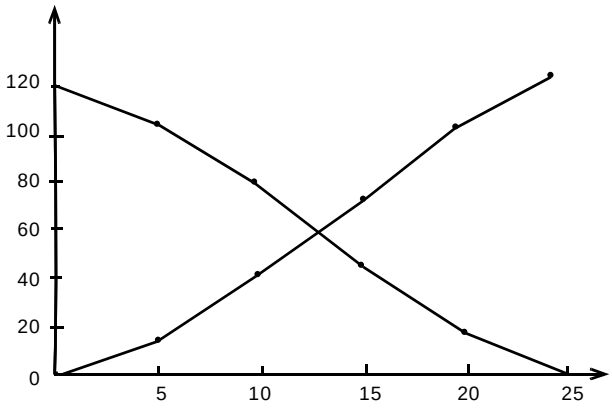
$$\sum |X_i - M_e| < \sum |X_i - A| \qquad A \neq M_e$$

Frekans serileri ve sınıflı seriler için ise,

$$\sum f_i |X_i - M_e| < \sum f_i |X_i - A| \qquad A \neq M_e$$

olacaktır.

Frekans serilerinde medyan formül ile hesaplanabildiği gibi artan ve azalan birikimli frekans eğrileri yardımı ile de hesaplanabilir. Bu hesaplamada artan ve azalan birikimli frekans eğrilerinin kesişim noktası medyanı vermektedir. Yukarıda verilen örneğin, artan ve azalan birikimli frekans eğrileri ve medyanı şekil 3.4 de görülmektedir.



Şekil 3.4

3.7. BÖLENLER

Serileri iki, dört, on ve yüz eşit parçaya ayıran değerler genel olarak bölenler olarak adlandırılmaktadır. Seriyi iki eşit parçaya bölmek için hesaplanan değerlere medyan, dört eşit parçaya bölmek için hesaplanan değerlere kartil, on eşit parçaya bölmek için hesaplanan değerlere desil ve yüz eşit parçaya bölmek için hesaplanan değerlere santil adı verilmektedir. Bir seri için 3 kartil, 9 desil ve 99 santil hesaplanmaktadır. Ancak uygulamalarda desil ve santil yaygın olarak kullanılmadığından ve medyanda daha önce açıklanmış olduğundan burada sadece kartiller anlatılacaktır. Seriyi dört eşit parçaya bölen kartiller birinci kartil, ikinci kartil ve üçüncü kartil olarak isimlendirilirler.

Seri birimlerinin n/4'ünün değeri birinci kartilden, n/2'sinin değeri ikinci kartilden ve 3n/4 'ünün değeri üçüncü kartilden küçük olacaktır. Birinci kartil $Q_{1/4}$, ikinci kartil $Q_{2/4}$, üçüncü kartil ise $Q_{3/4}$ şeklinde ifade edilmektedir. Serinin ikinci kartili aynı zamanda meydanıdır.

Basit serilerde kartiller medyana benzer şekilde hesaplanmaktadır. Yani seriyi ilk birimden başlayarak n/4 son birimden başlayarak 3n/4 şeklinde bölen değer birinci kartil, ilk birimden başlayarak 3n/4, son birimden başlayarak n/4 şeklinde bölen değer ise üçüncü kartil olmaktadır.

Frekans serilerinde kartillerin bulunabilmesi için medyanda olduğu gibi artan birikimli frekansların hesaplanması gerekir. Bu frekanslar hesaplandıktan sonra birinci kartil için seri birimlerinin n/4 değeri bulunarak bu değerin artan birikimli frekanslar sütununda hangi aralıkta yer aldığı belirlenir. Bu şekil de belirlenen aralığın büyük değeri karşısındaki X değeri serinin birinci kartil değerini verecektir. Üçüncü kartilin hesabı içinde aynı işlemler seri birimlerinin 3n/4 değeri bulunarak yapılmaktadır.

ÖRNEK: Aşağıdaki frekans serisinin kartillerini hesaplayınız.

X_i	f_i	Artan Birikimli Frekanslar
20	8	8
21	23	31
22	32	63
23	36	99
24	27	126
25	14	140
	140	

$$\frac{n}{4} = \frac{140}{4} = 35$$

35 artan birikimli frekanslar sütununda 31-63 aralığında bulunduğundan $Q_{1/4}$ =22,

$$\frac{n}{2} = \frac{140}{2} = 70$$

70 artan birikimli frekanslar sütununda 63-99 aralığında bulunduğundan $Q_{1/2}$ =Me=23,

$$\frac{3n}{4} = \frac{3(140)}{4} = 105$$

105 artan birikimli frekanslar sütununda 99-126 aralığında bulunduğundan $Q_{3/4}$ =24 olacaktır.

Sınıflı serilerde kartillerin hesaplanmasında medyanın hesabında kullanılan formüle benzer bir formül ile enterpolasyon işlemi yapılmaktadır. medyanın hesabı için kullandığımız formülü kartillerin hesabı için şöyle genelleştirebiliriz.

$$Q_{h/r} = I_a + \frac{n\frac{h}{r} - \sum_{i=1}^{k-1} f_i}{f_b}.c$$

Yukarıda görülen bu formül kartillerin hesaplanmasında kullanılırken r = 4 olacaktır. Aynı formül desil ve santiller hesaplamasında da kullanılmaktadır. Ancak formül desiller için kullanılırsa r =10 ve santiller için kullanılırsa r = 100 olacaktır. Formülde görülen h ise hesaplanan bölene göre 1,2,3,... değerlerini almaktadır. Örneğin 3. kartil için h= 3, 6. desil için h= 6 ve 85. santil için h= 85 olacaktır.

Diğer taraftan hesaplanan bölene göre formülde;

I_a: Bölen sınıfının alt sınırını

$\sum_{i=1}^{k-1} f_i$: Bölen sınıfından bir önceki sınıfın birikimli frekansını

f_b: Bölen sınıfının frekansını
c: Sınıf aralığını ifade etmektedir.

ÖRNEK: Aşağıda verilen sınıflı serinin kartillerini hesaplayınız.

Sınıflar	f_i	Birikimli Frekanslar
100-120	22	22
120-140	37	59
140-160	43	102

160-180	28	130
180-200	20	150
	150	

$$Q_{h/r} = I_a + \frac{n\frac{h}{r} - \sum_{i=1}^{k-1} f_i}{f_b}.c$$

1. Kartil

$$\frac{n}{4} = \frac{150}{4} = 37{,}5$$

olduğundan 120-140 sınıfı 1. kartil sınıfıdır.

$$Q_{1/4} = 120 + \frac{37{,}5 - 22}{37}.20 = 120 + \frac{310}{37} = 128{,}37$$

2. Kartil (Medyan):

$$\frac{2n}{4} = \frac{n}{2} = \frac{150}{2} = 75$$

olduğundan 140-160 sınıfı 2. Kartil veya medyan sınıfıdır.

$$Q_{2/4} = M_e = 140 + \frac{75 - 59}{43}.20 = 140 + \frac{320}{43} = 147{,}44$$

3. Kartil:

$$\frac{3n}{4} = \frac{3(150)}{4} = 112{,}5$$

olduğundan 160-180 sınıfı 3. Kartil sınıfıdır.

$$Q_{3/4} = 160 + \frac{112{,}5 - 102}{28}.20 = 160 + \frac{210}{28} = 167{,}50$$

3.8. UYGUN ORTALAMANIN SEÇİMİ

Aritmetik ortalama uygulamalarda en çok kullanılan ortalamadır. Tek olarak kullanılan ortalama kelimesi aritmetik ortalama anlamında kullanılmaktadır. Aritmetik

ortalamanın hesaplanmasında serinin bütün değerleri kullanıldığından seride aşırı düşük veya yüksek değerlerin bulunması durumunda aritmetik ortalama bu değerlerden etkilendiğinden, seriyi temsil etme özelliğini kaybedebilir.

Geometrik ortalama aritmetik, harmonik ve kareli ortalamalara göre serideki aşırı değerlerden daha az etkilenmektedir. Bu nedenle serinin bütün değerlerinin kullanıldığı bir ortalama hesaplanacaksa ve seride aşırı değerler varsa, geometrik ortalamanın kullanılması uygun olabilir. Seride sıfır ve negatif değerlerin bulunması durumunda geometrik ortalama hesaplanamaz. Seride sıfır değerleri varsa harmonik ortalama hesaplanamazken ters işaretli birimler varken harmonik ortalama hatalı sonuçlar verebilir.

Kareli ortalama hesaplanırken serinin değerlerinin kareleri alındığından seride ters işaretli birimler bulunduğunda kareli ortalama hesaplanması uygun olmaktadır. Seri birimlerinin tersleri söz konusu olduğunda harmonik ortalama hesaplanır.

Bir seri için harmonik, aritmetik, geometrik ve kareli ortalama birlikte hesaplandığında, en küçük ortalama harmonik ortalama, en büyük ortalama ise kareli ortalamadır. Bu durumda

$$H < G < \overline{X} < K$$

yazılabilir.

Mod ve medyanın hesaplanmasında serinin bütün değerleri kullanılmadığından, seride aşırı düşük veya yüksek değerler bulunduğunda mod veya medyanın kullanılmaları uygun olabilir. Ayni nedenle, seride açık sınıflar bulunduğunda mod veya medyan diğer ortalamalara göre daha uygun ölçü olmaktadır. Çünkü bu gibi durumlarda, aritmetik, geometrik ve kareli ortalamalardan biri kullanılacaksa, açık sınıfların sınıf orta noktalarının belirlenmesi sorun yaratabilir.

3.9. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: İşletmenin 10 günlük buzdolabı satışları aşağıda verilmiştir.

Günler	Satışlar (Adet)
1	8
2	13
3	17
4	11
5	9
6	16
7	21
8	23
9	14
10	18

Buzdolabı satışlarının,

a) Aritmetik ortalamasını

b) Kareli ortalamasını

c) Geometrik ortalamasını hesaplayınız.

d) Harmonik ortalama

e) Hesapladığınız ortalamaları karşılaştırınız.

Çözüm:

X_i	X_i^2	$\log X_i$	$1X_i$
8	64	0,90308	0,1250
13	169	1,11394	0,0769
17	289	1,23044	0,0588
11	121	1,04139	0,0909
9	81	0,95424	0,1111
16	256	1,20411	0,0625
21	441	1,32221	0,0476
23	529	1,36172	0,0434
14	196	1,14612	0,0714
18	324	1,25527	0,0555
150	2470	11,53252	0,7431

a) $\overline{X} = \dfrac{\sum X_i}{n} = \dfrac{150}{10} = 15$

b) $K = \sqrt{\dfrac{\sum X_i^2}{n}} = \sqrt{\dfrac{2470}{10}} = \sqrt{247} = 15{,}71$

c) $G = \sqrt[n]{\prod_{i=1}^{n} X_i}$

$$\log G = \frac{1}{n}\sum \log X_i = \frac{1}{10}(11{,}53252) = 1{,}153252$$

G = 14,23

d) $H = \dfrac{n}{\sum \frac{1}{x_i}} = \dfrac{10}{0{,}7431} = 13{,}45$

e) 13,45<14,23<15<15,7

$$H < G < \overline{X} < K$$

ÖRNEK 2: Bir malın 40 ayrı satış yerindeki satış fiyatları aşağıda verilmiştir. (1= 1 TL)

Fiyat	Satış Yeri Sayısı
2000	2
2200	14
2250	9
2400	8
2500	5
2600	2
	40

Satış fiyatının,
a) Aritmetik ortalamasını
b) Kareli ortalamasını
c) Geometrik ortalamasını hesaplayınız.
d) Harmonik ortalamasını hesaplayınız.
e) Hesapladığınız ortalamaları karşılaştırınız.

Çözüm:

X_i	f_i	$f_i X_i$	$f_i X_i^2$	$\log X_i$	$f_i \log X_i^2$	f_i / X_i
2000	2	4000	8000000	3,30103	6,60206	0,0010
2200	14	30800	67760000	3,34242	46,79388	0,0064
2250	9	20250	45562500	3,35218	30,16962	0,0040
2400	8	19200	46080000	3,38021	27,04168	0,0033
2500	5	12500	31250000	3,39794	16,98970	0,0020
2600	2	5200	13520000	3,41497	6,82994	0,0008
	40	91950	212172500			0,0175

a) $\overline{X} = \dfrac{\sum f_i X_i}{\sum f_i} = \dfrac{91950}{40} = 2298{,}75$

b) $K = \sqrt{\dfrac{\sum f_i X_i^2}{f_i}} = \sqrt{\dfrac{212172500}{40}} = \sqrt{5304312{,}5} = 2303{,}10$

c) $G = \sqrt[\Sigma f_i]{\prod X_i^{f_i}}$

$$\log G = \frac{1}{\sum f_i} \sum f_i \log X_i = \frac{1}{40}(134{,}42688) = 3{,}360672$$

G = 2294,41

d) $H = \dfrac{\sum f_i}{\sum f_i / x_i} = \dfrac{40}{0{,}0175} = 2285{,}71$

e) 2285,71<2294,41<2298,75<2303,10

$$H < G < \overline{X} < K$$

ÖRNEK 3: Bir işyerinde satılan ve 1 kg. olması gereken 700 paket tartılmış ve paketlerin ağırlıklarının eksik veya fazla olduğu belirlenmiştir. Tartılan paketlerin ağırlık farkları aşağıda verilmiştir.

Ağırlık Farkları	f_i
-15 - -10	126
-10 - -5	220
-5 – 0	171
0 – 5	108
5 –10	61
10 – 15	14
	700

Ağırlık farkları dağılımının,
a) Aritmetik ortalamasını
b) Kareli ortalamasını
c) Geometrik ortalamasını hesaplayınız.
d) Hesapladığınız ortalamaları karşılaştırınız.

Çözüm:

Ağırlık Farkları	f_i	X_i	$f_i X_i$	$f_i X_i^2$
-15 - -10	126	-12,5	-1575,0	19687,5
-10 - -5	220	-7,5	-1650,0	12375,00
-5 – 0	171	-2,5	-427,5	1068,75
0 – 5	108	2,5	270,0	675,00
5 –10	61	7,5	457,5	3431,25
10 – 15	14	12,5	175,0	2187,50
	700		-2750,0	39425,00

a) $\overline{X} = \dfrac{\sum f_i X_i}{\sum f_i} = \dfrac{-2750}{700} = -3{,}92$

$$K = \sqrt{\frac{\sum f_i X_i^2}{f_i}} = \sqrt{\frac{39425}{700}} = \sqrt{56{,}32} = 7{,}50$$

b) Seride negatif işaretli birimler olduğundan geometrik ortalama hesaplanamaz.

c) - 3,92 < 7,50

$$\overline{X} < K$$

ÖRNEK 4: Aynı holdingin bünyesinde yer alan 7 işletmenin 2004 yılı karları aşağıda verilmiştir.

İşletmeler	Karlar (1000 TL)
A	1,2
B	3,4
C	8,6
D	0,8
E	7,3
F	9,7
G	4,1

Serinin,
a) Modunun
b) Medyanını
c) Kartillerini hesaplayınız.

Çözüm: X_i : 0,8 ; 1,2 ; 3,4 ; 4,1 ; 7,3 ; 8,6 ; 9,7

a) Seri modsuzdur.

b) $\dfrac{n+1}{2} = \dfrac{7+1}{2} = 4$ $\qquad M_e = 4{,}1$

c) $\dfrac{n}{4} = \dfrac{7}{4} = 1{,}75$ $\qquad Q_1 = \dfrac{0{,}8+1{,}2}{2} = 1$

$Q_2 = M_e$

d) $\dfrac{3n}{4} = \dfrac{3(7)}{4} = 5{,}25$ $\qquad Q_3 = \dfrac{7{,}3+8{,}6}{2} = 7{,}95$

ÖRNEK 5: Bir belediyenin son bir ay içinde verdiği inşaat ruhsatlarının kat sayılarına göre dağılımı aşağıda verilmiştir.

Kat Sayıları	Verilen Ruhsat Sayısı
1	19
2	28

3	33
4	50
5	122
6	71
7	34
8	12
	369

Dağılımın
a) Modunun
b) Medyanını
c) Kartillerini hesaplayınız.

Çözüm:

X_i	f_i	Artan Birikimli Frekanslar
1	19	19
2	28	47
3	33	80
4	50	130
5	122	252
6	71	323
7	34	357
8	12	369
	369	

a) Seri tek modlu olduğundan mod uygun ölçü olacaktır. En yüksek frekans 122 olduğundan serinin modu,

$M_o = 5$

bulunur.

b) $\frac{n}{2} = \frac{369}{2} = 184{,}5$

$M_e = 5$

c) $\frac{n}{4} = \frac{369}{4} = 92{,}25$

$Q_{1/4} = 4$

$Q_{2/4} = M_e = 5$

$$\frac{3n}{4} = \frac{3(369)}{4} = 276{,}75$$

$Q_{3/4} = 6$

ÖRNEK 6: Bir banka şubesinin Ocak ayında 90 kişiye verdiği tüketici kredilerinin miktarlarına göre dağılımı aşağıda verilmiştir. (1 = 1.000 T.L.)

Sınıflar	f_i
2,5 - 7,5	6
7,5 - 12,5	15
12,5 - 17,5	18
17,5 - 22,5	27
22,5 - 27,5	13
27,5 - 32,5	11
	90

Kredilerin dağılımının,
a) Modunu
b) medyanını
c) Kartillerini hesaplayınız.

Çözüm:

Sınıflar	f_i	Artan Birikimli Frekanslar
2,5 - 7,5	6	6
7,5 - 12,5	15	21
12,5 - 17,5	18	39
17,5 - 22,5	27	66
22,5 - 27,5	13	79
27,5 - 32,5	11	90
	90	

a) $M_o = I_a + \frac{\Delta_1}{\Delta_1 + \Delta_2} c$

Mod Sınıfı : 17,5 – 22,5

Δ_1 = 27 – 18 = 9

Δ_2 = 27 – 13 = 14

c= 5

$$M_o = 17{,}5 + \frac{9}{9+14}.5$$

$$= 17{,}5 + \frac{45}{23} = 19{,}45$$

b) $M_e = I_a + \dfrac{\dfrac{n}{2} - \sum_{i=1}^{k-1} f_i}{f_m}.c$

Medyan Sınıfı: 17,5 22,5

$\sum_{i=1}^{k-1} f_i = 39$

$f_m = 27$

c = 5

$$M_e = 17{,}5 + \frac{45-39}{27}.5$$

$$= 17{,}5 + \frac{30}{27} = 18{,}61$$

c) $Q_{h/r} = I_a + \dfrac{n\dfrac{h}{r} - -\sum_{i=1}^{k-1} f_i}{f_b}.c$

$$\frac{n}{4} = \frac{90}{4} = 22{,}5$$

1. Kartil Sınıfı : 12,5 – 17,5

$\sum_{i=1}^{k-1} f_i = 21$

$f_b = 18$

c = 5

$$Q_{1/4} = 12{,}5 + \frac{22{,}5-21}{18}.5$$

$$= 12{,}5 + \frac{7{,}5}{18} = 12{,}91$$

$Q_{2/4} = M_e = 18,61$

$$\frac{3n}{4} = \frac{3(90)}{4} = 67,5$$

3. Kartil Sınıfı : 22,5 – 27,5

$$\sum_{i=1}^{k-1} f_i = 66$$

$f_b = 13$
$c = 5$

$$Q_{3/4} = 22,5 + \frac{67,5 - 66}{13}.5$$

$$= 22,5 + \frac{7,5}{13} = 23,07$$

ÖRNEK 7: Bir işletme işe alacağı personel için üç sınav yapmakta ve sınavda başarılı olanları mülakata almaktadır. İşletmeye başvuran bir kişiye ait sonuçlar ile tartılar aşağıda verilmiştir.

Testler	Alınan Puanlar	Tartılar
Genel Yetenek	69	2
Bilim	84	4
Yabancı Dil	76	1
Mülakat	70	3

Verilen puanların tartılı aritmetik ortalamasını hesaplayınız.

Çözüm:

X_i	t_i	$X_i t_i$
69	2	138
84	4	336
76	1	76
70	3	210
	10	760

$$\overline{X}_t = \frac{\sum_{i=1}^{n} X_i t_i}{\sum_{i=1}^{n} t_i}$$

$$\overline{X}_t = \frac{760}{10} = 76$$

ÖRNEK 8: Bir malın ortalama satış fiyatini belirlemek için 42 satış yerinden satış fiyatları ve günlük satış miktarları alınmıştır.

Fiyatlar (1=1 TL)	Satış Yeri Sayısı	Satış Miktarları
58	6	462
60	14	541
62	17	214
65	4	172
70	1	88
	42	1477

Satış miktarlarını tartı alarak, malin satış fiyatının tartılı aritmetik ortalamasını hesaplayınız.

Çözüm:

X_i	f_i	t_i	$f_i t_i$	$f_i X_i t_i$
58	6	462	2772	160776
60	14	541	7574	454440
62	17	214	3638	227375
65	4	172	688	44720
70	1	88	88	6160
	42	1477	14760	893471

$$\overline{X}_t = \frac{\sum_{i=1}^{k} t_i f_i X_i}{\sum_{i=1}^{k} t_i f_i}$$

$$\overline{X}_t = \frac{893471}{14760} = 60{,}5332$$

ÖRNEK 9: Bir mamulün 105 imalat yerindeki maliyetlerinin dağılımları ile, imalat yerlerinin günlük üretim miktarlarının toplamları aşağıda verilmiştir.

Maliyet Sınıfları (1=1 TL)	İmalat Yeri Sayısı	Günlük Üretim Miktarları (1=1000 adet)
5,5 - 6,0	27	158
6,0 - 6,5	35	113
6,5 - 7,0	29	71
7,0 -7,5	8	18
7,5 - 8,0	6	14
	105	

Mamulün maliyetinin tartılı aritmetik ortalamasını üretim miktarlarını tartı olarak hesaplayınız.

Çözüm:

Sınıflar	f_i	t_i	X_i	f_it_i	$f_iX_it_i$
5,5 - 6,0	27	158	5,75	4266	24529,50
6,0 - 6,5	35	113	6,25	3955	24718,78
6,5 - 7,0	29	71	6,75	2059	13898,25
7,0 -7,5	8	18	7,25	144	1044,00
7,5 - 8,0	6	14	7,75	84	651,00
	105			10508	64841,5

$$\overline{X}_t = \frac{\sum_{i=1}^{k} t_i f_i X_i}{\sum_{i=1}^{k} t_i f_i}$$

$$\overline{X}_t = \frac{64841{,}50}{10508} = 6{,}17$$

ÖRNEK 10: Bir işletmede 128 işçi çalışmaktadır ve ortalama aylık ücret 2,856 TL. Olarak hesaplanmıştır. İşçi Ücretlerine yapılacak zam için, işçiler aylık 1,5 TL. artı % 45 zam istemekte, işveren ise aylık 2.- TL artı % 25 zam teklif etmektedir.

a) Her iki durumda ortalama aylık ücreti ve aylık ortalama artış oranlarını

b) Her iki durumun işverene getireceği aylık maliyet yükünü ve toplam ücret artış oranlarını hesaplayınız.

Çözüm:

a) Yeni ortalamayı $\overline{Y}$ ile ifade edelim.

1,5 + % 45 zam yapılırsa, eski ortalama 1,5 TL ve eski ortalamanın $1+0{,}45=1{,}45$ katı artış olacağından,

$$\begin{aligned}\overline{Y} &= 1{,}5+1{,}45\overline{X}\\ &= 1{,}5+1{,}45(2{,}856)\\ &= \underbrace{5{,}6412.-TL}_{5{,}6412}\end{aligned}$$

Artış oranı = $\dfrac{5{,}6412-2{,}856}{2{,}856}=0{,}9752$ olacaktır.

2 + % 25 zam yapılırsa, aynı şekilde,

$$\begin{aligned}\overline{Y} &= 2+1{,}25\overline{X}\\ &= 2+1{,}25(2{,}856)\\ &= 5{,}57.-TL\end{aligned}$$

Artış oranı = $\dfrac{5{,}57-2{,}856}{2{,}856}=0{,}9502$ olacaktır.

b) 128 işçiye ödenen toplam aylık ücret,

$$\sum_{i=1}^{128} X_i = 128(2{,}856) = 365{,}568.-TL \text{ dir.}$$

1,5.- TL + % 45 zam yapılırsa ortalama ücret 5641200.- TL. olacağından ödenecek toplam ücret,

$$\sum_{i=1}^{128} Y_i = 128(5{,}6412) = 722{,}0736.-TL$$

ve 2 + % 25 zam yapılırsa ortalama ücret 5,57.- TL olacağından ödenecek toplam ücret,

$$\sum_{i=1}^{128} Y_i = 128(5{,}57) = 712{,}96.-TL \text{ olacaktır.}$$

İşverene birinci durum,

722,0736 - 365,568 = 356,5056.- TL.

İkinci durum ise,

712,96 - 365,568 = 347,392.- TL.

aylık yük getirecektir.

$$\text{Birinci Durumda Artış Oranı} = \frac{356,5056}{365,568} = 0,9752$$

$$\text{İkinci Durumda Artış Oranı} = \frac{347,392}{365,568} = 0,9502$$

olacaktır.

BÖLÜM 4

DEĞİŞKENLİK VE ASİMETRİ ÖLÇÜLERİ

Bir önceki bölümde serileri temsil etmek ve birbirleri ile karşılaştırmak amacı ile hesaplanan ortalamalar açıklanmıştı. Ortalamalar, serilerin karşılaştırılmasında her zaman yeterli ölçüler olmayabilirler. Bu nedenle serilerin karşılaştırılmasında, dağılma ölçüleri olarak da adlandırılan değişkenlik ölçülerinden ve serilerin dağılma şekillerinin karşılaştırılmasında da eğrilik ölçüleri olarak da adlandırılan asimetri ölçülerinden yararlanılmaktadır. Asimetri ölçülerinin hesaplanmasında, değişkenlik ölçüleri de kullanılacağından, bu bölümde asimetri ölçülerine dağılma ölçülerinden sonra yer verilmiştir.

4.1. DEĞİŞKENLİK ÖLÇÜLERİ

Serilerin karşılaştırılmalarında ortalamaların yeterli olmadığı durumlarda değişkenlik ölçülerinden yararlanılacağından daha önce söz edilmişti. Çünkü bazı durumlarda serilerin ortalamaları birbirine yakın değerde olsa da serilerin dağılımları birbirinden farklı olabilir. Bu farkı belirleyebilmek için değişkenlik ölçülerinden yararlanılır.

Konuyu basit bir örnekle açıklayalım.

X_i ve Y_i iki farklı, basit seri olsun.

X_i	Y_i
0	2
1	2
2	3
3	4
9	4
15	15

Bu serilerin aritmetik ortalamaları

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{15}{5} = 3$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{15}{5} = 3$$

olacaktır. Yukarıda görülen bu serilerin verilmediğini ve sadece aritmetik ortalamalarının bilindiğini varsayalım. Seriler verilmediğinden, dağılımlarının birbirine yakın veya birbirinden farklı olabileceği düşünülebilir. Bu dağılımların farkını belirlemek için değişkenlik ölçülerini hesaplamamız gerekmektedir.

Değişkenlik ölçüleri seri değerlerinin dağılmalarının ölçüsüdür. Bu ölçüler seri değerlerinin ortalama etrafında ne kadar yoğun dağıldıklarını belirtirler. Değişkenlik ölçülen nispi ölçüler olup her hangi bir seri için hesaplandıklarında fazla bir anlam ifade etmezler. Bu ölçüler ancak birbirleri ile karşılaştırıldıklarında, hesaplandıkları serilerin hangisinde dağılımın ortalama etrafında daha yoğun olduğunu belirtirler. Diğer bir ifade ile değişkenlik ölçüsü küçük olan seriler, değişkenlik ölçüsü büyük olan serilere göre ortalama etrafında daha yoğun bir dağılım gösterirler. Bu nedenle değişkenlik ölçüleri analizde kullanılan verinin homojenliği ile de ilgilidirler. Homojenlik seri değerlerinin birbirine yakınlığını ifade etmektedir. Değişkenlik ölçüsü küçük olan seriler diğer serilere göre daha homojendirler.

Serilerin değişkenliklerini belirlemek için çeşitli değişkenlik ölçüleri kullanılabilir. Ancak bu bölümde değişkenlik ölçülerinden, hesaplanması en kolay olan değişim aralığı ile istatistiksel analizlerde en çok kullanılan ölçü olan standart sapma anlatılacaktır.

4.1.1. Değişim Aralığı

Değişim veya dağılma aralığı olarak adlandırılan bu ölçünün hesaplanması için uzun matematiksel işlemler gerekmez. Bu nedenle iki veya daha fazla serinin değişkenliklerini karşılaştırmada kullanılabilir. Fakat değişim aralığı serilerin dağılımını belirlemede kesin bir ölçü değildir.

Değişim aralığı bir serinin maksimum ve minimum değerleri arasındaki farka eşittir. Değişim aralığı ,

$D.A = X_{max} - X_{min}$

şeklinde gösterilebilir.

Açıklamalarımızın başında verilen örnekte $\overline{X}$ = 3 ve $\overline{Y}$ = 3 olarak bulunmuştu. Bu serilerin değişim aralıkları ise,

$DA_x = X_{max.} - X_{min.} = 9 - 0 = 9$

$DA_y = Y_{max.} - Y_{min.} = 4 - 2 = 2$

olarak hesaplanmıştır.

Serilerin ortalamaları ile değişim aralıklarıda bilindiğinden, Y serisinde dağılımın X serisine göre ortalama etrafında daha yoğun olduğunu söyleyebiliriz.

Değişim aralığı, frekans serilerinde X_i sütununun maksimum ve minimum değerleri arasındaki farka, sınıflı serilerde ise ilk sınıfın alt sınır değeri ile son sınıfın üst sınır değeri arasındaki farka eşittir.

Değişim aralığı aynı, fakat dağılımları birbirinden farklı basit, frekanslı veya sınıflı seriler olabilir. Örneğin daha önce verilen örnekle ilgili bir Z serisi Z_i: 2,3,3,3,4 şeklinde yazılabilir. Bu serinin değişim aralığı DA_2 = 4-2=2 dir ve Y serisinin değişim aralığına eşittir. Oysa serilerin dağılımları birbirinden farklıdır. Bu nedenle değişim aralığı kesin bir ölçü değildir.

4.1.2.Standart Sapma

Seri değerlerinin aritmetik ortalamadan farklarının kareli ortalamasına standart sapma denilir. Standart sapma σ (küçük sigma) ile gösterilir. Standart sapmanın karesine varyans adi verilir. Varyans V(X) veya σ^2 sembolleri ile ifade edilir.

Basit serilerde standart sapma

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}}$$

formülü ile hesaplanmaktadır.

Verilen bu formülün açılarak, kısaltılması ile

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}}$$

$$= \sqrt{\frac{\sum\left(X_i^2 - 2X_i\overline{X}\right) + \overline{X}^2}{n}}$$

$$= \sqrt{\frac{\sum X_i^2}{n} - \frac{2\overline{X}\sum X_i}{n} + \frac{\sum \overline{X}^2}{n}}$$

$$= \sqrt{\frac{\sum X_i^2}{n} - 2\overline{X}\overline{X} + \frac{n\overline{X}^2}{n}}$$

$$= \sqrt{\frac{\sum X_i^2}{n} - 2\overline{X}^2 + \overline{X}^2}$$

$$= \sqrt{\frac{\sum X_i^2}{n} - (\overline{X})^2}$$

ifadesi elde edilmiştir.

İfadede yer alan $\frac{\sum_{i=1}^{n} X_i^2}{n}$ kareli ortalamanın karesi olduğundan

$$\sigma = \sqrt{K^2 - \overline{X}^2}$$

olarak yazılabilir. Standart sapma hesaplanırken farkların kareleri alındığından standart sapma negatif olamaz. Daha önce aritmetik ortalamanın özellikleri açıklanırken, seri birimlerinin aritmetik ortalamadan farklarının kareleri toplamının minimum olduğu gösterilmişti. Standart sapma aritmetik ortalamanın bu özelliğine dayanarak hesaplanan bir ölçüdür.

ÖRNEK:

X_1 : 3, 5, 6, 9, 9, 10

serisinin standart sapmasını ve varyansını hesaplayınız.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{3+5+6+9+9+10}{6} = \frac{42}{6} = 7$$

$$\sigma = \sqrt{\frac{\sum \left(X_i - \bar{X}\right)^2}{n}}$$

$$\sigma = \sqrt{\frac{(3-7)^2 + (5-7)^2 + (6-7)^2 + (9-7)^2 + (9-7)^2 + (10-7)^2}{6}}$$

$$= \sqrt{\frac{(-4)^2 + (-2)^2 + (-1)^2 + (2)^2 + (2)^2 + (3)^2}{6}}$$

$$= \sqrt{\frac{16+4+1+4+4+9}{6}}$$

$$= \sqrt{\frac{38}{6}} = \sqrt{6,3333} = 2,1566$$

veya

$$\sigma = \sqrt{K^2 - \bar{X}^2}$$

$$K = \sqrt{\frac{\sum X_i^2}{n}}$$

$$K^2 = \frac{\sum X_i^2}{n}$$

$$K^2 = \frac{3^2 + 5^2 + 6^2 + 9^2 + 9^2 + 10^2}{6}$$

$$= \frac{9+25+36+81+81+100}{6}$$

$$= \frac{332}{6} = 55,3333$$

$$\sigma = \sqrt{K^2 - \bar{X}^2}$$

$$= \sqrt{55,3333 - (7)^2}$$

$$= \sqrt{55,3333 - 49}$$

$$= \sqrt{6,3333} = 2,5166$$

Her iki formül ile elde edilen sonuçların aynı olduğu görülmektedir.

Serinin varyansı ise,

$$V(X) = \sigma^2 = (2{,}5166)^2 = 6{,}3333$$

olacaktır.

Standart sapması hesaplanacak seri frekans serisi ise, seri birimlerinin değerlerinin aritmetik ortalamadan farklarının kareleri toplamı

$$\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2$$

ve $n = \sum_{i=1}^{k} f_i$ olacağından standart sapma,

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}}$$

olarak elde edilmektedir. Bu formülün açılarak, kısaltılması ile

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$= \sqrt{\frac{\sum f_i \left(X_i^2 - 2X_i\overline{X} + \overline{X}\right)^2}{\sum f_i}}$$

$$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \frac{\sum f_i 2X_i\overline{X}}{\sum f_i} + \frac{\sum f_i \overline{X}^2}{\sum f_i}}$$

$$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \frac{2\overline{X}\sum f_i X_i}{\sum f_i} + \frac{\overline{X}^2 \sum f_i}{\sum f_i}}$$

$$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - 2.\overline{X}.\overline{X} + \overline{X}^2}$$

$$= \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \overline{X}^2}$$

ifadesi elde edilmiştir. Bu ifadede,

$\frac{\sum f_i X_i^2}{\sum f_i}$ kareli ortalamanın karesi olduğundan,

$$\sigma = \sqrt{K^2 - \overline{X}^2}$$

olarak yazılabilir. Böylece formülün frekans serileri içinde geçerli olduğu görülmektedir.

ÖRNEK: Aşağıdaki frekans serisinin standart sapmasını hesaplayınız.

X_i	f_i	$f_i X_i$	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	$f_i\left(X_i - \overline{X}\right)^2$	$f_i X_i^2$
2	3	6	-2,5	6,25	18,75	12
3	6	18	-1,5	2,25	13,50	54
4	7	28	-0,5	0,25	1,75	112
5	16	80	0,5	0,25	4,00	400
6	8	48	1,5	2,25	18,00	288
	40	180			56,00	866

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{180}{40} = 4{,}5$$

$$\sigma = \sqrt{\frac{\sum f_i\left(X_i - \overline{X}\right)^2}{\sum f_i}} = \sqrt{\frac{56}{40}}$$

$$= \sqrt{1{,}4} = 1{,}18$$

$$\sigma = \sqrt{K^2 - \overline{X}^2} = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \overline{X}^2}$$

$$= \sqrt{\frac{866}{40} - (4{,}5)^2} = \sqrt{21{,}65 - 20{,}25}$$

$$= \sqrt{1{,}40} = 1{,}18$$

$V(X) = \sigma^2 = 1,4$

Her iki formül ile elde edilen sonuçların aynı olduğu görülmektedir.

Sınıflı serilerde standart sapma frekans serilerinde kullanılan

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma = \sqrt{K^2 - \overline{X}^2} = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}}$$

formülleri ile hesaplanır. Ancak formüllerde görülen X_i yerine sınıf orta değerleri kullanılmaktadır.

ÖRNEK: Aşağıdaki sınıflı serinin standart sapmasını hesaplayınız.

Sınıflar	f_i	X_i	$f_i X_i$	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	$f_i \left(X_i - \overline{X}\right)^2$	$f_i X_i^2$
1-3	3	2	6	-3,5	12,25	36,75	12
3-5	5	4	20	-1,5	2,45	12,25	80
5-7	7	6	42	0,5	0,25	1,75	252
7-9	4	8	32	2,5	6,25	25,00	256
9-11	1	10	10	4,5	20,25	20,25	100
	20		110			95,00	700

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{110}{20} = 5,5$$

$$\sigma = \sqrt{\frac{\sum f_i \left(X_i - \overline{X}\right)^2}{\sum f_i}} = \sqrt{\frac{95}{20}}$$

$$= \sqrt{4,75} = 2,17$$

$$\sigma = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \overline{X}^2}$$

$$= \sqrt{\frac{700}{20} - (5{,}5)^2} = \sqrt{35 - 30{,}25}$$

$$= \sqrt{4{,}75} = 2{,}17$$

V(X) = $\sigma^2 = 4{,}75$

Görüldüğü gibi her iki formül ile aynı sonuç elde edilmiştir.

Standart Sapmanın Özellikleri

- Bir serinin bütün değerlerine sayı eklenir veya çıkarılırsa oluşturulacak yeni serinin standart sapması değişmez. X_i serisinin bütün değerlerinden a sabit sayısı eklenerek veya çıkarılarak yeni bir Y_i serisi oluşturuluyorsa,

$Y_i = X_i + a$ veya $Y_i = X_i - a$ ise $\sigma_Y = \sigma_X$ olacaktır.

$Y_i = X_i + a$ olduğunu düşünelim. Bu durumda $\overline{Y} = \overline{X} + a$ olduğu aritmetik ortalamanın özelliklerinden dolayı bilinmektedir.

Bu durumda Y serisinin standart sapması

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}{n}}$$

olacaktır.

Y_i ve $\overline{Y}$ yerine yukarıda verilen ifadeler konulup, gerekli kısaltmalar yapılırsa

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left[\left(X_i + a\right) - \left(\overline{X} - a\right)\right]^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i + a - \overline{X} - a\right)^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}}$$

eşitliği elde edilir.

X serisi frekans serisi veya sınıflı seri ise,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(Y_i - \overline{Y}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{k} f_i \left[\left(X_i + a\right) - \left(\overline{X} - a\right)\right]^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i + a - \overline{X} - a\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} f_i \left(X_i - \overline{X}\right)^2}{\sum_{i=1}^{k} f_i}} = \sigma_X$$

olacaktır.

$Y_i = X_i - a$ dönüşümü içinde, aynı işlemler yapılırsa $\sigma_Y = \sigma_X$ olduğu görülecektir.

ÖRNEK: Aşağıda verilen X_i serisinin bütün değerlerine 5 ekleyerek yeni bir Y_i serisi oluşturunuz ve $\sigma_Y = \sigma_X$ olduğunu gösteriniz.

X_i	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	$Y_i = X_i + 5$	$Y_i - \overline{Y}$	$\left(Y_i - \overline{Y}\right)^2$
4	-4	16	9	-4	16
5	-3	9	10	-3	9
8	0	0	13	0	0
11	3	9	16	3	9
12	4	16	17	4	16
40			65		50

$$\bar{X} = \frac{\sum X_i}{n} = \frac{40}{5} = 8$$

$$\sigma_X = \sqrt{\frac{\sum \left(X_i - \bar{X}\right)^2}{n}} = \sqrt{\frac{50}{5}}$$

$$= \sqrt{10} = 3{,}16$$

$$\bar{Y} = \bar{X} + 5 = 8 + 5 = 13$$

$$\sigma_Y = \sqrt{\frac{\sum \left(Y_i - \bar{Y}\right)^2}{n}} = \sqrt{\frac{50}{5}}$$

$$= \sqrt{10} = 3{,}16$$

$\sigma_Y = \sigma_X$ = 3,16 olduğu görülmektedir.

- Bir serinin bütün değerlerinin aynı sayı ile çarpılması veya bölünmesi ile oluşturulacak yeni serinin standart sapması, ilk serinin standart sapmasının aynı sayı ile çarpımına veya bölümüne eşittir.

X serisinin bütün birimlerinin değerleri a sayısı ile çarpılarak veya bölünerek Y serisi oluşturulduğunda,

$Y_i = a.X_i$ ise $\sigma_Y = a.\sigma_X$

$Y_i = \frac{X_i}{a}$ ise $\sigma_Y = \frac{\sigma_X}{a}$

olacaktır.

$Y_i = a.X_i$ olduğunu varsayalım. Diğer taraftan $\bar{Y} = a\bar{X}$ olduğu aritmetik ortalamanın özelliklerinden dolayı bilinmektedir. Bu durumda Y serisinin standart sapması,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}{n}}$$

olacaktır.

Yi ve $\bar{Y}$ yerine yukarıda verilen ifadeler konulup, gerekli kısaltmalar yapıldığında

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n}\left(aX_i - a\overline{X}\right)^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n}a^2\left(X_i - \overline{X}\right)^2}{n}}$$

$$\sigma_Y = \sqrt{\frac{a^2\sum\left(X_i - \overline{X}\right)^2}{n}}$$

$$\sigma_Y = a.\sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{n}} = a.\sigma_X$$

elde edilir. X serisi sınıflı seri veya frekans serisi ise,

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{k} f_i\left(Y_i - \overline{Y}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{k} f_i\left(aX_i - a\overline{X}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Y = \sqrt{\frac{a^2\sum f_i\left(X_i - \overline{X}\right)^2}{\sum f_i}}$$

$$\sigma_Y = a.\sqrt{\frac{\sum f_i\left(X_i - \overline{X}\right)^2}{\sum f_i}}$$

olacaktır.

$Y_i = X_i / a$ dönüşümü içinde, aynı işlemler yapılırsa $\sigma_Y = \sigma_X$ /a olduğu görülecektir.

ÖRNEK: Aşağıda verilen X_i serisinin bütün değerlerini 2 ile çarparak bir Y serisi oluşturunuz ve $\sigma_Y = 2\sigma_X$ olduğunu gösteriniz.

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$Y_i = 2X_i$	$Y_i - \overline{Y}$	$(Y_i - \overline{Y})^2$
4	-6	36	8	-12	144
7	-3	9	14	-6	36
10	0	0	20	0	0
11	1	1	22	2	4
18	8	64	36	16	256
50		110	100		440

$$\overline{X} = \frac{\sum X_i}{n} = \frac{50}{5} = 10$$

$$\sigma_X = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{110}{5}}$$

$$= \sqrt{22} = 4{,}69$$

$$\overline{Y} = 2\overline{X} = 2(10) = 20$$

$$\sigma_Y = \sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n}} = \sqrt{\frac{440}{5}}$$

$$= \sqrt{88} = 9{,}38$$

$\sigma_Y = 2\sigma_X = 2(4{,}69) = 9{,}38$

İki farklı hesaplama ile de aynı sonuçların bulunduğu görülmektedir.

-Bir serinin bütün değerlerinden, serinin aritmetik ortalamasının çıkartılması ile elde edilecek farkların, serinin standart sapmasına bölünmesi ile oluşturulacak seriye standart seri denilir. Bu seriyi oluşturan değerlere ise standart değer adı verilir. Standart serinin aritmetik ortalaması 0 ve standart sapması 1'e eşittir.

$$Z_i = \frac{X_i - \overline{X}}{\sigma} \quad \text{ise} \quad \overline{Z} = 0 \quad \text{ve} \quad \sigma_Z = 1 \text{ 'dir.}$$

X_i serisi basit seri ise, Z_i serisinin ortalaması,

$$\overline{Z} = \frac{\sum Z_i}{n}$$

olacaktır. Z_i değeri formülde yerine konulduğunda

$$\overline{Z} = \frac{\sum_{i=1}^{n} \frac{X_i - \overline{X}}{\sigma}}{n} = \frac{\frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \overline{X})}{n}$$

olacak ve

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$

olduğundan

$$\overline{Z} = 0$$

olarak elde edilecektir.

Z serisinin standart sapması ise,

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)^2}{n}}$$

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma} - 0\right)^2}{n}}$$

$$\sigma_Z = \sqrt{\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$$

$$\sigma_Z = \sqrt{\frac{1}{\sigma_2}\sigma^2} = 1$$

olacaktır.

X serisi frekans serisi veya sınıflı seri ise,

$$\bar{Z} = \frac{\sum f_i Z_i}{\sum f_i}$$

$$\bar{Z} = \frac{\sum_{i=1}^{k} f_i \left(\frac{X_i - \bar{X}}{\sigma}\right)}{\sum_{i=1}^{k} f_i} = \frac{\frac{1}{\sigma}\sum_{i=1}^{k} f_i (X_i - \bar{X})}{\sum_{i=1}^{k} f_i}$$

olacak ve $\sum f_i (X_i - \bar{X}) = 0$ olduğundan

$$\bar{Z} = 0$$

olarak elde edilecektir.

Z serisinin standart sapması ise

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(Z_i - \bar{Z}\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(\frac{X_i - \bar{X}}{\sigma} - 0\right)^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Z = \sqrt{\frac{\frac{1}{\sigma^2}\sum_{i=1}^{k} f_i (X_i - \bar{X})^2}{\sum_{i=1}^{k} f_i}}$$

$$\sigma_Z = \sqrt{\frac{1}{\sigma_2}\sigma^2} = 1$$

olarak bulunacaktır.

ÖRNEK: Aşağıda verilen basit seriyi standart seriye dönüştürerek, standart serinin aritmetik ortalamasını ve standart sapmasını hesaplayınız.

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$Z_i - \overline{Z}$	$(Z_i - \overline{Z})^2$
8	-6	36	-1,5	2,25
12	-2	4	-0,5	0,25
14	0	0	0	0
16	2	4	0,5	0,25
20	6	36	1,5	2,25
		80		5,00

$$\overline{X} = \frac{\sum X_i}{n} = \frac{70}{5} = 14$$

$$\sigma_X = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{80}{5}}$$

$$= \sqrt{16} = 4$$

$$Z_i = \frac{X_i - \overline{X}}{\sigma}$$

$$Z_1 = \frac{8-14}{4} = -1{,}5$$

$$Z_2 = \frac{12-14}{4} = -0{,}5$$

$$Z_3 = \frac{14-14}{4} = 0$$

$$Z_4 = \frac{16-14}{4} = 0{,}5$$

$$Z_5 = \frac{20-14}{4} = 1{,}5$$

$$\overline{Z} = \frac{\sum Z_i}{n} = \frac{0}{5} = 0$$

$$\sigma_Z = \sqrt{\frac{\sum_{i=1}^{n} (Z_i - \overline{Z})^2}{n}} = \sqrt{\frac{5}{5}} = 1$$

ÖRNEK: Aşağıda verilen frekans serisini standart seriye dönüştürerek, standart serinin aritmetik ortalamasını ve standart sapmasını hesaplayınız.

X_i	f_i	f_iX_i	$X_i - X$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$	Z_i	f_iZ_i	$(Z_i - \overline{Z})^2$	$f_i(Z_i - \overline{Z})^2$
4	14	56	-2,4	5,76	80,64	-1,2	-16,8	1,44	20,16
6	19	114	-0,4	0,16	3,04	-0,2	-3,8	0,04	0,76
8	10	80	1,6	2,56	25,60	0,8	8,0	0,64	6,40
10	7	70	3,6	12,96	90,72	1,8	12,6	3,24	22,68
	50	320			200,00				50,00

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{320}{50} = 6{,}4$$

$$\sigma_X = \sqrt{\frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i}} = \sqrt{\frac{200}{50}}$$

$$= \sqrt{4} = 2$$

$$Z_i = \frac{X_i - \overline{X}}{\sigma}$$

$$Z_1 = \frac{4 - 6{,}4}{2} = -1{,}2$$

$$Z_2 = \frac{6 - 6{,}4}{2} = -0{,}2$$

$$Z_3 = \frac{8 - 6{,}4}{2} = 0{,}8$$

$$Z_4 = \frac{10 - 6{,}4}{2} = 1{,}8$$

$$\overline{Z} = \frac{\sum f_i Z_i}{\sum f_i} = \frac{0}{50} = 0$$

$$\sigma_Z = \sqrt{\frac{\sum f_i \left(Z_i - \overline{Z}\right)^2}{\sum f_i}} = \sqrt{\frac{50}{50}} = \sqrt{1} = 1$$

-Birden fazla serinin değerlerinin toplamları veya farkları yeni bir seri oluşturduğunda, seriler bağımsız ise, yeni serinin varyansı kendini oluşturan serilerin varyansları toplamına eşittir. Serilerin bağımsız olmaması durumunda kovaryans adi verilen ölçülerinde dikkate alınması gerekecektir.

Kovaryans iki değişkenin birlikte değişiminin ölçüsüdür ve X ve Y gibi iki değişken için $\sigma_{(X,Y)}$ veya Kov (X, Y) şeklinde ifade edilir. X ve Y değişkenleri için kovaryans,

$$Kov(X,Y) = \sigma_{X,Y} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n}$$

formülü ile hesaplanır. $\sigma_{X,Y} = 0$ ise değişkenler bağımsızdır, yani aralarında ilişki yoktur. $\sigma_{X,Y} < 0$, yani negatif işaretli ise değişkenler arasında ters yönlü ilişki vardır, yani değişkenlerden birinin değeri artarken diğerininki azalmaktadır. $\sigma_{X,Y} > 0$, yani pozitif işaretli ise değişkenler arasında doğru yönlü ilişki vardır, yani değişkenlerden birinin değeri artarken diğerinin değeri de artmaktadır.

$Z_i = X_i + Y_i$ ise $\overline{Z} = \overline{X} + \overline{Y}$

olduğunu biliyoruz. Z_i serisinin varyansı,

$$\sigma_Z^2 = \frac{\sum \left(Z_i - \overline{Z}\right)^2}{n}$$

olduğundan yukarıda verilen ifadeler bu formülde yerine konulduğunda,

$$\sigma_Z^2 = \frac{\sum \left[\left(X_i + Y_i\right) - \left(\overline{X} + \overline{Y}\right)\right]^2}{n}$$

$$= \frac{\sum \left(X_i + Y_i - \overline{X} - \overline{Y}\right)^2}{n}$$

$$= \frac{\sum \left[\left(X_i - \overline{X}\right) + \left(Y_i - \overline{Y}\right)\right]^2}{n}$$

$$= \frac{\sum \left[\left(X_i - \overline{X}\right)^2 + \left(Y_i - \overline{Y}\right)^2 + 2\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)\right]}{n}$$

$$= \frac{\sum (X_i - \bar{X})^2}{n} + \frac{\sum (Y_i - \bar{Y})^2}{n} + \frac{2(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

olarak bulunur. Burada $\sigma_{XY} = 0$ ise,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$$

olarak elde edilmektedir.

ÖRNEK: Aşağıda verilen X_i ve Y_i serilerinin toplamı ile, $Z_i = X_i + Y_i$ serisini oluşturarak; X,Y ve Z değişkenlerinin varyansları arasındaki ilişkiyi gösteriniz.

X_i	Y_i	X_i^2	Y_i^2	$Z_i = X_i + Y_i$	Z_i^2	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$
4	2	16	4	6	36	-4	-3	12
7	3	49	9	10	100	-1	-2	2
8	5	64	25	13	169	0	0	0
13	10	169	100	23	529	5	5	25
32	20	298	138	52	834			39

$$\bar{X} = \frac{\sum X_i}{n} = \frac{32}{4} = 8$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{20}{4} = 5$$

$$\sigma_Y^2 = \frac{\sum Y_i^2}{n} - (\bar{Y})^2 = \frac{138}{4} - (5)^2 = 9{,}5$$

$$\sigma_X^2 = \frac{\sum X_i^2}{n} - (\bar{X})^2 = \frac{298}{4} - (8)^2 = 10{,}5$$

$$\bar{Z} = \frac{\sum Z_i}{n} = \frac{52}{4} = 13$$

$$\bar{Z} = \bar{X} + \bar{Y} = 8 + 5 = 13$$

$$\sigma_Z^2 = \frac{\sum Z_i^2}{n} - (\bar{Z})^2 = \frac{834}{4} - (13)^2 = 39{,}5$$

$$\sigma_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{39}{4} = 9{,}75$$

$9{,}75 > 0$ olduğundan,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

$$= 9{,}5 + 10{,}5 + 2(9{,}75) = 39{,}5$$

olarak bulunur.

$Z_i = X_i - Y_i$ ise $\bar{Z} = \bar{X} - \bar{Y}$

olduğundan ve Z_i serisinin varyansı bilindiğinden

$$\sigma_Z^2 = \frac{\sum (Z_i - \bar{Z})^2}{n}$$

olduğundan yukarıda verilen ifadeleri bu formülde yerine koyarsak,

$$\sigma_Z^2 = \frac{\sum \left[(X_i - Y_i) - (\bar{X} - \bar{Y}) \right]^2}{n}$$

$$= \frac{\sum (X_i - Y_i - \bar{X} + \bar{Y})^2}{n}$$

$$= \frac{\sum \left[(X_i - \bar{X}) - (Y_i - \bar{Y}) \right]^2}{n}$$

$$= \frac{\sum \left[(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2 - 2(X_i - \bar{X})(Y_i - \bar{Y}) \right]}{n}$$

$$= \frac{\sum (X_i - \bar{X})^2}{n} + \frac{\sum (Y_i - \bar{Y})^2}{n} - \frac{2(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$$

olarak bulunur. Burada $\sigma_{XY} = 0$ ise,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$$

olacaktır.

ÖRNEK: Bir önceki örnekte verilen X_i ve Y_i serilerinden, $Z_i = X_i - Y_i$ serisini oluşturarak; X, Y ve Z değişkenlerinin varyansları arasındaki ilişkiyi gösteriniz.

X_i	Y_i	$Z_i = X_i - Y_i$	Z_i^2
4	2	2	4
7	3	4	16
8	5	3	9
13	10	3	9
32	20	12	38

$$\overline{X} = 8\,;\ \overline{Y} = 5\,;\ \sigma_X^2 = 10{,}5\,;\ \sigma_Y^2 = 9{,}5 \text{ ve } \sigma_{XY} = 9{,}75$$

olarak daha önce bulunmuştu.

$$\overline{Z} = \frac{\sum Z_i}{n} = \frac{12}{4} = 3$$

$$\overline{Z} = \overline{X} - \overline{Y} = 8 - 5 = 3$$

$$\sigma_Z^2 = \frac{\sum Z_i^2}{n} - (\overline{Z})^2 = \frac{38}{4} - (3)^2 = 0{,}5$$

$\sigma_{XY} > 0$ olduğundan,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$$

$$= 9{,}5 + 10{,}5 - 2(9{,}75) = 0{,}5$$

olarak bulunur.

İkiden fazla serinin değerlerinin toplamları veya farklarından oluşturulacak yeni serinin varyansı için gerekli formüller benzer şekilde elde edilebilir. Örneğin,

$Z_i = X_i + Y_i - T_i$ ise,

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_T^2 + 2\sigma_{XY} + 2\sigma_{XT} - 2\sigma_{YT}$$

olacaktır.

-Bir seri farklı serilerin birleştirilmesi ile oluşturuluyorsa, yeni serinin varyansı kendini oluşturan serilerin aritmetik ortalamalarının oluşturduğu serinin varyansları toplamına eşittir.

m birimli X_L serisi ile r birimli Y_j serisinin birleştirilerek n = m+r birimli Z_i serisi oluşturduğunu varsayalım. Bu seriler,

$X_L : X_1, X_2, ..., X_m$

$Y_j : Y_1, Y_2, ..., Y_r$

$Z_i : X_1, X_2, ..., X_m \ , \ Y_1, Y_2, ..., Y_r$

olacaktır. Bu durumda,

$$\sigma_Z^2 = \frac{m\sigma_X^2 + r\sigma_Y^2}{m+r} = \frac{m(\overline{X}-\overline{Z})^2 + r(\overline{Y}-\overline{Z})^2}{m+r}$$

$$\sigma_Z^2 = \frac{m\sigma_X^2 + r\sigma_Y^2}{n} = \frac{m(\overline{X}-\overline{Z})^2 + r(\overline{Y}-\overline{Z})^2}{n}$$

olarak elde edilecektir. Diğer taraftan,

$$\sigma_Z^2 = \frac{\sum (Z_i - \overline{Z})^2}{n}$$

olduğundan ve Z_i serisi X_L ve Y_j serilerinin birleştirilmesi ile oluşturulduğundan,

$$\sigma_Z^2 = \frac{\sum_{L=1}^{m} (X_i - \overline{Z})^2 + \sum_{j=1}^{r} (Y_i - \overline{Z})^2}{n}$$

olacaktır. $\sum_{L=1}^{m} (X_i - \overline{Z})^2$ ifadesine $\pm \overline{X}$ ve $\sum_{j=1}^{r} (Y_i - \overline{Z})^2$ ifadesine $\pm \overline{Y}$ ekleyelim.

$$\sigma_Z^2 = \frac{\sum_{L=1}^{m} (X_L - \overline{X} + \overline{X} - \overline{Z})^2 + \sum_{j=1}^{r} (Y_j - \overline{Y} + \overline{Y} - \overline{Z})^2}{n}$$

$$\sigma_Z^2 = \frac{\sum_{L=1}^{m} [(X_L - \overline{X}) + (\overline{X} - \overline{Z})]^2 + \sum_{j=1}^{r} [(Y_j - \overline{Y}) + (\overline{Y} - \overline{Z})]^2}{n}$$

$$\sigma_Z^2 = \frac{\sum_{L=1}^{m} [(X_L - \overline{X})^2 + (\overline{X} - \overline{Z})^2 + 2(X_L - \overline{X})(\overline{X} - \overline{Z})] + \sum_{j=1}^{r} [(Y_j - \overline{Y})^2 + (\overline{Y} - \overline{Z})^2 + 2(Y_j - \overline{Y})(\overline{Y} - \overline{Z})]}{n}$$

$$\sigma_Z^2 = \frac{\sum_{L=1}^{2}\left(X_L - \bar{X}\right)^2 + m\left(\bar{X} - Z\right)^2 + 2\left(\bar{X} - \bar{Z}\right)\sum_{L=1}^{m}\left(X_L - \bar{X}\right)}{n}$$

$$+ \frac{\sum_{j=1}^{r}\left(Y_j - \bar{Y}\right)^2 + r\left(\bar{Y} - Z\right)^2 + 2\left(\bar{Y} - \bar{Z}\right)\sum_{j=1}^{r}\left(Y_j - \bar{Y}\right)}{n}$$

$\sum_{L=1}^{m}\left(X_L - \bar{X}\right)^2 = 0$ ve $\sum_{j=1}^{r}\left(Y_j - \bar{Y}\right) = 0$ olduğundan,

$$\sigma_Z^2 = \frac{\sum_{L=1}^{m}\left(X_L - \bar{X}\right)^2 + \sum_{j=1}^{r}\left(Y_j - \bar{Y}\right)^2 + m\left(\bar{X} - \bar{Z}\right)^2 + r\left(\bar{Y} - \bar{Z}\right)^2}{n}$$ olur.

$\sigma_X^2 = \frac{\sum_{L=1}^{m}\left(X_L - \bar{X}\right)^2}{m}$ olduğundan

$$\sum_{L=1}^{m}\left(X_L - \bar{X}\right)^2 = m\sigma_X^2$$

$\sigma_Y^2 = \frac{\sum_{j=1}^{r}\left(Y_j - \bar{Y}\right)^2}{r}$ olduğundan

$$\sum_{j=1}^{r}\left(Y_j - \bar{Y}\right)^2 = r\sigma_Y^2$$

olur. Bu eşitlikler ile,

$$\sigma_Z^2 = \frac{m\sigma_X^2 + r\sigma_Y^2}{n} = \frac{m(\bar{X} - \bar{Z})^2 + r(\bar{Y} - \bar{Z})^2}{n}$$

olarak elde edilir.

ÖRNEK: Aşağıda verilen X_L ve Y_j serilerini birleştirerek bir Z_i serisi oluşturunuz ve varyansları arasındaki ilişkiyi gösteriniz.

X_L	Y_j
6	1
9	4
10	5

11	8
36	9
	27

Zi : 6; 9; 10; 11; 1; 4; 5; 8; 9

$$\bar{Z} = \frac{\sum Z_i}{n} = \frac{6+9+10+11+1+4+5+8+9}{9} = \frac{63}{9} = 7$$

$$\bar{X} = \frac{\sum X_L}{m} = \frac{36}{4} = 9$$

$$\bar{Y} = \frac{\sum Y_j}{r} = \frac{27}{5} = 5{,}4$$

$$\sigma_Z^2 = \frac{\sum \left(Z_i - \bar{Z}\right)^2}{n}$$

$$\sigma_Z^2 = \frac{(6-7)^2 + (9-7)^2 + (10-7)^2 + (11-7)^2 + (1-7)^2 + (4-7)^2 + (5-7)^2 + (8-7)^2 + (9-7)^2}{9}$$

$$= \frac{84}{9} = 9{,}3333$$

$$\sigma_X^2 = \frac{\sum_{L=1}^{m} \left(X_L - \bar{X}\right)^2}{m}$$

$$= \frac{(6-9)^2 + (9-9)^2 + (10-9)^2 + (11-9)^2}{4}$$

$$= \frac{14}{4} = 3{,}5$$

$$\sigma_Y^2 = \frac{\sum_{j=1}^{r} \left(Y_j - \bar{Y}\right)^2}{r}$$

$$= \frac{(1-5{,}4)^2 + (4-5{,}4)^2 + (5-5{,}4)^2 + (8-5{,}4)^2 + (9-5{,}4)^2}{5}$$

$$= \frac{41{,}2}{5} = 8{,}24$$

$$\sigma_Z^2 = \frac{m\sigma_X^2 + r\sigma_Y^2}{n} = \frac{m(\overline{X} - \overline{Z})^2 + r(\overline{Y} - \overline{Z})^2}{n}$$

$$\sigma_Z^2 = \frac{4(3,5) + 5(8,24)}{9} + \frac{4(9-7)^2 + 5(5,4-7)^2}{9}$$

$$= \frac{14 + 41,2}{9} + \frac{16 + 12,8}{9}$$

$$= \frac{84}{9} = 9,3333$$

- Bir ana kütlenin varyansı σ^2 örnek alınarak tahmin ediliyorsa, örnek varyansı anakütle varyansının sapmasız tahmincisi olmadığından, tahminci olarak S^2 kullanılır.

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

olarak hesaplanır ve anakütle varyansının sapmasız tahmincisidir. Bu nedenle,

$$\sigma^2 = S^2$$

olarak tahmin edilir.

Örnek varyansı ise,

$$\sigma_X^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

olduğundan

$$S^2 = \frac{n\sigma_X^2}{n-1}$$

olacaktır.

4.1.3. Değişim Katsayısı

Serilerin dağılımlarının değişkenlik ölçüleri yardımı ile karşılaştırılabileceklerinden daha önce söz edilmişti. Ancak bu ölçüler mutlak değerler ile hesaplandıklarından, karşılaştırılmak istenen serilerin değerlerinin farklı ölçü birimleri ile ifade edildiği veya seri değerlerinin büyüklüklerinin farklı olduğu durumlarda kullanılmaları uygun

değildir. Örneğin, ölçümler, litre, metre, kilogram vb. cinsinden olabilir. Bu gibi durumlarda serilerin dağılımlarının karşılaştırılmasında değişim katsayısı kullanılabilir.

Değişim katsayısı oransal bir ölçüdür ve

$$DK = \frac{\sigma}{\overline{X}}.100$$

formülü ile hesaplanmaktadır. Bu hesaplama ile mutlak değerler, oranlara dönüştürüldüğünden, ölçü birimlerinin etkisi giderilmiş olmaktadır. Serilerin değişim katsayıları hesaplandığında, değişim katsayısı küçük olan serilerde, birimlerin ortalama etrafında daha uygun dağıldıkları sonuca varılmaktadır.

ÖRNEK: Aritmetik ortalaması 8 ve varyansı 9 olan serinin değişim katsayısını hesaplayınız.

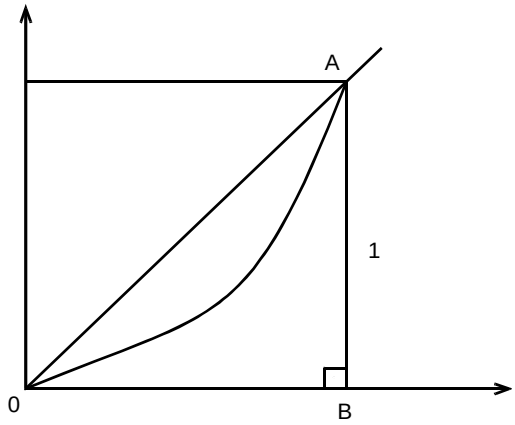
$$\sigma = \sqrt{9} = 3 \qquad DK = \frac{\sigma}{\overline{X}}.100 = \frac{3}{8}.100 = \%37{,}5$$

4.1.4. Toplanma Oranı

Toplam değerini bütün birimlere eşit olarak dağıldığı dağılımlarda eşit dağılım söz konusudur. Dağılımı eşit olan serilerde, serinin bütün birimlerinin değerinin aritmetik ortalamaya eşit olması gerekmektedir. Serinin toplam değeri $\sum f_i X_i$ ve toplam birim sayısı $\sum f_i$ ise toplam değerin toplam birim sayısına bölümü aritmetik ortalamaya eşit olacaktır. Seri birimlerinin değerleri hesaplanan aritmetik ortalamadan farklılaştıkça dağılımın eşitliği bozulacak ve belirli birimlerde toplanma başlayacaktır. Serilerdeki bu toplanmanın ölçüsü toplanma oranı ile hesaplanmaktadır.

Toplanma oranının hesaplanabilmesi için artan birikimli oransal frekanslardan yararlanılır. Bilindiği gibi artan birikimli oransal frekanslar bir seride belirli bir değerden daha küçük değer alan birimlerin oranını vermektedir. Serinin toplam değeri ve frekanslar için artan birikimli oransal frekanslar hesaplanırsa, bu frekanslar yardımı ile dağılımın eşit olup olmadığı belirlenebilir. Artan birikimli oransal frekanslar p_i, artan birikimli oransal toplam değeri q_i ile ifade edersek, seride eşit dağılım söz konusu ise p_i ve q_i değerleri eşit olacaktır. Bu p_i ve q_i değerleri grafik üzerinde gösterilerek, seride toplanma olup olmadığı belirlenebilir. Artan birikimli oransal frekansları gösteren p_i değerleri yatay eksende, artan birikimli oransal toplam değeri gösteren q_i değeri düşey eksende gösterilerek çizilen grafikte, dağılım eşit olduğunda bir tek noktanın elde edildiği görülecek ve bu noktanın orijin ile birleştirilmesiyle eksenlere 45^0'lik açı yapan bir doğru elde edilecektir. dağılımın eşit olmadığı durumlarda bu doğru bir eğri şeklini alacaktır. Bu eğriye "Lorenz eğrisi" veya "Toplanma eğrisi" adı verilmektedir.

Şekil 4.1.'de verilen grafikte eksenlere 45^0'lik açı yapan doğru ile Lorenz eğrisi görülmektedir.



Şekil 4.1

Şekil 4.1'de görüldüğü gibi, orijine 45^0'lik açı yapan doğru veya köşegen arasında kalan alana toplanma alanı, toplanma alanının köşegen ile yatay eksen arasında kalan üçgenin alanına oranına ise, toplanma oranı adını alır. Eşit dağılım durumunda Lorenz eğrisi köşegen ile çakıştığından toplanma oranı 0, tam toplanma durumunda ise toplanma alanı 0AB Üçgeninin alanına eşit olacağından, toplanma oranıda (1)'e eşit olacaktır. Bu nedenle toplanma oranı (0) ile (1) arasında değişecektir.

Toplanma oranı grafik ile hesaplanabileceği gibi aşağıdaki formül yardımı ile de hesaplanabilir.

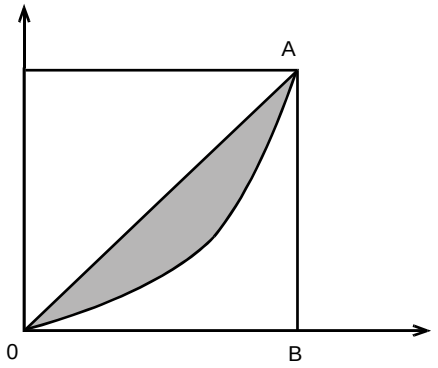
$$TO = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

Toplanma oranının hesaplanmasını bir örnek ile gösterelim.

ÖRNEK: Aşağıdaki serinin toplanma oranını hesaplayarak grafik üzerinde gösteriniz.

Sınıflar	f_i	X_i	f_iX_i	Birikimli Frekans	Birikimli f_iX_i	p_i	qi	piqi
1-3	78	2	156	78	156	0,171	0,070	0,101
3-5	175	4	700	253	856	0,556	0,388	0,168
5-7	142	6	852	395	1708	0,868	0,774	0,094
7-9	51	8	408	446	2116	0,980	0,959	0,021
9-11	9	10	90	455	2206	1,000	1,000	0,000
	455		2206			2,575		0,384

$$TO = \frac{\sum_{i=1}^{4}(p_i - q_i)}{\sum_{i=1}^{4} p_i} = \frac{0{,}384}{2{,}575} = 0{,}419$$



Şekil 4.2

Toplama oranı yukarıda görülen formül ile hesaplanabildiği gibi Şekil 4.2'den de hesaplanabilir. Bu durumda toplanma oranı,

$$TO = \frac{Taralı\ alan}{\overset{\Delta}{OAB}\ alanı}$$

ifadesi ile hesaplanır.

Toplanma oranı milli gelir, servet, sermaye, ücret ve toprak dağılımlarının dengeli olup olmadığının belirlenmesinde kullanılabilir. Yukarıda söz edilen dağılımlar için hesaplanan toplanma oranının (0)'a yaklaşması durumunda dağılımın daha dengeli, (1)'e yaklaşması durumunda ise daha dengesiz olduğu sonucuna varılmaktadır.

4.2. ASİMETRİ ÖLÇÜLERİ

Serileri temsil etmek için hesaplanan ortalamaların, serilerin karşılaştırılmasında yeterli olmadıkları durumlarda değişkenlik ölçülerinin hesaplanması gerektiğinden daha önce söz edilmişti.

Değişkenlik ölçüleri serilerin değişkenliğini ölçebilmesine rağmen, onların dağılma şekilleri hakkında bilgi veremezler. Oysa ortalamaları ve değişkenlik ölçüleri birbirine eşit veya yakın olan serilerin dağılımları birbirinden farklı olabilir. Aşağıda iki farklı seri ile bu serilerin aritmetik ortalamaları ve standart sapmaları verilmiştir.

X_i	f_i	Y_j	f_j
1	3	1	6

2	17	2	8
3	10	3	15
5	4	4	5
6	4	6	4
7	2	7	2
	40		40

$\overline{X} = 3{,}125$ $\overline{Y} = 3{,}125$

$\sigma_X = 1{,}67$ $\sigma_Y = 1{,}61$

Örnekteki serilerin aritmetik ortalamaları birbirine eşit, standart sapmaları ise birbirine çok yakın olmasına rağmen serilerin dağılımlarının birbirinden farklı olduğu görülmektedir. Bu nedenle, serilerin sadece değişkenlik ölçüleri değil, dağılma şekillerinindi belirlenmesi gerekmektedir.

Serilerin dağılma şekillerini belirlemek için asimetri ölçüleri adı verilen ölçüler hesaplanmaktadır. Bu ölçüler yardımı ile dağılımın serinin belirli değerlerinde yoğun olup olmadığı belirlenebilir.

Asimetri ölçüleri ile dağılma şekilleri belirlenecek serilerin bir kısmi, belirli bir değere göre simetrik dağılım göstermektedirler. Böyle serilere simetrik seri adı verilir. Dağılımı simetrik olmayıp belirli değerlerde yoğunlaşan serilere ise asimetrik seri adı verilmektedir. Asimetrik seriler dağılımın yoğunlaştığı değerlere göre sağa veya sola asimetrik olabilirler.

Asimetri ölçülerinin hesaplanmasının bir diğer nedeni de serilerin dağılımının, normal dağılıma uygun olup olmadığının belirlenmesidir. Dağılımı çan eğrisine benzeyen, tek maksimumlu ve simetrik olan normal dağılım istatistikte önemli bir yere sahiptir ve bazı yöntemlerin uygulanması dağılımların normal dağılım olduğu varsayımına dayanmaktadır. Bu nedenle dağılımların normal dağılıma uygun olup olmadıklarının belirlenmesi gerekebilir.

Bir dağılımın normal dağılım olarak kabul edilebilmesi için simetri durumunun bulunması ve basıklık ölçüsünün hesaplanması gerekecektir. Basıklık ölçüsü ile bir dağılımın basıklığının, normal dağılım basıklığından farklı olup olmadığı belirlenmektedir. Bir serinin dağılımının simetrik olup olmadığının belirlenmesinde ise en kolay yol serinin grafiğini çizmektir. Ancak bu yolla incelemede bazı hatalar meydana gelebileceğinden, simetri durumunun asimetri ölçülerinden biri ile incelenerek, karara varılması daha doğru olacaktır. Bu bölümde ortalamalar, kartiller ve momentler ile hesaplanan asimetri ölçüleri anlatılacaktır.

4.2.1. Ortalamalar ve Kartiller Arası Farklar ile Hesaplanan Asimetri Ölçüleri

Simetrik dağılım gösteren tek modlu serilerde aritmetik ortalama, mod ve medyan birbirine eşittir. Bu nedenle, herhangi bir serinin aritmetik ortalaması, modu ve medyanı birbirine eşit yani,

$\overline{X} = M_e = M_o$ ise

seri simetriktir diyebiliriz. Bu değerler arasındaki fark arttıkça serinin eğikliği de artacaktır.

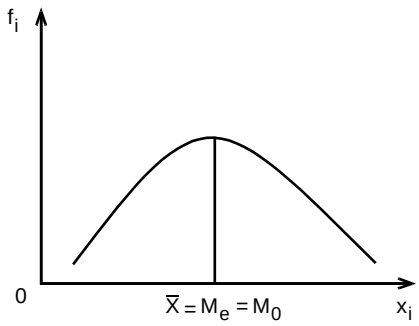
$\overline{X} < M_e < M_o$

olduğunda seri sola eğik,

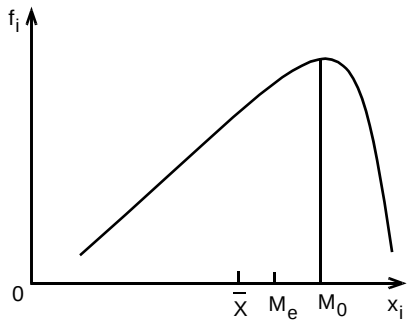
$\overline{X} > M_e > M_o$

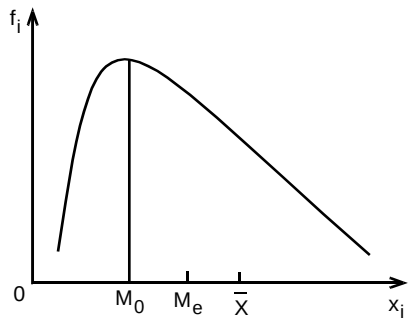
olduğunda ise seri sağa eğik olacaktır.

Şekil 4.3. (a) da simetrik, Şekil 4.3. (b)de sola eğik ve Şekil 4.3. (c)'de sağa eğik seriler görülmektedir.



(a)





(b) **(c)**

Şekil 4.3

Sağa ve sola eğikliği fazla olmayan serilerde, aritmetik ortalama ile mod arasındaki fark, aritmetik ortalama ile medyan arasındaki farkın yaklaşık üç katına eşittir. Bu eşitlik,

$$\overline{X} - M_o \cong 3(\overline{X} - M_e)$$

olarak gösterilebilir ve bu eşitlikten yararlanılarak Pearson asimetri ölçüleri olarak adlandırılan ölçüler hesaplanabilir. Bu ölçülerin hesaplanmasında farklı ölçü birimlerinin etkisini gidermek ve oransal bir ölçü elde edebilmek için, aritmetik ortalama ile mod veya medyan arasındaki farklar standart sapmaya bölünmektedir. α_1 ve α_2 sembolleri ile göstereceğimiz Pearson asimetri ölçülerinin formülleri aşağıda verildiği gibidir.

$$\alpha_1 = \frac{\overline{X} - M_o}{\sigma}$$

$$\alpha_2 = \frac{3(\overline{X} - M_e)}{\sigma}$$

Bu formüllere göre,
Seri simetrik ise ,

$\overline{X} = M_e = M_o$ olacağından $\alpha_1 = 0$ $\alpha_2 = 0$

Seri sola eğik ise,

$\overline{X} < M_e < M_o$ olacağından $\alpha_1 < 0$ $\alpha_2 < 0$

Seri sağa eğik ise,

$\overline{X} > M_e > M_o$ olacağından $\alpha_1 > 0$ $\alpha_2 > 0$

olacaktır.

Bu nedenle α_1 veya α_2 , sıfır ise serinin simetrik olduğu, pozitif ise sağa eğik olduğu, negatif ise sola eğik olduğu belirlenmiş olacaktır. Diğer bir ifade ile asimetri ölçülerinin sıfır olması simetriyi, negatif olması sola ve pozitif olması sağa eğikliği ifade etmektedir.

Asimetri ölçüleri ortalamalar arası farka dayanarak hesaplanabildikleri gibi, kartiller arası farka dayanarak da hesaplanabilirler. Bu durumda simetrik serilerde 3. kartil ile 2.kartil arasındaki fark, 2.kartil ile 1.kartil arasındaki farka eşit olduğundan, serilerin simetrik olup olmadığı veya eğikliğinin yönü belirlenebilir. Bu açıklamalara göre,

$$Q_3 - Q_2 = Q_2 - Q_1$$

ise seri simetrik,

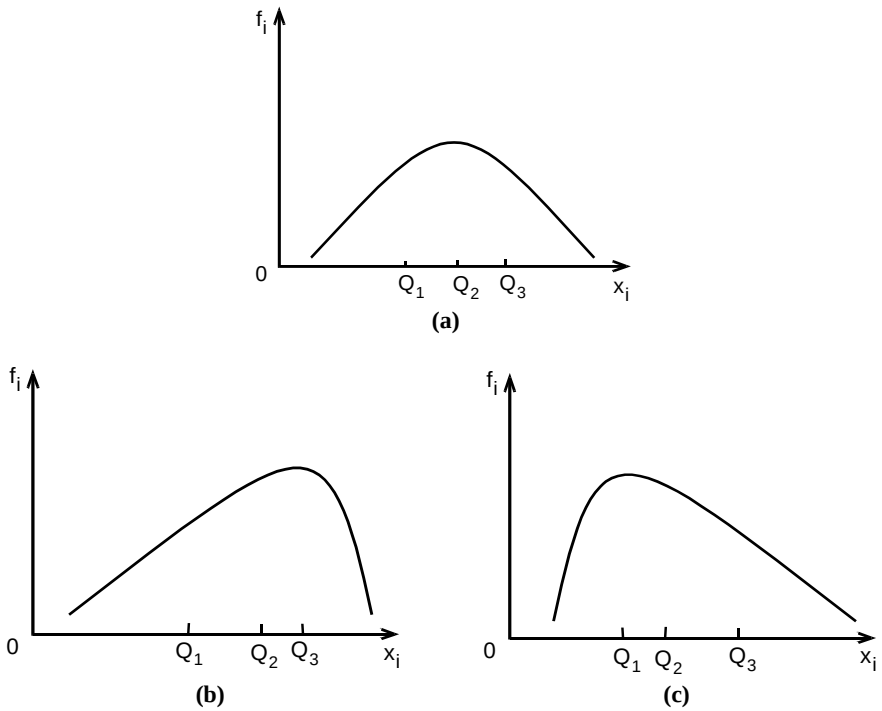
$$Q_3 - Q_2 < Q_2 - Q_1$$

ise seri sola eğik,

$$Q_3 - Q_2 > Q_2 - Q_1$$

ise seri sağa eğik olacaktır.

Kartiller arası farklara göre eğiklik Şekil 4.4'de görülmektedir.



Şekil 4.4

Şekilde görüldüğü gibi

$(Q_3 - Q_2) - (Q_2 - Q_1) = 0$ ise seri simetrik (şekil 4.4.a),

$(Q_3 - Q_2) - (Q_2 - Q_1) < 0$ ise sola eğik (şekil 4.4.b),

$(Q_3 - Q_2) - (Q_2 - Q_1) > 0$ ise sağa eğik (şekil 4.4.c)

olacaktır. Serilerin simetri ve eğiklik durumlarını incelemek için yapılan hesaplamalarda, farklı ölçü birimlerinin etkisini gidermek ve oransal bir ölçü elde edebilmek için, kartiller

arası farklar kartiller arası farklar toplamına bölünür. Bu şekilde elde edilen ölçü Bowley asimetri ölçüsü olarak adlandırılmaktadır. Bowley asimetri ölçüsü,

$$A_S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

olarak gösterilir. Formüldeki parantezler açıldığında ölçü,

$$A_S = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

olarak elde edilir.

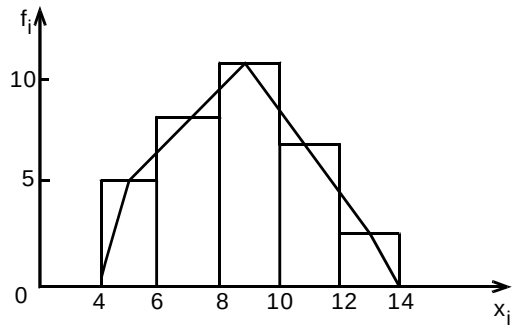
Bowley asimetri ölçüsüne göre elde edilen sonuç (+1) ile (-1) arasında değişmektedir. Sonuç (0) ise seri simetrik, (-) ise seri sola asimetrik ve (+) ise seri sağa asimetrik olacaktır.

ÖRNEK: Aşağıda bir sınıflı seri verilmiştir.

Sınıflar	f_i	Artan Birikimli Frekanslar	X_i	f_iX_i	$f_iX_i^2$
4-6	5	5	5	25	125
6-8	8	13	7	56	392
8-10	12	25	9	108	972
10-12	7	32	11	77	847
12-14	3	35	13	39	507
	35			305	2843

Bu serinin simetrik olup olmadığını,

- Grafiğini çizerek,
- Mod, medyan ve aritmetik ortalamasını karşılaştırarak,
- Pearson katsayılarını ve
- Bowley asimetri ölçüsünü hesaplayarak inceleyiniz.



Şekil 4.5

Şekil 4.5'den dağılımın, simetrik kabul edilebileceği görülmektedir.

$$M_0 = I_a + \frac{\Delta_1}{\Delta_1 + \Delta_2}.c$$

$$M_0 = 8 + \frac{4}{4+5}.2 = 8{,}88$$

$$M_e = I_a + \frac{\frac{n}{2} - \sum_{i=1}^{k-1} f_b}{f_m}.c$$

$$\frac{n}{2} = \frac{35}{2} = 17{,}5$$

$$M_e = 8 + \frac{17{,}5 - 13}{12}.2 = 8{,}75$$

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{305}{35} = 8{,}71$$

$\overline{X} < M_e < M_o$ olduğundan seri sola asimetriktir.

$$\sigma = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - (\overline{X})^2}$$

$$\sigma = \sqrt{\frac{2843}{35} - (8{,}71)^2} = 2{,}31$$

$$\alpha_1 = \frac{\overline{X} - M_o}{\sigma}$$

$$\alpha_1 = \frac{8,71 - 8,88}{2,31} = 0,07$$

$\alpha_1 < 0$ olduğundan seri sola asimetriktir.

$$\alpha_2 = \frac{3(\overline{X} - M_e)}{\sigma}$$

$$\alpha_2 = \frac{3(8,71 - 8,75)}{2,31} = -0,05$$

$\alpha_2 < 0$ olduğundan seri sola asimetriktir.

$$Q_{h/r} = I_a + \frac{n\frac{h}{r} - \sum_{i=1}^{k-1} f_i}{f_b}.c$$

$$\frac{n}{4} = \frac{35}{4} = 8,75 \qquad \frac{3n}{4} = \frac{3(35)}{4} = 26,25$$

$$Q_1 = 6 + \frac{8,75 - 5}{8}.2 = 6,93$$

$$Q_2 = M_e = 8,75$$

$$Q_3 = 10 + \frac{26,25 - 25}{7}.2 = 10,35$$

$$A_S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

$$= \frac{(10,35 - 8,75) - (8,75 - 6,93)}{(10,35 - 8,75) + (8,75 - 6,93)} = \frac{1,6 - 1,82}{1,6 + 1,82} = -0,06$$

Sonuç negatif olduğundan, seri sola asimetriktir. Ancak asimetri katsayısının değeri sıfıra yakın olduğundan seri simetrik kabul edilebilir.

4.2.2. Momentler ile Hesaplanan Asimetri ve Basıklık Ölçüleri

Daha önce belirtildiği gibi bir serinin simetri durumunun incelenmesinde, momentler ile hesaplanan asimetri ölçülerinin kullanılması durumunda daha doğru sonuçlar elde edilmektedir. Ayrıca momentler ile bir dağılımın basıklığının normal dağılıma uygun olup olmadığının belirlenmesinde kullanılan basıklık ve yükseklik ölçüleri de

hesaplanabilmektedir. Momentler ile hesaplanan asimetri ve basıklık ölçülerini hesaplamadan önce, kısaca momentlerden söz edilmesi uygun olacaktır.

4.2.2.1. Momentler

Seri değerlerinin belirli bir orijine göre çeşitli derecelerden kuvvetlerinin ortalamalarına moment adı verilmektedir. Momentler genel olarak basit seriler için,

$$ {}_a m_q = \frac{\sum_{i=1}^{n} (X_i - a)^q}{n} $$

frekans serileri ve sınıflı seriler için,

$$ {}_a m_q = \frac{\sum_{i=1}^{k} f_i (X_i - a)^q}{\sum_{i=1}^{k} f_i} $$

formülleri ile ifade edilebilir. Formüllerde görülen a için farklı değerler kullanılabilirse de, uygulamada genellikle $a = 0$ veya $a = \overline{X}$ alınmaktadır. a = 0 olduğunda hesaplanan momentlere orijine veya sıfıra göre momentler (m), $a = \overline{X}$ olduğunda hesaplanan momentlere ise ortalamaya göre momentler (μ) adı verilir. Formülde görülen q ise momentin derecesini gösterir ve uygulamada genellikle birinci, ikinci, üçüncü ve dördüncü dereceden momentler kullanılmaktadır.

Orjine göre momentler, basit seriler, frekans serileri ve sınıflı serilerde aşağıda verilen formüller ile hesaplanmaktadır.

Basit Seri	Frekans Serisi ve Sınıflı Seri
$m_1 = \frac{\sum X_i}{n} = \overline{X}$	$m_1 = \frac{\sum f_i X_i}{\sum f_i} = \overline{X}$
$m_2 = \frac{\sum X_i^2}{n} = K^2$	$m_2 = \frac{\sum f_i X_i^2}{\sum f_i} = K^2$
$m_3 = \frac{\sum X_i^3}{n}$	$m_3 = \frac{\sum f_i X_i^3}{\sum f_i}$
$m_4 = \frac{\sum X_i^4}{n}$	$m_4 = \frac{\sum f_i X_i^4}{\sum f_i}$

Görüldüğü gibi orjine göre birinci moment aritmetik ortalamaya, orjine göre ikinci moment ise kareli ortalamanın karesine eşittir.

Ortalamaya göre momentler ise basit seriler, frekans serileri ve sınıflı serilerde aşağıda verilen formüller ile hesaplanmaktadır.

Basit	**Frekans Serisi ve Sınıflı Seri**
$\mu_1 = \frac{\sum (X_i - \overline{X})}{n} = 0$	$\mu_1 = \frac{\sum f_i (X_i - \overline{X})}{\sum f_i} = 0$
$\mu_2 = \frac{\sum (X_i - \overline{X})^2}{n} = \sigma^2$	$\mu_2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i} = \sigma^2$
$\mu_3 = \frac{\sum (X_i - \overline{X})^3}{n}$	$\mu_3 = \frac{\sum f_i (X_i - \overline{X})^3}{\sum f_i}$
$\mu_4 = \frac{\sum (X_i - \overline{X})^4}{n}$	$\mu_4 = \frac{\sum f_i (X_i - \overline{X})^4}{\sum f_i}$

Daha önce aritmetik ortalamanın özelliklerinde belirtildiği gibi, seri birimlerinin aritmetik ortalamadan farklarının toplamı sıfır olduğundan ortalamaya göre birinci moment sıfıra eşit olacaktır. Ortalamaya göre ikinci moment ise varyansa eşittir.

Yukarıda verilen ortalamaya göre moment formüllerinin kuvvetleri alınır ve gerekli kısaltmalar yapılırsa,

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

formülleri elde edilir. Serilerin orijine göre momentleri biliniyorsa, yukarıda verilen formüller ile ortalamaya göre momentleri hesaplanabilir.

ÖRNEK: Aşağıda verilen frekans serisinin orijine ve ortalamaya göre momentlerini hesaplayınız.

X_i	f_i	X_if_i	$X_if_i^2$	$X_if_i^3$	$X_if_i^4$	$X_i - \overline{X}$	$f_i(X_i - \overline{X})$	$f_i(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^4$
1	12	12	12	12	12	-2	-24	48	-96	192
2	17	34	68	136	272	-1	-17	17	-17	17
3	20	60	180	540	1620	0	0	0	0	0
4	21	84	336	1344	5376	1	21	21	21	21

5	10	50	250	1250	6250	2	20	40	80	160
	80	240	846	3282	13530		0	126	-12	390

- Orjine göre momentler

$$m_1 = \frac{\sum f_i X_i}{\sum f_i} = \frac{240}{80} = 3 = \overline{X}$$

$$m_2 = \frac{\sum f_i X_i^2}{\sum f_i} = \frac{846}{80} = 10{,}575 = K^2$$

$$m_3 = \frac{\sum f_i X_i^3}{\sum f_i} = \frac{3282}{80} = 41{,}025$$

$$m_4 = \frac{\sum f_i X_i^4}{\sum f_i} = \frac{13530}{80} = 169{,}125$$

- Ortalamaya göre momentler

$$\mu_1 = \frac{\sum f_i (X_i - \overline{X})}{\sum f_i} = \frac{0}{80} = 0$$

$$\mu_2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i} = \frac{126}{80} = 1{,}575 = \sigma^2$$

$$\mu_3 = \frac{\sum f_i (X_i - \overline{X})^3}{\sum f_i} = \frac{-12}{80} = -0{,}15$$

$$\mu_4 = \frac{\sum f_i (X_i - \overline{X})^4}{\sum f_i} = \frac{390}{80} = 4{,}875$$

-Orjine göre momentler ile ortalamaya göre momentlerin hesaplanması

$$\mu_2 = m_2 - m_1^2$$
$$= 10{,}575 - (3)^2$$
$$= 1{,}575$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$
$$= 41{,}025 - 3\,(3)(10{,}575) + 2(3)^3$$
$$= 41{,}025 - 95{,}175 + 54$$

= - 0,15

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$
$$= 169{,}125\text{-}4(3)(41{,}025)+ 6(3)^2(10{,}575) - 3(3)^4$$
$$= 169{,}125 - 492{,}3 + 571{,}05 - 243$$
$$= 4{,}875$$

4.2.2.2. Momentler ile Hesaplanan Asimetri Ölçüsü

Momentler ile asimetri ölçüsü hesaplanırken, ortalamaya göre 3. moment, ortalamaya göre 2. momentin 3/2. kuvvetine bölünür ve elde edilen asimetri ölçüsü α_3 ile ifade edilir.

α_3 katsayısı aşağıda verilen formül ile hesaplanır.

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}}$$

$\mu_2 = \sigma^2$ olduğundan

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

olacaktır.

Momentler ile hesaplanan α_3 katsayısı,

$\alpha_3 = 0$ ise seri simetrik,

$\alpha_3 < 0$ ise seri sola asimetrik,

$\alpha_3 > 0$ ise seri sağa asimetrik

olacaktır.

ÖRNEK: Aşağıda verilen serinin simetrik olup olmadığını inceleyiniz.

Sınıflar	f_i	X_i	f_iX_i	$X_i - \overline{X}$	$f_i(X_i - \overline{X})$	$f_i(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^3$
0-4	15	2	30	-7	-105	735	-5145
4-8	44	6	264	-3	-132	396	-1188
8-12	30	10	300	1	30	30	30
12-16	18	14	252	5	90	450	2250
16-20	13	18	234	9	117	1053	9477
	120		1080		0	2664	5424

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{1080}{120} = 9$$

$$\mu_2 = \sigma_2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{\sum f_i} = \frac{2664}{120} = 22{,}2$$

$$\sigma = \sqrt{22{,}2} = 4{,}71$$

$$\mu_3 = \frac{\sum f_i \left(X_i - \overline{X}\right)^3}{\sum f_i} = \frac{5424}{120} = 45{,}2$$

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = \frac{45{,}2}{(4{,}71)^3} = \frac{45{,}2}{104{,}48} = 0{,}43$$

$\alpha_3 > 0$ olduğundan seri sağa asimetriktir.

4.2.2.3. Momentler ile Hesaplanan Basıklık Ölçüsü

Momentler ile basıklık ölçüsü hesaplanırken, ortalamaya göre 4. moment, ortalamaya göre 2. momentin karesine bölünür ve elde edilen basıklık ölçüsü α_4 ile ifade edilir.

α_4 katsayısı aşağıda verilen formül ile hesaplanır.

$$\alpha_4 = \frac{\mu_4}{\mu_2^2}$$

$\mu_2 = \sigma^2$ olduğundan

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

olacaktır. Momentler ile hesaplanan basıklık ölçüsünde serilerin basıklığı normal dağılımın basıklığı ile karşılaştırılmaktadır. Normal dağılım için α_4 = 3 olduğundan,

α_4 = 3 ise seri normal,
α_4 < 3 ise seri basık,
α_4 > 3 ise seri sivri veya yüksek

olacaktır.

ÖRNEK: Daha önce asimetrisini incelediğiniz sınıflı serinin basıklığını inceleyiniz.

Sınıflar	f_i	X_i	$X_i - \overline{X}$	$f_i(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^4$
0-4	15	2	-7	-5145	36015
4-8	44	6	-3	-1188	3564
8-12	30	10	1	30	30
12-16	18	14	5	2250	11250
16-20	13	18	9	9477	85293
	120			5424	136152

$\overline{X} = 9$

$\mu_4 = \sigma^2 = 22{,}2$ olarak hesaplanmıştı.

$$\mu_4 = \frac{\sum f_i \left(X_i - \overline{X}\right)^4}{\sum f_i} = \frac{136152}{120} = 1134{,}6$$

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = \frac{1134{,}6}{(22{,}2)^4} = \frac{1134{,}6}{492{,}84} = 2{,}230$$

$\alpha_4 < 3$ olduğundan seri basıktır.

4.3. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir işletme 8 satış personeli ile ansiklopedi pazarlamaktadır. Personelin bir haftalık satışları aşağıda verilmiştir.

Personel	Satışlar (Adet)
A	12
B	21
C	35
D	8
E	14
F	7
G	26
H	17

Serinin,

a) Değişim aralığını,
b) Varyans ve standart sapmasını hesaplayınız.

Çözüm:

a) DA = $X_{max.}$- X_{min} = 35- 7 = 28

b)

X_i	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$
12	-5,5	30,25
21	3,5	12,25
35	17,5	306,25
8	-9,5	90,25
14	-3,5	12,25
7	-10,5	110,25
26	8,5	72,25
17	-0,5	0,25
140	0	634,00

$$\overline{X} = \frac{\sum X_i}{n} = \frac{140}{8} = 17{,}5$$

$$\sigma^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n} = \frac{634}{8} = 79{,}25$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{79{,}25} = 8{,}90$$

ÖRNEK 2: Bir süper markete tesadüfi olarak seçilen bir günde 880 müşteri gelmiş ve gelen müşterilerin aileleri ile ilgili bir anket yapılmıştır. Müşterilerin ankette yer alan "Evde kaç kişi yaşıyorsunuz?" sorusuna verdikleri cevaplar aşağıdaki gibi düzenlenmiştir.

Evde Yaşayan Kişi Sayıları	Müşteri Sayıları
1	30
2	177
3	209
4	264
5	114
6	71
7	15
	880

Dağılımın,

a)Değişim aralığını,
b)Varyans ve standart sapmasını,
c)S değerini hesaplayınız.

Çözüm:

X_i	f_i	f_iX_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$
1	30	30	-2,6	6,76	202,80
2	177	354	-1,6	2,56	453,12
3	209	627	-0,6	0,36	75,24
4	264	1056	0,4	0,16	42,24
5	114	570	1,4	1,96	223,44
6	71	426	2,4	5,76	408,96
7	15	105	3,4	11,56	173,40
	880	3168			1579,20

a) DA = $X_{max.} - X_{min} = 7-1 = 6$

b) $$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{3168}{880} = 3,6$$

$$\sigma_2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i} = \frac{1579,20}{880} = 1,7945$$

$$\sigma = \sqrt{1,7945} = 1,3395$$

c)

$$S^2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i - 1} = \frac{1579,20}{880 - 1} = 1,7965$$

$$S = \sqrt{1,7965} = 1,3403$$

ÖRNEK 3: Bir turistik otele son 3 ayda gelen müşterilerin otelde kaldıkları gün sayılarına göre, düzenlenen seri aşağıda verilmiştir.

Kalınan Gün Sınıfları	Müşteri Sayısı
1-6	184
6-11	161

11-16	86
16-21	21
21-26	8
	460

Serinin,
a)Değişim aralığını,
b)Varyans ve standart sapmasını,
c) S değerini hesaplayınız.

Çözüm:

Sınıflar	f_i	X_i	f_iX_i	$f_iX_i^2$
1-6	184	3,5	644,0	2254,00
6-11	161	8,5	1368,5	11632,25
16-21	86	13,5	1161,0	15673,50
21-26	21	18,5	388,5	7187,25
	8	23,5	188,0	4418,00
	460		3750,0	41165,00

a) $DA = X_{max}.- X_{min} = 26\text{-}1=25$

b) $$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{3750}{460} = 8{,}15$$

$$\sigma_2 = \frac{\sum f_i X_i^{\,2}}{\sum f_i} - (\overline{X})^2 = \frac{41165}{} - (8{,}15)^2$$

$$= 89{,}4891 - 66{,}4225 = 23{,}0666$$

$$\sigma = \sqrt{23{,}0666} = 4{,}8027$$

c) $$S^2 = \frac{n\sigma^2}{n-1} = \frac{460(23{,}0666)}{460-1} = 23{,}1168$$

ÖRNEK 4: A işletmesinin 2009 yılı aylık satışlarının ortalaması 108,2 milyon TL. Ve standart sapması 19,7 milyon TL.dır. Ayni mali satan B işletmesinin ise 2009 yılı aylık satışlarının ortalamasının 465 adet ve standart sapmasının 38,4 adet olduğu belirlenmiştir. Bu dağılımları karşılaştırınız.

Çözüm:

$$DK_A = \frac{\sigma_A}{\bar{X}_A}.100 = \frac{19,7}{108,2}100 = \%18,20$$

$$DK_B = \frac{\sigma_B}{\bar{X}_B}.100 = \frac{38,4}{465}100 = \%8,25$$

B işletmesinde A işletmesine göre, aritmetik ortalama etrafında daha yoğun bir dağılım olduğu görülmektedir.

ÖRNEK 5: Bir serinin aritmetik, kareli ve geometrik ortalamaları aşağıda verilmiştir.

16,8 ; 11,7 ; 13,4

Serinin standart sapmasını hesaplayınız.

Çözüm: Bir seride aritmetik, kareli ve geometrik ortalamalar hesaplandığında,

$$G < \bar{X} < K$$

olduğundan,
G= 11.,7
$\bar{X}$ = 13,4
K= 16,8 olacaktır.

$$\sigma = \sqrt{K^2 - (\bar{X})^2}$$

$$= \sqrt{(16,8)^2 - (13,4)^2}$$

$$= \sqrt{282,24 - 179,56} = \sqrt{102,68} = 10,13$$

ÖRNEK 6: Bir seri için aşağıdaki değerler hesaplanmıştır. Serinin standart sapmasını hesaplayınız.

$$\sum_{i=1}^{180} X_i = 1476$$

$$\sum_{i=1}^{180} X_i^2 = 14580$$

Çözüm:

$$\sigma = \sqrt{\frac{\sum X_i^2}{n} - \left(\frac{\sum X_i}{n}\right)^2}$$

$$= \sqrt{\frac{14580}{180} - \left(\frac{1476}{180}\right)^2}$$

$$= \sqrt{81 - 67{,}24} = \sqrt{13{,}76} = 3{,}7$$

ÖRNEK 7: Aşağıda verilen değerlerden yararlanarak serinin aritmetik ortalamasını hesaplayınız.

$$\sum_{i=1}^{50} (X_i - \overline{X})^2 = 7650$$

$$\sum_{i=1}^{50} X_i^2 = 36450$$

Çözüm:

$$\sigma^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n} = \frac{7650}{50} = 153$$

$$K^2 = \frac{\sum X_i^2}{n} = \frac{36450}{50} = 729$$

$$\sigma^2 = K^2 - \overline{X}^2$$

$$\overline{X}^2 = K^2 - \sigma^2$$

$$\overline{X}^2 = 729 - 153 = 576$$

$$\overline{X} = \sqrt{576} = 24$$

ÖRNEK 8: Aşağıda bir seri için hesaplanan değerler verilmiştir. Serinin kareli ortalamasını hesaplayınız.

$$\sum_{i=1}^{64} (X_i - \overline{X})^2 = 409600$$

$$\sum_{i=1}^{64} X_i = 9600$$

Çözüm:

$$\sigma^2 = \frac{\sum \left(X_i - \overline{X}\right)^2}{n} = \frac{409600}{64} = 6400$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{9600}{64} = 150$$

$$\sigma^2 = K^2 - \overline{X}^2$$

$$K^2 = \sigma^2 + \overline{X}^2$$

$$K^2 = 6400 + (150)^2 = 28900$$

$$K = \sqrt{28900} = 170$$

ÖRNEK 9: Bir işletmede 220 işçi çalışmaktadır. İşçilerin aylık net ücretlerinin aritmetik ortalaması 1972800.- TL ve standart sapması 886500 TL. olarak belirlenmiştir. İşçilere 1200000 TL. Artı % 28 zam yapılması halinde, yeni Ücret dağılımının aritmetik ortalamasını ve standart sapmasını hesaplayınız.

Çözüm:

$$\overline{Y} = 1200000 + 1{,}28\ \overline{X}$$

$$= 1200000 + 1{,}28\ (1972800)$$
$$= 3725184.\text{- TL.}$$

$$\sigma_y = 1{,}28\ \sigma_X$$

$$= 1{,}28\ (886500)$$
$$= 1134720.\text{- TL.}$$

ÖRNEK 10: Bir bölgede yaşayan ailelerin aylık ortalama gelirlerini belirlemek amacı ile iki ayrı araştırma yapılmıştır. Birinci araştırma 175 aileyi kapsamaktadır. Bu araştırma sonucunda aylık gelir ortalaması 3,5 milyon ve standart sapma 0,9 milyon lira olarak bulunmuştur. İkinci araştırma ise 280 aileyi kapsamaktadır. Bu araştırmaya göre ise aylık gelir ortalaması 3,2 milyon ve standart sapma 1,1 milyon lira olarak bulunmuştur. İki araştırmanın verilerini birleştirerek aylık gelirlerin ortalamasını ve

standart sapmasını hesaplayınız.

Çözüm:

$n_1 = 175 \quad \bar{X}_1 = 3,5 \qquad \sigma_1 = 0,9$

$n_2 = 280 \quad \bar{X}_2 = 3,2 \qquad \sigma_2 = 1,1$

$$\bar{Y} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{175(3,5) + 280(3,2)}{175 + 280}$$

$$= \frac{1508,5}{455} = 3,3$$

$$\sigma_Y^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1(\bar{X}_1 - \bar{Y})^2 + n_2(\bar{X}_2 - \bar{Y})^2}{n_1 + n_2}$$

$$= \frac{175(0,9)^2 + 2808(1,1)^2}{175 + 280} + \frac{175(3,5 - 3,3)^2 + 280(3,2 - 3,3)^2}{175 + 280}$$

$$= \frac{480,55}{455} + \frac{9,8}{455} = 1,08$$

$$\sigma_Y^2 = 1,038$$

ÖRNEK 11: Yapılan bir araştırma ile eşlerin yıllık gelirleri incelenmiş ve 120 aileyi kapsayan araştırma sonuçlarına göre erkeklerin gelirlerinin ortalaması 48 milyon ve standart sapması 7 milyon lira, kadınların gelirlerinin ortalaması 21 milyon ve standart sapması 6 milyon lira olarak bulunmuştur. Bu ailelerin yıllık gelirlerinin ortalamasını ve standart sapmasını hesaplayınız. (Kovaryans 18 olarak hesaplanmıştır.)

Çözüm:

$$\bar{Y} = \bar{X}_E + \bar{X}_K$$

$$= 48 + 21 = 67$$

$$\sigma_Y^2 = \sigma_E^2 + \sigma_K^2 + 2\sigma_{EK}$$

$$= (7)^2 + (6)^2 + 2(18)$$

$$= 49 + 36 + 2(18) = 121$$

$$\sigma_Y = \sqrt{121} = 11$$

ÖRNEK 12: Aşağıda 60 öğrencinin 5 üzerinden aldıkları matematik notları verilmiştir.

Notlar	Öğrenci Sayısı
1	7
2	13
3	21
4	10
5	9
	60

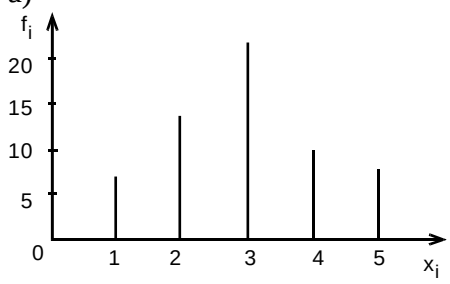
Bu not dağılımının, asimetri durumunu,

a)Grafiğini çizerek,
b)Aritmetik ortalama, mod ve medyan yardımı ile,
c)Pearson katsayıları ile,
d)Bowley asimetri ölçüsü ile inceleyiniz.

Çözüm:

X_i	f_i	f_iX_i	Artan Birikimli Frekanslar	$f_iX_i^2$
1	7	7	7	7
2	13	26	20	52
3	21	63	41	189
4	10	40	51	160
5	9	45	60	225
	60	181		633

a)



Grafikten serinin dağılımının simetrik dağılıma yakın olduğu anlaşılmaktadır.

b) $\overline{X} = \dfrac{\sum f_i X_i}{\sum f_i} = \dfrac{181}{60} = 3{,}01$

$M_0 = 3$

$$\frac{n}{2} = \frac{60}{2} = 30$$

$M_e = 3$

$$M_0 = M_e \cong \overline{X}$$

olduğundan dağılım simetrik olarak kabul edilebilir.

c) $$\sigma = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - (\overline{X})^2}$$

$$= \sqrt{\frac{633}{60} - (3{,}01)^2} = \sqrt{1{,}4899} = 1{,}22$$

$$\alpha_1 = \frac{\overline{X} - M_0}{\sigma} = \frac{3{,}01 - 3}{1{,}22} = 0{,}008$$

$$\alpha_2 = \frac{3(\overline{X} - M_e)}{\sigma} = \frac{3(3{,}01 - 3)}{1{,}22} = 0{,}024$$

α_1 ve $\alpha_2 > 0$ olduğundan seri hafif sağa asimetriktir. Simetrik kabul edilebilir.

d) $$\frac{n}{4} = \frac{60}{4} = 15$$

$Q_{1/4} = 2$

$Q_{1/4} = M_e = 3$

$$\frac{3n}{4} = \frac{3(60)}{4} = 45$$

$Q_{3/4} = 4$

$$A_S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

$$= \frac{(4-3)-(3-2)}{(4-3)+(3-2)} = 0$$

olduğundan seri simetriktir.

ÖRNEK 13:Aşağıda bir fabrikada çalışan 240 işçinin Ocak ayı toplam fazla mesai saatlerinin dağılımı verilmiştir.

Fazla Mesai Saati Sınıfları	İşçi Sayıları
0-20	23
20-40	45
40-60	121
60-80	37
80-100	14
	240

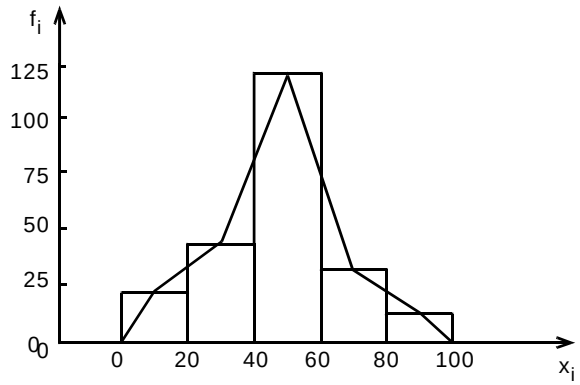
Serinin simetri durumunu,

a)Grafiğini çizerek,
b)Aritmetik ortalama, mod ve medyan yardımı ile,
c)Pearson katsayıları ile,
d)Bowley asimetri ölçüsü ile inceleyiniz.

Çözüm:

Sınıflar	f_i	X_i	f_iX_i	Artan Birikimli Frekanslar	f_iX^2
0-20	23	10	230	23	2300
20-40	45	30	1350	68	40500
40-60	121	50	6050	189	302500
60-80	37	70	2590	226	181300
80-100	14	90	1260	240	113400
	240		11480		640000

a)



Grafikten serinin dağılımının simetrik dağılıma yakın olduğu anlaşılmaktadır.

b) $\overline{X} = \dfrac{\sum f_i X_i}{\sum f_i} = \dfrac{11480}{240} = 47{,}83$

$$M_0 = I_a + \frac{\Delta_1}{\Delta_1 + \Delta_2}.c$$

$$M_0 = 40 + \frac{76}{76+84}20 = 49{,}5$$

$$M_e = I_a + \frac{\frac{n}{2} - \sum_{i=1}^{k} f_i}{f_m}.c$$

$$\frac{n}{2} = \frac{240}{2} = 120$$

$$M_e = 40 + \frac{120-68}{121}.20 = 48{,}59$$

$$47{,}83 < 48{,}59 < 49{,}5$$

$$\overline{X} < M_e < M_o$$

olduğundan seri sola asimetriktir.

c)

$$\sigma = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - (\overline{X})^2}$$

$$= \sqrt{\frac{640000}{240} - (47{,}83)^2} = \sqrt{378{,}9577} = 19{,}46$$

$$\alpha_1 = \frac{\overline{X} - M_0}{\sigma} = \frac{47{,}83 - 49{,}5}{19{,}46} = -0{,}085$$

$$\alpha_2 = \frac{3(\overline{X} - M_e)}{\sigma} = \frac{3(47{,}83 - 4859)}{19{,}46} = -0{,}117$$

α_1 ve $\alpha_2 < 0$ olduğundan seri sola asimetriktir.

d) $$Q_{h/r} = I_a + \frac{n\frac{h}{r} - \sum_{i=1}^{k-1} f_i}{f_m}.c$$

$$\frac{n}{4} = \frac{240}{4} = 60$$

$$Q_{1/4} = 20 + \frac{60 - 23}{45} 20 = 36{,}44$$

$Q_{2/4} = M_e = 48{,}59$

$$\frac{3n}{4} = \frac{3(240)}{4} = 180$$

$$Q_{3/4} = 40 + \frac{180 - 68}{121} 20 = 58{,}51$$

$$A_S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

$$= \frac{(58,51 - 48,59) - (48,59 - 36,44)}{(58,51 - 48,59) + (48,59 - 36,44)}$$

$$= \frac{9,92 - 12,15}{9,92 + 12,15} = \frac{-2,23}{22,07} = -0,101$$

olduğundan seri sola asimetriktir.

ÖRNEK 14: Aşağıda 50 personelin tesadüfi olarak seçilen bir hafta içinde işe geç geldikleri gün sayıları verilmiştir.

Gün Sayısı	Personel Sayısı
0	12
1	21
2	9
3	6
4	2
	50

Bu serinin asimetri durumu ve basıklığını inceleyiniz.

Çözüm:

X_i	f_i	f_iX_i	$X_i - \overline{X}$	$f_i(X_i - \overline{X})$	$f_i(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^4$
0	12	0	-1,3	-15,6	20,28	-26,364	34,2732
1	21	21	-0,3	-6,3	1,89	-0,567	0,1701
2	9	18	0,7	6,3	4,41	3,087	2,1609
3	6	18	1,7	10,2	17,34	29,478	50,1126
4	2	8	2,7	5,4	14,58	39,366	106,2882
	50	65		0	58,50	45,000	193,0050

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{65}{50} = 1,3$$

$$\mu_2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i} = \frac{58,50}{50} = 1,176$$

$$\sigma = \sqrt{1,176} = 1,0844$$

$$\mu_3 = \frac{\sum f_i (X_i - \overline{X})^3}{\sum f_i} = \frac{45}{50} = 0{,}9$$

$$\mu_4 = \frac{\sum f_i (X_i - \overline{X})^4}{\sum f_i} = \frac{193{,}005}{50} = 3{,}8601$$

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = \frac{3{,}8601}{(1{,}176)^3} = 2{,}7911$$

$\alpha_3 > 0$ olduğundan seri sağa asimetriktir.

ÖRNEK 15: Aşağıda bir işletmede çalışan 150 personelin aylık net ücretlere göre Dağılımı verilmiştir. (1= 1000000.- TL)

Ücret Sınıfları	İşçi Sayısı
1-3	26
3-5	59
5-7	38
7-9	18
9-11	9
	150

Dağılımın asimetri ve basıklık ölçülerini hesaplayınız.

Çözüm:

Sınıflar	f_i	X_i	f_iX_i	$X_i - \overline{X}$	$f_i(X_i - \overline{X})$	$f_i(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^4$
1-3	26	2	52	-3	-78	234	-702	2106
3-5	59	4	236	-1	-59	59	-59	59
5-7	38	6	228	1	38	38	38	38
7-9	18	8	144	3	54	162	486	1458
9-11	9	10	90	5	45	225	1125	5625
	150		750		0	718	888	9286

$$\overline{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{750}{150} = 5$$

$$\mu_2 = \frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i} = \frac{718}{150} = 4{,}7866$$

$$\sigma = \sqrt{4,7866} = 2,1878$$

$$\mu_3 = \frac{\sum f_i (X_i - \overline{X})^3}{\sum f_i} = \frac{888}{150} = 5,92$$

$$\mu_4 = \frac{\sum f_i (X_i - \overline{X})^4}{\sum f_i} = \frac{9286}{150} = 61,9066$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{5,92}{(2,1878)^3} = 0,56$$

$\alpha_3 > 0$ olduğundan seri sağa asimetriktir.

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{61,9066}{(4,7866)^2} = 2,70$$

$\alpha_4 < 3$ olduğundan seri basıktır.

ÖRNEK 16: Aşağıda 80 birimli bir serinin orjine göre momentleri verilmiştir.

$m_1 = 4$; $m_2 = 25$; $m_3 = 164$; $m_4 = 1263$

Serinin asimetri ve basıklık durumunu inceleyiniz.

Çözüm:

$$\mu_2 = m_2 - m_1^2$$
$$= 25 - (4)^2 = 9$$

$$\sigma = \sqrt{9} = 3$$

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3$$
$$= 164 - 3(4)(25) + 2(4)^3 = -8$$

$$\mu_4 = m_4 - 4m_1 m_2 + 6m_1^2 m_2 - 3m_1^4$$
$$= 1263 - 4(4)(164) + 6(4)^2(25) - 3(4)^4 = 271$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{-8}{(3)^3} = -0,29$$

$\alpha_3 < 0$ olduğundan seri sola asimetriktir.

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{271}{(9)^2} = 3,34$$

$\alpha_4 > 3$ olduğundan seri sivridir.

ÖRNEK 17: Aşağıda 200 kişinin 2009 yılı yıllık gelirlerinin dağılımı verilmiştir. (1=100 milyon TL)

Gelir Sınıfları	Kişi Sayısı
7,5-12,5	25
12,5-17,5	50
17,5-22,5	85
22,5-27,5	40
	200

Toplanma oranını hesaplayınız.

Çözüm:

Sınıflar	f_i	X_i	Artan Birikimli Frekanslar	f_iX_i	Birikimli f_iX_i	p_i	q_i	p_i-q_i
7,5-12,5	25	10	25	250	250	0,125	0,067	0,058
12,5-17,5	50	15	75	750	1000	0,270	0,270	0,105
17,5-22,5	85	20	160	1700	2700	0,729	0,729	0,071
22,5-27,5	40	25	200	1000	3700	1,000	1,000	0
	200					1,300		0,234

$$TO = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = \frac{0,234}{1,3} = 0,18$$

Toplanma oranı 0,18 olarak hesaplanmıştır. Bu sonuca göre gelir dağılımının adil olduğu söylenemez.

BÖLÜM 5

İNDEKSLER

Bazı istatistiksel araştırmalarda bir veya birkaç değişkenin mekan ve zaman içinde gösterdiği değişmelerin oransal olarak belirlenmesi istenebilir. Bu gibi durumlarda indekslerden yararlanılır. İndeksler birçok alanlarda kullanılabilmelerine rağmen daha çok iktisat ve işletme alanlarında kullanılmaktadır. İndeksler yardımı ile bir ülkede bölgeler arası üretim, fiyat, gelir dağılımı veya nüfus artışının zaman içinde gösterdiği değişmeler belirlenebilir.

5.1. İNDEKS KAVRAMI

İstatistik oranların bir türü olan indeksler kısaca sayısal değerlerin birbirlerine oranlanması ile elde edilir ve hesaplanabilmeleri için en az iki değerin bilinmesi gerekmektedir.

Bir olaya ait bir veya birden fazla değişkenin, farklı yerlerdeki veya zaman içindeki oransal değişimini belirlemek amacı ile hesaplanan sayılar, "indeks sayılar" olarak tanımlanır. İndeks sayılar, kısaca "indeksler" diye adlandırılmaktadır. İndeksler, fiyat, üretim, yatırım, ücret, satış değişmelerinin belirlenmesi gibi farklı olaylarda kullanılabilir.

İndeksler, hesaplandıkları seri türlerine göre zaman ve mekan indeksleri; hesaplandıkları madde sayısına göre basit ve bileşik indeksler olmak üzere ikiye ayrılırlar. Zaman ve mekan indeksleri aynı zamanda basit veya bileşik indeksler olabilir.

Zaman indeksleri hesaplanmalarında temel alınan devreye göre sabit ve değişken esaslı indeksler olarak, bileşik indeksleri ise hesaplanmalarına konu olan maddelerin önem derecelerine göre tartılı ve tartısız bileşik indeksler olarak ikiye ayrılırlar.

5.2. MEKÅN İNDEKSLERİ

Mekan serilerinin mekan içindeki oransal değişimini belirlemek için hesaplanan indekslere mekan indeksleri denir. Mekan indeksleri, nüfus, üretim, tüketim, fiyat gibi çeşitli değişkenlerin bölgeler, iller gibi mekan birimlerine göre oransal değişimlerinin belirlenmesi için kullanılırlar.

Mekan indeksleri seriyi oluşturan değerlerden her birinin serinin aritmetik ortalanmasına oranlanması ile hesaplanırlar.

$$I_i = \frac{X_i}{\overline{X}}.100$$

Bu formülde;

I_i : Hesaplanan indeksleri
X_i : Seri değerlerini
$\overline{X}$: Seri değerlerinin aritmetik ortalamasını,

göstermektedir. Mekan indekslerinde aritmetik ortalamanın değeri 100 kabul edildiğinden bulunan sonuçlar, 100 den küçük veya büyük olması durumuna göre yorumlanırlar.

ÖRNEK: Beş ayrı bölgenin 2008 yılı buğday üretim miktarları aşağıda verilmiştir. (1=10.000 ton)

Bölgeler	Buğday Üretimi (1=10.000 ton) (X_i)
A	15
B	30
C	9
D	27
E	39
	120

Bölgelerin mekan indekslerini hesaplayınız.

Bölgelerin mekan indekslerini hesaplamak için serinin aritmetik ortalaması hesaplanacaktır.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{120}{5} = 24$$

Mekan indeksleri:

$$I_i = \frac{X_i}{\overline{X}}.100$$

$$I_A = \frac{X_A}{\overline{X}}.100 = \frac{15}{24}100 = 62{,}5$$

$$I_B = \frac{X_B}{\overline{X}}.100 = \frac{30}{24}100 = 125$$

$$I_C = \frac{X_C}{\overline{X}}.100 = \frac{9}{24}100 = 37{,}5$$

$$I_D = \frac{X_D}{\overline{X}}.100 = \frac{27}{24}100 = 112{,}5$$

$$I_E = \frac{X_E}{\overline{X}}.100 = \frac{39}{24}100 = 162{,}5$$

Bu sonuçlara göre, 2008 yılı buğday üretim miktarı, A,B,C,D,E bölgeleri ortalamasına göre, A bölgesinde % 37,5 (=100 - 62,5) C bölgesinde % 62,5 (=100 - 37,5) daha az, B bölgesinde % 12,5 (= 112,5 -100) ve E bölgesinde % 62,5 (= 162,5 -100) daha fazladır.

Mekan indekslerinde, seri değerleri aritmetik ortalamaya oranlanıp, 100 ile çarpıldığından hesaplanan indekslerin ortalamaları 100 olacaktır.

62,5+125+37,5+112,5+162,5 = 500

İndekslerin ortalaması $= \frac{500}{5} = 100$

olduğu görülmektedir.

5.3. ZAMAN İNDEKSLERİ

Zaman serilerinin oransal değişimini belirlemek için hesaplanan indekslere zaman indeksleri denir. Herhangi bir zaman birimine ait indeksin hesaplanmasında, zaman serisinin herhangi bir zaman biriminin değeri temel alınarak, diğerleri bu değere oranlanabileceği gibi, her zaman biriminin değeri bir önceki zaman biriminin değerine de oranlanabilir. Buna dayanarak zaman indeksleri "sabit esaslı" ve "değişken esaslı" olarak ikiye ayrılır.

5.3.1. Sabit Esaslı İndeksler

Zaman indekslerinde hesaplanacak zaman birimlerini devre olarak adlandırırsak; belirli bir devrenin değerini temel alıp, diğer devrelerdeki oransal değişmeleri bu değere göre belirlemek için hesaplanan indekslere, "sabit esaslı indeksler" denir.

Sabit esaslı indekslerde hesaplanmaya temel alınan devre, "temel devre" veya "esas devre" olarak adlandırılır ve değeri 100'e eşittir. Diğer devrelerin indeksleri ise, değerlerinin esas devre değerine oranlanması ile bulunacak ve esas devreye göre oransal artışları veya azalışları belirtecektir.

Bu indeksleri,

$$I_i = \frac{X_i}{X_0}.100$$

formülü ile hesaplayabiliriz.

Bu formülde;

I_i : i'nci devrenin indeks sayısını
X_i : indeksi hesaplanacak devrenin değerini
X_0 : esas devre değerini göstermektedir.

İndekslerde değişken genellikle X_i ile gösterilirse de, fiyat indekslerinde pi, miktar indekslerinde ise q_i değişkenleri X_i yerine kullanılmaktadır. Bu nedenle

sabit esaslı fiyat indeksleri,

$$I_i = \frac{p_i}{p_0}.100$$

sabit esaslı miktar indeksleri,

$$I_i = \frac{q_i}{q_0}.100$$

formülleri ile gösterilecektir.

ÖRNEK: Aşağıda bir malın 6 yıllık, satış fiyatları verilmiştir.

Yıllar	(p_i) A malı fiyatı (TL)	I_i (2004=100)	I_j (2007=100)
2004	35	100,0	125,0
2005	40	114,2	142,8
2006	32	91,4	114,2
2007	28	80,0	100,0
2008	50	142,8	178,5
2009	73	208,5	260,7

Bu verilerden yararlanarak,

2004 yılı temel yıl ve 2007 yılı temel yıl olarak alındığına göre 6 yıllık sabit esaslı indeksleri hesaplayınız.

2004= 100 (2004 temel yıl) ise:

Fiyat indeksleri;

$$I_i = \frac{p_i}{p_0}.100 \qquad p_0 = p_{04} = 35$$

$$I_{04} = \frac{p_{04}}{p_{04}}.100 = \frac{35}{35}100 = 100,0$$

$$I_{05} = \frac{p_{05}}{p_{04}}.100 = \frac{40}{35}100 = 114,2$$

$$I_{06} = \frac{p_{06}}{p_{04}}.100 = \frac{32}{35}100 = 91,4$$

$$I_{07} = \frac{p_{07}}{p_{04}}.100 = \frac{28}{35}100 = 80,0$$

$$I_{08} = \frac{p_{08}}{p_{04}}.100 = \frac{50}{35}100 = 142,8$$

$$I_{09} = \frac{p_{09}}{p_{04}}.100 = \frac{73}{35}100 = 208,5$$

olacaktır. Malın fiyatları, 2004 yılına göre 2005 yılında % 14,2, 2003 yılında % 42,8, 2009 yılında % 108,5 artarken; 2006 yılında % 8,6, 2007 yılında % 20 azalmıştır.

Aynı yıllara ait indeksleri (2007 = 100) 2007 yılı temel yıl olmak üzere hesaplayalım.

$$I_i = \frac{p_i}{p_0}.100 \qquad p_0 = p_{07} = 28$$

$$I_{04} = \frac{p_{04}}{p_{07}}.100 = \frac{35}{28}100 = 125,0$$

$$I_{05} = \frac{p_{05}}{p_{07}}.100 = \frac{40}{28}100 = 142,8$$

$$I_{06} = \frac{p_{06}}{p_{07}}.100 = \frac{32}{28}100 = 114,2$$

$$I_{07} = \frac{p_{07}}{p_{07}}.100 = \frac{28}{28}100 = 100,0$$

$$I_{08} = \frac{p_{08}}{p_{07}}.100 = \frac{50}{28}100 = 178,5$$

$$I_{09} = \frac{p_{09}}{p_{07}}.100 = \frac{73}{28}100 = 260,7$$

Yorum: Malın fiyatları 2007 yılına göre 2004 yılında % 25, 2005 yılında % 42,8 2006 yılında % 14,2 2008 yılında % 78,5 ve 2009 yılında % 160,7 oranında yükselmiştir.

Sonuçların karşılaştırılması: 2004 ve 2007 yılı temel yıl olarak hesapladığımız indeksleri karşılaştırdığımızda aralarında büyük farklar olduğunu görürüz. Bu nedenle sabit esaslı indekslerin hesaplanmasında, temel devrenin indekslerin hesaplanma amaçlarına göre belirlenmesi gerekecektir.

Sabit esaslı indekslerin hesaplanmasında, bir devre yerine, birkaç devrelik bir dönemde temel alınabilir. Bu durumda birkaç devrelik dönemin aritmetik ortalaması alınarak, bu ortalama esas devre değeri kabul edilir ve diğer devre değerleri hesaplanan bu ortalamaya oranlanır. Ortalamanın hesaplanması için alınacak devre sayısı, indeks hesaplanmasının amacına göre belirlenir. Bu durumda temel dönemi oluşturan devrelerin indekslerinin aritmetik ortalaması 100 olacaktır.

ÖRNEK: Aşağıda bir bölgenin yıllık kömür üretim miktarları verilmiştir.

Yıllar	Kömür Üretimi (1=10000 Ton)
2003	11
2004	18
2005	16
2006	19
2007	13
2008	21
2009	24

Bölgenin yıllık kömür üretimine ait sabit esaslı indeksleri 2003-2005 dönemi temel alarak hesaplayınız.

$$\bar{q}_0 = \frac{q_{03} + q_{04} + q_{05}}{3} = \frac{11+18+16}{3} = \frac{45}{3} = 15$$

$$I_i = \frac{q_i}{q_0}.100$$

$$I_{03} = \frac{q_{03}}{q_{03-05}}.100 = \frac{11}{15}100 = 73,3$$

$$I_{04} = \frac{q_{04}}{q_{03-05}}.100 = \frac{18}{15}100 = 120,0$$

$$I_{05} = \frac{q_{05}}{q_{03-05}}.100 = \frac{16}{15}100 = 106,7$$

$$I_{06} = \frac{q_{06}}{q_{09-05}}.100 = \frac{19}{15}100 = 126,6$$

$$I_{07} = \frac{q_{07}}{q_{03-05}}.100 = \frac{13}{15}100 = 86,66$$

$$I_{08} = \frac{q_{08}}{q_{03-05}}.100 = \frac{21}{15}100 = 140,0$$

$$I_{09} = \frac{q_{09}}{q_{03-05}}.100 = \frac{24}{15}100 = 160,0$$

Yorum: Bölgenin yıllık kömür üretim miktarı 2003-2005 dönemine (2003-2005 yılları ortalamasına) göre; 2003'de % 16,7 daha az; 2004'de % 20, 2005'de % 6,7 2006'da % 26,6 2008'de % 40, 2009'da % 60 daha fazla 2007'de ise % 13,4 daha az olmuştur.

- Burada temel dönem yılları indekslerinin ortalaması,

$$I_{03-05} = \frac{I_{03} + I_{04} + I_{05}}{3} = \frac{73,3 + 120,0 + 106,7}{3}$$
$$= \frac{300}{3} = 100$$

olarak bulunur.

5.3.2. Değişken Esaslı İndeksler

Zaman indekslerinde, indeksi hesaplanacak devre değerinin bir önceki devre değerine oranlanması ile hesaplanacak indeksler "değişken esaslı indeksler" olarak adlandırılır.

Sabit esaslı indekslerdeki temel devre, değişken esaslı indekslerde indeksi hesaplanacak her devre için değişecektir ve değişken esaslı indeksler her devre indeksinin bir önceki devreye göre oransal artışını veya azalışını gösterecektir. Bu indekslerde hesaplanacak her devre için bir önceki devre 100 kabul edilmektedir.

Değişken esaslı indeksler,

$$I_i = \frac{X_i}{X_{i-1}} 100$$

formülü ile hesaplanmaktadır.

Sabit esaslı indekslerden farklı olarak formülde, X_{i-1} bir önceki devre değerini ifade etmektedir.

Değişken esaslı fiyat ve miktar indeksleri,

$$I_i = \frac{p_i}{p_{i-1}} 100$$

ve

$$I_i = \frac{q_i}{q_{i-1}} 100$$

formülü ile hesaplanmaktadır.

ÖRNEK: Aşağıda bir fabrika da üretilen yıllık buzdolabı miktarları verilmiştir.

Yıllar	Üretim Miktarı (q_i) (1=1000 Adet)	I_i (D.E)
2005	5	-
2006	4	80
2007	3	75
2008	6	200
2009	8	133

Bu verilerden yararlanarak değişken esaslı indeksleri hesaplayınız.

$$I_i = \frac{q_i}{q_{i-1}} 100$$

2005 yılı indeksi bir önceki devrenin değeri bilinmediğinden hesaplanamaz.

$$I_{06} = \frac{4}{5} 100 = 80$$

$$I_{07} = \frac{3}{4}100 = 75$$

$$I_{08} = \frac{6}{3}100 = 200$$

$$I_{09} = \frac{8}{6}100 = 133$$

Yorum: fabrikanın yıllık üretim miktarı, 2006 yılında 2005 yılına göre % 20 azalmış; 2007 yılında 2006 yılına göre % 25 azalmış; 2008 yılında 2007 yılına göre % 100 artmış; 2009 yılında 2008 yılına göre % 33 artmıştır.

Yukarıda incelenen ve ayrı ayrı örnek üzerinde hesaplanan sabit ve değişken esaslı indeksleri aynı seri üzerinde de hesaplayabiliriz.

ÖRNEK: Aşağıdaki serinin sabit ve değişken esaslı indekslerini hesaplayınız.

Devreler	X_i	I_i (S.E; 1=100)	I_i (D.E)
1	20	100	100
2	16	80	80
3	18	90	112,5
4	25	125	138,88
5	28	140	112
6	35	175	125

Sabit esaslı indeksler:

$$I_i = \frac{X_i}{X_0}100 \rightarrow (1 = 100)$$

$$I_1 = \frac{20}{20}100 = 100$$

$$I_2 = \frac{16}{20}100 = 80$$

$$I_3 = \frac{18}{20}100 = 90$$

$$I_4 = \frac{25}{20}100 = 125$$

$$I_5 = \frac{28}{20}100 = 140$$

$$I_6 = \frac{35}{20}100 = 175$$

Değişken esaslı indeksler:

$$I_i = \frac{X_i}{X_{i-1}} 100$$

$$I_1 = -$$

$$I_2 = \frac{16}{20} 100 = 80$$

$$I_3 = \frac{18}{16} 100 = 112{,}5$$

$$I_4 = \frac{25}{18} 100 = 138{,}88$$

$$I_5 = \frac{28}{25} 100 = 112$$

$$I_6 = \frac{35}{28} 100 = 125$$

5.3.3. Sabit Esaslı İndekslerden Değişken Esaslı İndekslerin Hesaplanması

Yapılan araştırma için değişken esaslı indeksler gerekiyorsa, ancak mevcut kaynaklardan sabit esaslı indeksler elde edilmişse, bu sabit esaslı indekslerin, değişken esaslı indekslere çevrilmesi gerekecektir.

Sabit esaslı indeksler,

$$I_{i,D} = \frac{I_{i,S}}{I_{i-1,S}} .100$$

formülü ile değişken esaslı indekslere dönüştürülmektedir.

ÖRNEK: Bir önceki örnekte hesaplanan sabit esaslı indeksler yardımı ile değişken esaslı indeksleri hesaplayınız.

Devreler	I_i (1=100)
1	100
2	80
3	90
4	125
5	140
6	175

$$I_{1,D} = -$$

$$I_{2,D} = \frac{80}{100}100 = 80$$

$$I_{3,D} = \frac{90}{80}100 = 112{,}5$$

$$I_{4,D} = \frac{125}{90}100 = 138{,}88$$

$$I_{5,D} = \frac{140}{125}100 = 112$$

$$I_{6,D} = \frac{175}{140}100 = 125$$

Görüldüğü gibi, bundan önceki örnekte gerçek verilerle hesaplanan sonuçlar ile, sabit esaslı indeksler yardımı ile hesaplanan sonuçlar aynıdır.

5.3.4. Değişken Esaslı İndekslerden Sabit Esaslı İndekslerin Hesaplanması

Yapılan araştırma için sabit esaslı indeksler gerekiyorsa, ancak mevcut kaynaklardan değişken esaslı indeksler elde edilmişse, bu değişken esaslı indekslerin sabit esaslı indekslere çevrilmesi gerekecektir.

Değişken esaslı indeksler

$$I_i^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}.\cdots.\frac{I_{i,D}}{100}100$$

formülü ile sabit esaslı indekslere dönüştürülebilir.

ÖRNEK: Bir önceki örnekte hesaplanan değişken esaslı indeksler yardımı ile sabit esaslı indeksleri hesaplayınız.

Devreler	I_i (D.E)
1	-
2	80
3	112,5
4	138,88
5	112
6	125

$$I_1^S = 100$$

$$I_2^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}100 = I_1^S \frac{I_{2,D}}{100} = 100\frac{80}{100} = 80$$

$$I_3^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}.\frac{I_{3.D}}{100}100 = I_2^S \frac{I_{3,D}}{100} = 80\frac{112,5}{100} = 90$$

$$I_4^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}.\frac{I_{3.D}}{100}\frac{I_{4,D}}{100}100$$

$$= I_3^S \frac{I_{4,D}}{100} = 90\frac{138,88}{100} = 125$$

$$I_5^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}.\frac{I_{3.D}}{100}\frac{I_{4,D}}{100}\frac{I_{5,D}}{100}100$$

$$= I_4^S \frac{I_{5,D}}{100} = 125\frac{112}{100} = 140$$

$$I_6^S = \frac{I_{1,D}}{100}.\frac{I_{2,D}}{100}.\frac{I_{3.D}}{100}\frac{I_{4,D}}{100}\frac{I_{5,D}}{100}\frac{I_{6,D}}{100}100$$

$$= I_5^S \frac{I_{6,D}}{100} = 140\frac{125}{100} = 175$$

Görüldüğü gibi yapılan dönüşümle önceki örnekte gerçek verilerle hesaplanan sonuçlar ile, sabit esaslı indeksler yardımı ile hesaplanan sonuçlar aynıdır.

5.3.5. Sabit Esaslı İndekslerde Temel Devrenin Değiştirilmesi

Herhangi bir devrenin esas alınması ile hesaplanan sabit esaslı indekslerde temel devre değiştirilebilir. Temel devreyi değiştirmek için kullanılacak formül aşağıdaki gibi olacaktır.

$$I_{i,y} = \frac{I_i}{I_{0,y}}100$$

Bu formülde;

$I_{i,y}$: Yeni temel devreye göre hesaplanacak indeksi

I_i : Sabit esaslı indeksi

$I_{o,y}$: Yeni temel devrenin indeksini göstermektedir.

ÖRNEK:

Yıllar	X_i	I_i (2000=100)
2005	5	100
2006	6	120
2007	10	200
2008	8	160
2009	12	240

Yukarıda 2005 yılı temel yıl olarak düzenlenmiş indekslerle, yapılan dönüşümün kontrol edilebilmesi için bu indekslerin hesaplandığı gerçek değerler verilmiştir.

Temel yılı 2007 olarak değiştirelim.

$$I_{05,y} = \frac{I_{05}}{I_{07}}100 = \frac{100}{200}100 = 50$$

$$I_{06,y} = \frac{I_{06}}{I_{07}}100 = \frac{120}{200}100 = 60$$

$$I_{07,y} = \frac{I_{07}}{I_{07}}100 = \frac{200}{200}100 = 100$$

$$I_{08,y} = \frac{I_{08}}{I_{07}}100 = \frac{160}{200}100 = 80$$

$$I_{09,y} = \frac{I_{09}}{I_{07}}100 = \frac{240}{200}100 = 120$$

Aynı indeksleri verilen gerçek değerleri ile hesaplayalım.

$$I_i = \frac{X_i}{X_0}100 \qquad (2007=100)$$

$$I_{05} = \frac{X_{05}}{X_{07}}100 = \frac{5}{10}100 = 50$$

$$I_{06} = \frac{X_{06}}{X_{07}}100 = \frac{6}{10}100 = 60$$

$$I_{07} = \frac{X_{07}}{X_{07}}100 = \frac{10}{10}100 = 100$$

$$I_{08} = \frac{X_{08}}{X_{07}}100 = \frac{8}{10}100 = 80$$

$$I_{09} = \frac{X_{09}}{X_{07}}100 = \frac{12}{10}100 = 120$$

Sonuçların temel devre değiştirilmesi ile elde edilen sonuçlarla aynı olduğu görülmektedir.

5.4. BİLEŞİK İNDEKSLER

İndekslerin hesaplandıkları madde sayısına göre basit ve bileşik indeksler olarak ikiye ayrıldığından daha önce söz etmiştik. Basit indeksler bir tek madde için hesaplanan indekslerdir. Buraya kadar yaptığımız açıklamalar basit indeksler ile ilgilidir. Bu bölümde ise bileşik indekslerle ilgili açıklamalar yer alacaktır.

Birden fazla madde için hesaplanan indekslere bileşik indeksler denilmektedir. Bir tek maddenin bir kaç madde yerine kullanılması sakıncalı olacağından, bu gibi durumlarda bileşik indeksler hesaplanmaktadır. Bileşik indeksler tartılı ve tartısız bileşik indeksler olmak üzere iki grupta incelenebilir. Ayrıca bileşik indeksler mekan ve zaman serileri için de hesaplanabilir.

5.4.1. Tartısız Bileşik İndeksler

Tartısız bileşik indeksler, indeks hesabına girecek olan maddelerin özelliklerine bağlı olarak farklı yöntemlerle hesaplanabilir.

5.4.1.1. Toplam Yöntemi

Bu yöntem de indeksler, indeks hesabına alınacak maddelerin, her devre için toplam değerleri bulunarak hesaplanır. Böylece toplam yönteminde sabit esaslı indeks formülü,

$$I_i = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_0} 100$$

Değişken esaslı indeks formülü ise,

$$I_i = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_{i-1}} 100$$

olacaktır. Bu formüllerde X yerine, fiyat indekslerinde p ve miktar indekslerinde q kullanılacaktır.

ÖRNEK: Aşağıda verilen 4 madde için tartısız bileşik fiyat indekslerini toplam yöntemi ile hesaplayalım.

Maddeler	YILLAR 2006	2007	2008	2009
A	10	10	12	16
B	125	105	100	130
C	5	6	6	9
D	26	30	20	35
Toplam	166	151	138	190

Sabit esaslı indeksler: (2006=100)

$$I_i = \frac{\sum p_i}{\sum p_0} 100$$

$$I_{06} = \frac{p_{06}}{p_{06}} 100 = \frac{166}{166} 100 = 100,0$$

$$I_{07} = \frac{p_{07}}{p_{06}} 100 = \frac{151}{166} 100 = 90,9$$

$$I_{08} = \frac{p_{08}}{p_{06}} 100 = \frac{138}{166} 100 = 83,1$$

$$I_{09} = \frac{p_{09}}{p_{06}} 100 = \frac{190}{166} 100 = 114,4$$

Dört maddenin fiyatları 2006 yılına göre 2007 yılında % 9,1 ve 2008 yılında % 16,9 azalmış, 2009 yılında ise % 14,04 artmıştır.

Değişken esaslı indeksler:

$$I_i = \frac{\sum p_i}{\sum p_{i-1}} 100$$

$$I_{06} = \frac{p_{06}}{p_{00}} 100 = -$$

$$I_{07} = \frac{p_{07}}{p_{06}} 100 = \frac{151}{166} 100 = 90,9$$

$$I_{08} = \frac{p_{08}}{p_{07}} 100 = \frac{138}{151} 100 = 91,3$$

$$I_{09} = \frac{p_{09}}{p_{08}} 100 = \frac{190}{138} 100 = 137,6$$

Dört maddenin fiyatları, 2007 yılında 2006'ya göre % 9,1 ve 2008 yılında 2007 yılına göre % 8,7 azalmış, 2009 yılında ise 2008 yılına göre % 37,6 artmıştır.

5.4.1.2. Ortalamaların Kullanıldığı Yöntemler

Toplam yönteminde, indeks hesabında kullanılan maddelerin değerlendirilmesinde kullanılan ölçüler farklı olabilir. Farklı ölçülerin kullanıldığı durumlarda farklı sonuçlar elde edilecektir. Örneğin, bir önceki örnekte A malının çay olduğunu ve 100 gr.lık paket fiyatının verildiğini düşünürsek; fiyatı kilograma çevirmekle A malının fiyatı ile bileşik indekslerin hesaplanmasında kullanılan toplam fiyatları ve sonuç değişecektir. Bu değişiklikten kaçınmak için bileşik indekslerin hesabında ortalama kullanılmaktadır. Ortalama kullanımda, gerçek değerlerin yerine, her değişken için hesaplanan basit indekslerin ortalaması alınarak, ölçülerin farklılıkları giderilmiş olur.

- **Aritmetik Ortalama**

Tartısız bileşik indekslerin hesaplanmasında, değişkenlerin basit indekslerinin aritmetik ortalamasının kullanılması durumunda,

$$I_{ib} = \frac{\sum_{j=1}^{n} I_{j,i}}{n}$$

formülü uygulanacaktır.

Bu formülde;

I_{ib} : Bileşik indeksi
I_{ji} : Basit indeksleri
n : Madde sayısını

göstermektedir.

Yukarıda görülen formülü tartısız bileşik indeks için yazarsak;

Sabit esaslı indeksler için;

$$I_{ib} = \frac{\sum_{j=1}^{n} \frac{p_{ji}}{p_{j0}} 100}{n} = \frac{100}{n} \sum_{j=1}^{n} \frac{p_{ji}}{p_0}$$

Değişken esaslı indeksler için,

$$I_{ib} = \frac{\sum_{j=1}^{n} \frac{p_{ji}}{p_{j(i-1)}} 100}{n} = \frac{100}{n} \sum_{j=1}^{n} \frac{p_{ji}}{p_{j(i-1)}}$$

formülleri elde edilir.

Formüllerden de anlaşılacağı gibi burada iki farklı yolla hesaplama olanağı vardır.

- İlk olarak her devre için basit indeksleri toplayıp değişken sayısına bölerek
- İkinci olarak $\frac{p_i}{p_{i-1}}$ veya $\frac{p_i}{p_0}$ oranlarını hesaplayıp ve bu oranların toplamını $\frac{100}{n}$ ile çarparak bileşik indeksleri hesaplayabiliriz.

Bu hesaplama şekilleri, basit indekslerin aritmetik ortalaması veya oranların aritmetik ortalaması yöntemleri olarak adlandırılabilir. Bu yöntemlerden her ikisinin kullanılması durumunda aynı sonuçlar elde edilecektir.

ÖRNEK: A, B ve C malları yıllık ortalama fiyatları aşağıda verilmiştir. (1=1 TL)

	YILLAR		
Maddeler	2007	2008	2009
A	3	5	6
B	100	110	130
C	48	48	60

Yukarıda verilen fiyatlardan yararlanarak bileşik fiyat indekslerini basit indekslerin ortalaması ve oranların ortalaması yöntemleri ile hesaplayınız.

Basit indekslerin aritmetik ortalaması yöntemi:

Sabit esaslı (2007=100)

$$I_i = \frac{p_i}{p_0} 100$$

A Malı

$$I_{07} = \frac{3}{3} 100 = 100$$

$$I_{08} = \frac{5}{3}100 = 166,6$$

$$I_{09} = \frac{6}{3}100 = 200$$

B Malı

$$I_{07} = \frac{100}{100}100 = 100$$

$$I_{08} = \frac{110}{100}100 = 110$$

$$I_{09} = \frac{130}{100}100 = 130$$

C Malı

$$I_{07} = \frac{48}{48}100 = 100$$

$$I_{08} = \frac{48}{48}100 = 100$$

$$I_{09} = \frac{60}{48}100 = 125$$

Sonuçları bir tabloda toplayalım.

		(2007=100)	
Maddeler	2007	2008	2009
A	100	166,6	200
B	100	110	130
C	100	100	125

Bileşik indeksler: (2007=100)

$$I_i = \frac{\sum_{i=1}^{n} I_{ij}}{n}$$

$$I_{07} = \frac{100+100+100}{3} = \frac{300}{3} = 100$$

$$I_{08} = \frac{166,6+110+100}{3} = \frac{376,6}{3} = 125,53$$

$$I_{09} = \frac{200+130+125}{3} = \frac{455}{3} = 151,66$$

Değişken esaslı

$$I_i = \frac{p_i}{p_{i-1}} 100$$

A Malı

$$I_{07} = \frac{3}{3} 100 = 100$$

$$I_{08} = \frac{5}{3} 100 = 166,6$$

$$I_{09} = \frac{6}{5} 100 = 120$$

B Malı

$$I_{07} = \frac{100}{100} 100 = 100$$

$$I_{08} = \frac{110}{100} 100 = 110$$

$$I_{09} = \frac{130}{110} 100 = 118,1$$

C Malı

$$I_{07} = \frac{48}{48} 100 = 100$$

$$I_{08} = \frac{48}{48} 100 = 100$$

$$I_{09} = \frac{60}{48} 100 = 125$$

Sonuçları bir tabloda toplayalım.

Maddeler	2007	2008	2009
		(D.E)	
A	100	166,6	120
B	100	110	118,1
C	100	100	125

Bileşik indeksler:

$$I_{07} = \frac{100+100+100}{3} = \frac{300}{3} = 100$$

$$I_{08} = \frac{166,6+110+100}{3} = \frac{376,6}{3} = 125,53$$

$$I_{09} = \frac{120+118,1+125}{3} = \frac{363,1}{3} = 121$$

- **Oranların Aritmetik Ortalaması**

Oranların aritmetik ortalaması ile yapılacak hesaplamalar, basit indekslerin aritmetik ortalaması yöntemi ile aynı temele dayandığından aynı sonucu verecektir.

Sabit esaslı (2007=100)

Maddeler	p_{07}/p_{07}	p_{08}/p_{07}	p_{09}/p_{07}
		Fiyat Oranları	
A	3/3=1	5/3=1,666	6/3=2
B	100/100=1	110/100=1,1	130/100=1,3
C	48/48=1	48/48=1	60/48=1,25
Toplam	3	3,766	4,55

$$I_{ib} = \frac{100}{n}\sum_{j=1}^{n}\frac{p_{ji}}{p_0}$$

$$I_{07} = \frac{100}{3}(3) = 100$$

$$I_{08} = \frac{100}{3}(3,766) = 125,56$$

$$I_{09} = \frac{100}{3}(4,55) = 151,66$$

Değişken esaslı

2007 ve 2008 yılı tartısız bileşik indeksleri aynı yılların sabit esaslı indeksleri ile aynı sonucu vereceğinden sadece 2009 yılı indeksi hesaplanacaktır.

Maddeler	p_{09}/p_{07}
A	6/5=1,2
B	130/110=1,181
C	60/48=1,25
Toplam	3,631

$$I_{ib} = \frac{100}{n}\sum_{j=1}^{n}\frac{p_{ji}}{p_{j(i-1)}}$$

$$I_{09} = \frac{100}{3}(3,631) = 121,0$$

Görüldüğü gibi her iki yöntemde aynı sonuçları vermektedir.

Bileşik indekslerin hesaplanmasında aritmetik ortalama kullanılıyorsa, aritmetik ortalamanın serinin anormal değerlerinin etkisinde kalacağı unutulmamalıdır. Ayrıca tartısız bileşik indekslerin hesaplanmasında da aritmetik ortalamanın kullanıldığı durumlarda sabit esaslı indeksler ile değişken esaslı indeksler, veya değişken esaslı indeksler ile sabit esaslı indeksler hesaplanıyorsa bu hesaplanan indeksler gerçek değerlerinden sapacaktır. Aynı şekilde temel devre değiştirildiğinde de indeksler gerçek değerlerinden farklı olacaktır.

- **Geometrik Ortalama**

Bileşik indekslerin hesaplanmasında aritmetik ortalama kullanılabildiği gibi geometrik ortalamada kullanılabilir. Ancak hangi ortalamanın kullanılacağına karar verilmeden önce geometrik ve aritmetik ortalamanın özelliklerinin göz önünde bulundurulması gereklidir.

Geometrik ortalamanın hesaplanması, aritmetik ortalamaya göre daha zor olmasına rağmen indekslerin hesabında geometrik ortalama kullanımı aritmetik ortalamaya göre bazı yararlarda sağlayabilir. Geometrik ortalama serinin anormal değerlerinden aritmetik ortalama kadar etkilenmez. Ayrıca, indekslerin hesaplanmasında geometrik ortalama kullanılıyorsa, aritmetik ortalama kullanımında söz edilen sapmalar olmayacaktır.

Bileşik indekslerin hesaplanmasında geometrik ortalama kullanıldığında formüller,

Sabit esaslı indeksler için;

$$I_i = \sqrt[n]{\prod_{j=1}^{n}\frac{p_{ji}}{p_{j0}}}100$$

Değişken esaslı indeksler için

$$I_i = \sqrt[n]{\prod_{j=1}^{n} \frac{p_{ji}}{p_{j(i-1)}}} 100$$

olacaktır.

Hesaplamada logaritma kullanılıyorsa formüller.

$$\log I_i = \frac{1}{n} \sum \left(\log p_{ji} - \log p_{j0} \right) + \log 100$$

$$\log I_i = \frac{1}{n} \sum \left(\log p_{ji} - \log p_{j(i-1)} \right) + \log 100$$

olarak elde edilir.

ÖRNEK: Aşağıda A, B ve C mallarının fiyatları verilmiştir. (1=1 TL)

	YILLAR		
Maddeler	2007	2008	2009
A	20	19	27
B	75	75	80
C	550	600	590

2007 yılını temel yıl alarak 2009 yılı sabit ve değişken esaslı bileşik indekslerini geometrik ortalama ile hesaplayınız.

Sabit esaslı (2007=100)

$$\log I_{09} = \frac{1}{3}\left[\left(\log 27 + \log 80 + \log 590 \right) - \left(\log 20 + \log 75 + \log 550 \right) \right] + \log 100$$

$$= \frac{1}{3}\left[(1{,}43136 + 1{,}90309 + 2{,}77085) - (1{,}30103 + 1{,}87506 + 2{,}74036) \right] + 2$$

$$= \frac{1}{3}\left[6{,}10530 - 5{,}91645 \right] + 2$$

$$= \frac{1}{3}(0{,}18885) + 2 = 2{,}06295 \rightarrow \quad I_{09} = 115{,}6$$

Değişken esaslı

$$\log I_{09} = \frac{1}{3}\left[\left(\log 27 + \log 80 + \log 590\right) - \left(\log 19 + \log 75 + \log 600\right)\right] + \log 100$$

$$= \frac{1}{3}\left[\left(1{,}43136 + 1{,}90309 + 2{,}77085\right) - \left(1{,}27875 + 1{,}87506 + 2{,}77815\right)\right] + 2$$

$$= \frac{1}{3}\left[6{,}10530 - 5{,}93196\right] + 2$$

$$= \frac{1}{3}(0{,}17334) + 2 = 2{,}05778 \rightarrow \quad I_{09} = 114{,}2$$

5.4.2. Tartılı Birleşik İndeksler

Tartısız bileşik indeksler ile ilgili açıklamalarda, indekslerin hesaplanmasına giren maddelerin önem derecelerinin dikkate alınmadığı görülmektedir. Ancak bileşik indekslerin hesabına giren maddelerin önem derecelerinin farklı olduğu durumlarda tartısız bileşik indekslerin hesaplanması yanıltıcı sonuçlar verecektir. Bu nedenle bu gibi durumlarda tartılı bileşik indekslerin hesaplanması daha uygun olacaktır.

Maddelerin önem derecelerine göre dikkate alınarak hesaplanan bileşik indeksler tartılı bileşik indeks olarak adlandırılır. Tartı, malların önem derecelerini belirten ağırlıklardır. Tartılar tahmini olarak belirlenebilirse de, araştırma yapılarak belirlenmesi daha iyi sonuçlar verecektir.

Tartılı bileşik indeksler için, zaman indekslerinde, sabit ve değişken esaslı indeksler hesaplanabilir. Tartılar t ile ifade edildiğinde, sabit esaslı tartılı bileşik indeksler için

$$I_i = \frac{\sum_{j=1}^{n} \frac{X_{ji}}{X_{j0}} t_j}{\sum_{i=1}^{n} t_j} 100$$

Değişken esaslı bileşik indeksler için,

$$I_i = \frac{\sum_{j=1}^{n} \frac{X_{ji}}{X_{j(i-1)}} t_j}{\sum_{i=1}^{n} t_j} 100$$

formülleri kullanılacaktır. Formüllerden de anlaşılacağı gibi yapılan işlem, tartısız bileşik indekslerin tartılı aritmetik ortalamasının hesaplanmasıdır.

Tartılı bileşik indekslerde bir malın fiyat veya miktar indeksinin hesaplanmasında tartı olarak malın kıymeti (değeri) alınmaktadır. Malın kıymeti miktarı ile fiyatının çarpımı ile hesaplanır. Böylece malın kıymeti,

$$t = p.q$$

olarak gösterilebilir. Tartının belirlenmesinde $t = p_o.q_o$ alınırsa; her hangi bir devrenin fiyat indeksi,

$$I = \frac{\sum \frac{p_i}{p_0}.t}{\sum p_0 q_0}100 = \frac{\sum \frac{p_i}{p_0} p_0 q_0}{\sum p_0 q_0}100 = \frac{\sum p_i q_0}{\sum p_0 q_0}100$$

olacaktır. Bu formüle "Laspeyres" formülü denmektedir.

$t = p_o q_i$ alınırsa,

$$I = \frac{\sum p_i q_i}{\sum p_0 q_i}100$$

olarak elde edilir. Bu formüle "Paasche" formülü denmektedir.

$t= p_o(q_o+ q_i)$ alınırsa,

$$I = \frac{\sum p_i (q_0 + q_i)}{\sum p_0 (q_0 + q_i)}100$$

olarak elde edilir. Bu formülü de "Edgeworth Marshall" formülü denmektedir.

Laspeyres ve Paasche formüllerinin çarpımlarının kare kökü alınarak elde edilen formüle "Fisher" veya "İdeal İndeks" formülü denilmektedir.

$$I = \sqrt{I_L.I_P}$$

formülü açıldığında,

$$I = \sqrt{\frac{\sum p_i q_0}{\sum p_0 q_0}.100.\frac{\sum p_i q_i}{\sum p_0 q_i}}$$

$$I = 100.\sqrt{\frac{\sum p_i q_0}{\sum p_0 q_0}.\frac{\sum p_i q_i}{\sum p_0 q_i}}$$

şeklini almaktadır.

Sabit ve değişken esaslı, fiyat ve miktar indeksleri için, Laspeyres, Paasche, Edgeworth- Marshal ve Fisher formülleri aşağıda görülmektedir.

Laspeyres (I_L)

Fiyat indeksi:

Sabit esaslı

$$I = \frac{\sum p_i q_0}{\sum p_0 q_0} 100$$

Değişken esaslı

$$I = \frac{\sum p_i q_{i-1}}{\sum p_{i-1} q_{i-1}} 100$$

Miktar indeksi :

Sabit esaslı

$$I = \frac{\sum q_i p_0}{\sum q_0 p_0} 100$$

Değişken esaslı

$$I = \frac{\sum q_i p_{i-1}}{\sum q_{i-1} p_{i-1}} 100$$

Paasche (I_P)

Fiyat indeksi :

Sabit esaslı

$$I = \frac{\sum p_i q_i}{\sum p_0 q_i} 100$$

Değişken esaslı

$$I = \frac{\sum p_i q_i}{\sum p_{i-1} q_i} 100$$

Miktar indeksi:

Sabit esaslı

$$I = \frac{\sum q_i p_i}{\sum q_0 p_i} 100$$

Değişken esaslı

$$I = \frac{\sum q_i p_i}{\sum q_{i-1} p_i} 100$$

Edgeworth- Marshall (I_E)

Fiyat indeksi :

Sabit esaslı

$$I = \frac{\sum p_i(q_0 + q_i)}{\sum p_0(q_0 + q_i)} 100$$

Değişken esaslı

$$I = \frac{\sum p_i(q_{i-1} + q_i)}{\sum p_{i-1}(q_{i-1} + q_i)} 100$$

Miktar indeksi :

Sabit esaslı

$$I = \frac{\sum q_i(p_0 + p_i)}{\sum q_0(p_0 + p_i)} 100$$

Değişken esaslı

$$I = \frac{\sum q_i(p_{i-1} + p_i)}{\sum q_{i-1}(p_{i-1} + p_i)} 100$$

Fisher (= İdeal İndeks)

$$I = \sqrt{I_L . I_P}$$

ÖRNEK: Aşağıda A, B ve C mallarının üç yıllık fiyatları ve miktarları verilmiştir. (1=1 TL; 1=100 gram)

	A Malı		B Malı		C Malı	
Yıllar	Fiyat(p)	Miktar(q)	Fiyat(p)	Miktar(q)	Fiyat(p)	Miktar(q)
2007	5	10	20	60	15	25
2008	6	14	25	60	20	30
2009	6	15	40	50	30	35

Bu bilgilerden yararlanarak,

a) 2007 yılı temel yıl olmak üzere, 2009 yılı sabit esaslı fiyat indeksini, Laspeyres, Paasche, Edgeworth-Marshal ve Fisher formülleri ile hesaplayınız.

b) Aynı şekilde 2009 yılı değişken esaslı bileşik fiyat indeksini hesaplayınız.

a) 2009 yılı sabit esaslı indeksi (2007=100)

Laspeyres:

$$I = \frac{\sum p_i q_0}{\sum p_0 q_0} 100$$

$$I_{09} = \frac{(6.10)+(40.60)+(30.25)}{(5.10)+(20.60)+(15.25)}100$$

$$= \frac{60+2400+750}{50+1200+375}100 = \frac{3210}{1625}100 = 197,5$$

2009 yılında fiyatlar 2007 yılına göre %97,5 artmıştır.

Paasche :

$$I = \frac{\sum p_i q_i}{\sum p_0 q_i}100$$

$$I_{09} = \frac{(6.15)+(40.50)+(30.35)}{(5.15)+(20.50)+(15.35)}100$$

$$= \frac{90+2000+1050}{75+1000+525}100 = \frac{3140}{1600}100 = 196,6$$

2009 yılında fiyatlar 2007 yılına göre % 96,6 artmıştır.

Edgeworth - Marshall:

$$I = \frac{\sum p_i (q_0 + q_i)}{\sum p_0 (q_0 + q_i)}100$$

$$I_{09} = \frac{6(10+15)+40(60+50)+30(25+35)}{5(10+15)+20(60+50)+15(25+35)}100$$

$$I_{09} = \frac{150+4400+1800}{125+2200+900}100 = \frac{6350}{3225}100 = 196,8$$

2009 yılında fiyatlar 2007 yılına göre % 96,8 artmıştır.

Fisher:

$$I = \sqrt{I_L . I_P}$$

$$I_{09} = \sqrt{197,5.196,6} = \sqrt{38828,5} = 197,0$$

2009 yılında fiyatlar 2007 yılına göre % 97 artmıştır.

b) 2009 yılı değişken esaslı indeksi

Laspeyres :

$$I = \frac{\sum p_i q_{i-1}}{\sum p_{i-1} q_{i-1}} 100$$

$$I_{09} = \frac{(6.14) + (40.60) + (30.30)}{(6.14) + (25.60) + (20\text{x}30)} 100$$

$$= \frac{84 + 2400 + 900}{84 + 1500 + 600} 100 = \frac{3384}{2184} 100 = 154,9$$

2009 yılında fiyatlar 2008 yılına göre % 54,9 artmıştır.

Paasche :

$$I = \frac{\sum p_i q_i}{\sum p_{i-1} q_i} 100$$

$$I_{09} = \frac{(6.15) + (40.50) + (30.35)}{(6.15) + (25.50) + (20.35)} 100$$

$$= \frac{90 + 2000 + 1050}{90 + 1250 + 700} 100 = \frac{3140}{2040} 100 = 153,9$$

2009 yılında fiyatlar 2008 yılına göre % 53,9 artmıştır.

Edgeworth-Marshall

$$I = \frac{\sum p_i (q_{i-1} + q_i)}{\sum p_{i-1} (q_{i-1} + q_i)} 100$$

$$I_{09} = \frac{6(14+15) + 40(50+60) + 30(30+35)}{6(14+15) + 25(50+60) + 20(30+35)} 100$$

$$I_{07} = \frac{174 + 4400 + 1950}{174 + 2750 + 1300} 100 = \frac{6524}{4224} 100 = 154,45$$

2009 yılında fiyatlar 2008 yılına göre % 54,45 artmıştır.

Fisher :

$$I = \sqrt{I_L . I_P}$$

$$I_{09} = \sqrt{154,9.153,9} = \sqrt{23839,11} = 154,3$$

2009 yılında fiyatlar 2008 yılına göre % 54,3 artmıştır.

ÖRNEK: Aşağıda üç maddenin üç yıllık fiyatları ve miktarları verilmiştir. (1=1 TL, 1=100 gram)

	Yıllar					
	2007		2008		2009	
Maddeler	P	Q	P	Q	P	Q
A	10	3	12	2	13	5
B	55	10	60	10	80	8
C	7	20	8	30	10	30

p: Fiyat
q: Miktar

a) 2007 yılı temel yıl olmak üzere 2009 yılı bileşik miktar indeksini, Laspeyres, Paasche, Edgeworth-Marshall ve Fisher formülleri ile hesaplayınız.
b) 2008 ve 2009 yılları değişken esaslı bileşik miktar indekslerini aynı formüllerle hesaplayınız.

a) 2009 yılı sabit esaslı indeksi (2007 = 100)

Laspeyres :

$$I = \frac{\sum q_i p_0}{\sum q_0 p_0} 100$$

$$I_{09} = \frac{(5.10) + (8.55) + (30.7)}{(3.10) + (10.55) + (20.7)} 100$$

$$= \frac{50 + 440 + 210}{30 + 550 + 140} 100 = \frac{700}{720} 100 = 97,5$$

2009 yılında miktarlar 2007 yılına göre % 2.8 azalmıştır.

Paasche:

$$I = \frac{\sum q_i p_i}{\sum q_0 p_i} 100$$

$$I_{09} = \frac{(5.13) + (8.80) + (30.10)}{(3.13) + (10.80) + (20.10)} 100$$

$$= \frac{65 + 640 + 300}{39 + 800 + 200} 100 = \frac{1005}{1039} 100 = 96,7$$

2009 yılında miktarlar 2007 yılına göre % 3,3 azalmıştır.

Edgeworth - Marshall:

$$I = \frac{\sum q_i (p_0 + p_i)}{\sum q_0 (p_0 + p_i)} 100$$

$$I_{09} = \frac{5(10+13) + 8(55+80) + 30(7+10)}{3(10+13) + 10(55+80) + 20(7+10)} 100$$

$$I_{09} = \frac{115 + 1080 + 510}{69 + 1350 + 340} 100 = \frac{1705}{1759} 100 = 96,9$$

2009 yılında miktarlar 2007 yılına göre % 3,1 azalmıştır.

Fisher:

$$I = \sqrt{I_L . I_P}$$

$$I_{09} = \sqrt{97, 2.96, 7} = \sqrt{9399, 24} = 96,9$$

2009 yılında miktarlar 2007 yılına göre % 3,1 azalmıştır.

b) 2008 ve 2009 yılları değişken esaslı indeksleri

Laspeyres:

$$I = \frac{\sum q_i p_{i-1}}{\sum q_{i-1} p_{i-1}} 100$$

$$I_{08} = \frac{(2.10)+(10.55)+(30.7)}{(3.10)+(10.55)+(20.7)}100$$

$$= \frac{20+550+210}{30+550+140}100 = \frac{780}{720}100 = 108,3$$

$$I_{09} = \frac{(5.12)+(8.60)+(30.8)}{(2.12)+(10.60)+(30.8)}100$$

$$= \frac{50+480+240}{24+600+240}100 = \frac{770}{864}100 = 89,1$$

Paasche:

$$I = \frac{\sum q_i p_i}{\sum q_{i-1} p_i}100$$

$$I_{08} = \frac{(2.12)+(10.60)+(30.8)}{(3.12)+(10.60)+(20.8)}100$$

$$= \frac{24+600+240}{36+600+160}100 = \frac{864}{796}100 = 108,6$$

$$I_{09} = \frac{(5.13)+(8.80)+(30.10)}{(2.13)+(10.80)+(30.10)}100$$

$$= \frac{65+640+300}{26+800+300}100 = \frac{1005}{1126}100 = 89,2$$

Edgeworth – Marshall:

$$I = \frac{\sum q_i (p_{i-1} + p_i)}{\sum q_{i-1} (p_{i-1} + p_i)}100$$

$$I_{08} = \frac{2(10+12)+10(60+55)+30(7+8)}{3(10+12)+10(60+55)+20(7+8)}100$$

$$I_{08} = \frac{44+1050+450}{66+1050+300}100 = \frac{1544}{1416}100 = 109,0$$

$$I_{09} = \frac{5(12+13)+8(60+80)+30(8+10)}{2(12+13)+10(60+80)+30(8+10)}100$$

$$I_{09} = \frac{125+1120+540}{50+1400+540}100 = \frac{1785}{1990}100 = 89,6$$

Fisher:

$$I = \sqrt{I_L . I_P}$$

$$I_{08} = \sqrt{108,3.108,6} = \sqrt{11761,38} = 108,4$$

$$I_{09} = \sqrt{90,2.89,2} = \sqrt{8045,84} = 89,7$$

5.5. CARİ FİYATLAR İLE HESAPLANMIŞ VERİLERİN SABİT FİYATLAR İLE HESAPLANMIŞ VERİLERE İNDİRGENMESİ

Cari fiyatlar ile hesaplanmış veriler enflasyondan etkilendiklerinden, farklı devrelerin değerlerinin karşılaştırılmasında gerçek artış veya azalışlar belirlenemeyecektir. Gerçek artış veya azalışların belirlenebilmesi için enflasyonun etkisinin ortadan kaldırılması gerekir. Enflasyonun etkisinin ortadan kaldırılması için, cari fiyatlar ile hesaplanmış değerlerin, fiyat indekslerinin 100'e bölünmesi ile elde edilen değere bölünerek, sabit fiyatlar ile hesaplanmış değerlere indirgenmiş olması gerekir. Bu indirgeme işlemi

$$X_{i,S} = \frac{X_{i,C}}{I_i / 100} = \frac{X_{i,C}}{I_i} 100$$

formülü ile yapılmaktadır.

Bu formülde,
$X_{i,s}$: Sabit fiyatlar ile hesaplanan değerleri
$X_{i,c}$: Cari fiyatlar ile hesaplanan değerleri
I_i : Sabit esaslı fiyat indekslerini ifade etmektedir.

İndirgeme işleminin yapılabilmesi için indekslerin sabit esaslı olması gerekmektedir. Değişken esaslı indekslerin bulunması durumunda bu indeksler sabit esaslı indekslere dönüştürülecektir. Ayrıca sabit esaslı indeksin temel devresi, indirgemede kullanılacak devreden farklı olduğu durumlarda sabit esaslı indeksin temel devresinin değiştirilmesi gerekecektir.

ÖRNEK: Aşağıda bir işçinin 4 yıllık ortalama aylık gelirleri ile toptan eşya fiyat indeksleri verilmiştir. (1=1 TL)

Yıllar	Ortalama Aylık Gelir (1=1 TL)	Toptan Eşya Fiyat İndeksi
2006	500	100
2007	800	172
2008	1425	283
2009	2500	469

İşçinin 2007, 2008 ve 2009 yılı ortalama aylık gelirlerini, 2006 yılı fiyatlarına indirgeyiniz.

$$X_{i,S} = \frac{X_{i,C}}{I_i} 100$$

$$X_{07,06} = \frac{800}{172} 100 = 465,11$$

$$X_{08,06} = \frac{1452}{283} 100 = 513,07$$

$$X_{00,06} = \frac{2500}{469} 100 = 533,04$$

İşçinin 2006 yılı fiyatları ile hesaplanan 2007, 2008 ve 20094 yılları ortalama aylık gelirlerinin 2006 yılı ortalama aylık gelirinden daha düşük olduğu görülmektedir.

5.6. GELİR FİYAT VE MİKTAR İNDEKSLERİ ARASINDAKİ İLİŞKİ

Fiyat ile miktar çarpımı geliri vereceğinden, her devre için fiyat indeksi ile miktar indekslerinin çarpımının 100'e bölümü gelir indeksine eşit olacaktır.

Gelir indeksi

$$I_G = \frac{G_i}{G_0} 100$$

$G_i = p_i q_i$: İndeksi hesaplanacak devrenin geliri
$G_o = p_o q_o$: Temel devrenin geliri
olacağından,

$$I_G = \frac{p_i q_i}{p_0 q_0} 100$$

olacak ve fiyat indeksi,

$$I_F = \frac{p_i}{p_0} 100$$

miktar indeksi

$$I_M = \frac{q_i}{q_0} 100$$

olduğundan I_G eşitliği $\frac{100}{100}$ ile çarpıldığında

$$I_G = \frac{p_i}{p_0}\frac{q_i}{q_0}100\frac{100}{100}$$

$$I_G = \left(\frac{p_i}{p_0}100\right)\left(\frac{q_i}{q_0}100\right)\frac{1}{100}$$

$$I_G = \frac{I_F . I_M}{100}$$

olarak elde edilir.

5.7. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Aşağıda bir malın 2009 yılının ilk 6 ayındaki fiyatları verilmiştir.

Aylar	Fiyatlar (1=1 TL)
Ocak	500
Şubat	500
Mart	490
Nisan	545
Mayıs	600
Haziran	575

a) Mayıs ve Haziran ayı sabit esaslı indekslerini, ocak ayını esas devre alarak hesaplayınız.
b) Mayıs ve haziran aylarına ait değişken esaslı indeksleri hesaplayınız.
c) Sonuçları karşılaştırınız.

Çözüm:
a) Sabit esaslı indeksler

$$I_i = \frac{p_i}{p_0}100$$

$$I_{may} = \frac{p_{may}}{p_{ocak}}100 = \frac{600}{500}100 = 120$$

$$I_{haz.} = \frac{p_{haz.}}{p_{ocak}}100 = \frac{575}{500}100 = 115$$

b) Değişken esaslı indeksler

$$I_i = \frac{p_i}{p_{i-1}} 100$$

$$I_{may} = \frac{p_{may}}{p_{nis.}} 100 = \frac{600}{545} 100 = 110$$

$$I_{haz.} = \frac{p_{haz.}}{p_{may.}} 100 = \frac{575}{600} 100 = 95{,}83 \approx 96$$

c) Mayıs ayı:

A malının fiyatı Mayıs ayında, Ocak ayına göre %20 artarken; Nisan ayına göre %10 artmıştır.

Haziran ayı:

A malının fiyatı Haziran ayında, Ocak ayına göre %15 artarken; Mayıs ayına göre %4 düşmüştür.

ÖRNEK 2: Aşağıda 5 devrenin sabit esaslı indeksleri verilmiştir.

Devreler	SEİ(1=100)
1	100
2	87
3	126
4	219
5	171

a) 3. Devre temel devre olduğuna göre 2. ve 5. devrelerin sabit esaslı indekslerini hesaplayınız.
b) 3. ve 5. devrelerin değişken esaslı indekslerini hesaplayınız.

Çözüm:

a)

$$I_{i,y} = \frac{I_i}{I_0} 100$$

$$I_2 = \frac{I_2}{I_3} 100 = \frac{87}{126} 100 = 69{,}04$$

$$I_5 = \frac{I_5}{I_3} 100 = \frac{171}{126} 100 = 135{,}71$$

b)

$$I_{i,D} = \frac{I_{i,S}}{I_{i,1,S}} 100$$

$$I_{3,D} = \frac{I_{3,S}}{I_{2,S}} 100 = \frac{126}{87} 100 = 144{,}82$$

$$I_{5,D} = \frac{I_{5,S}}{I_{4,S}} 100 = \frac{171}{219} 100 = 78{,}08$$

ÖRNEK 3: Aşağıda 6 yıllık değişken esaslı indeksler verilmiştir.

Yıllar	DEI
2004	122
2005	147
2006	93
2007	138
2008	224
2009	166

2006 yılı temel yıl olduğuna göre 2008 yılı sabit esaslı indeksini hesaplayınız.

Çözüm:

$$I_{i,S} = \frac{I_{1,D}}{100} . \frac{I_{2,D}}{100} . \cdots . 100$$

$$I_{03,S} = \frac{122}{100} . \frac{147}{100} . \frac{93}{100} . \frac{138}{100} . \frac{224}{100} 100 = 515{,}56$$

2007 yılı sabit esaslı indeksi 2008 yılına göre oransal değişimi verdiğinden, hesaplanan 2008 yılı değişken esaslı indeksinin temel yılı 2006 yılıdır. Bu nedenle 2006 yılı sabit esaslı indeksinin de hesaplanması ve temel devrenin değiştirilmesi gerekmektedir.

$$I_{06,S} = \frac{122}{100} . \frac{147}{100} . \frac{93}{100} 100 = 166{,}78$$

$$I_{08} = \frac{I_{03}}{I_{01}} 100 = \frac{515{,}56}{166{,}78} 100 = 309{,}12$$

2006 yılı temel yıl olduğunda, 2008 yılı sabit esaslı indeksi 309,12 olarak hesaplanmıştır.

Bu çözüm yolu uzun bir yol olup, konunun açıklanması için verilmiştir. Aynı indeksin aşağıda görüldüğü gibi daha kısa yol ile hesaplanması mümkündür.

Temel yıl 2006 olacağına göre 2007 yılı değişken esaslı indeksi 2006 yılına göre oransal değişimi verecektir. Bu nedenle temel devre olmasını istediğimiz devreyi takip eden devreler için verilen formül uygulanırsa, yine aynı sonuçlar elde edilecektir.

$$I_{06,S} = \frac{I_{07,D}}{100} . \frac{I_{08,D}}{100} 100$$

$$= \frac{138}{100} . \frac{224}{100} . 100 = 309,12$$

ÖRNEK 4: Aşağıda 4 maddenin 3 yıllık fiyatları verilmiştir. (1=1 TL)

	Fiyatlar			
Yıllar	A	B	C	D
2007	250	110	1050	460
2008	300	90	1200	450
2009	400	160	1750	500

a) 2007 yılı temel yıl olduğuna göre 2009 yılı sabit esaslı tartısız bileşik fiyat indeksini "basit indekslerin aritmetik ortalaması" yönetimi ile
b) 2009 yılı değişken esaslı tartısız bileşik fiyat indeksini "oranların aritmetik ortalaması" yönetimi ile hesaplayınız.

Çözüm:

a) $$I_i = \frac{p_i}{p_0} 100$$

Sabit esaslı basit indeksler (2007 = 100)

$$I_{A,09} = \frac{4000}{2500} 100 = 160,0$$

$$I_{B,09} = \frac{1600}{1100} 100 = 145,4$$

$$I_{C,09} = \frac{17500}{10500} 100 = 166,6$$

$$I_{D,09} = \frac{5000}{4600} 100 = 108,6$$

Tartısız bileşik indeks

$$I_{i,B} = \frac{\sum I_i}{n}$$

$$I_{09,B} = \frac{160,0+145,4+166,6+108,6}{4} = \frac{580,6}{4} = 145,15$$

b)

$$I_i = \frac{100}{4}\left[\frac{4000}{3000}+\frac{1600}{900}+\frac{17500}{12000}+\frac{5000}{4500}\right]$$

$$= \frac{100}{4}\left[1,333+1,777+1,458+1,111\right]$$

$$= 25(5,679) = 141,975$$

ÖRNEK 5: Aşağıda 4 malın 5 yıllık fiyatları ve miktarları verilmiştir. (1=1 TL, 1=100 gram)

	Mallar							
	A		B		C		D	
Yıllar	p	q	p	q	P	q	p	q
2005	10	20	105	58	47	10	5	13
2006	8	23	105	60	50	9	6	15
2007	10	21	110	60	52	9	8	16
2008	12	18	108	65	48	12	7	13
2009	15	22	115	67	49	12	7	17

a) 2008 yılı sabit esaslı bileşik fiyat indeksini 2006 yılı temel yıl olduğuna göre Laspeyres formülü ile hesaplayınız.

b) 2009 yılının 2008 yılını temel alan bileşik miktar indeksini Paasche formülü ile hesaplayınız.

c) 2005 yılı sabit esaslı bileşik miktar indeksini 2008 yılı temel yıl olduğuna göre Fisher formülü ile hesaplayınız.

Çözüm:

a)

$$I = \frac{\sum p_i q_0}{\sum p_o q_0} 100$$

$$I_{08} = \frac{(12,23)+(108.60)+(48.9)+(7.15)}{(8.23)+(105.60)+(50.9)+(6.15)}100$$

$$= \frac{276+6480+432+105}{184+6300+450+90}100 = \frac{7293}{7024}100 = 103,829$$

2008 yılında fiyatlar 2006 yılına göre % 3,8 artmıştır.

b) $$I = \frac{\sum q_i p_i}{\sum q_0 p_i}100$$

$$I_{04} = \frac{(20.10)+(58.105)+(10.47)+(13.5)}{(18.10)+(65.105)+(12.47)+(13.5)}100$$

$$= \frac{200+6090+470+65}{180+6825+564+65}100 = \frac{6825}{7634}100 = 89,4$$

c)

$$I_L = \frac{\sum q_i p_0}{\sum q_0 p_0}100$$

$$I_{05} = \frac{(20.12)+(58.108)+(10.48)+(7.13)}{(12.18)+(108.65)+(48.12)+(7.13)}100$$

$$= \frac{240+6264+480+91}{216+7020+576+91}100 = \frac{7075}{7903}100 = 89,5$$

$$I_P = 89,4$$

$$I_F = I_{00} = \sqrt{89,5.89,4} = \sqrt{80001,3} = 89,44$$

2005 yılında fiyatlar 2008 yılına göre % 10,56 daha azdır.

ÖRNEK 6: Aşağıda bir kişinin 3 yıllık gelirleri ile yıllık enflasyon oranları verilmiştir. (1=1000 TL)

Yıllar	Yıllık Gelir (1=1000TL)	Yıllık Enflasyon Oranı
2007	25	0,75
2008	34	0,57
2009	63	0,68

2007 yılı fiyatları ile 2008 ve 2009 yılı gerçek gelirlerini hesaplayarak, cari gelir artışları ile karşılaştırınız.

Çözüm:

2008 yılında fiyat artışı 2007 yılına göre 0,57 olduğundan 2008 yılı değişken esaslı fiyat artış indeksi 157 ve aynı şekilde 2009 yılı değişken esaslı fiyat artış indeksi 168 olacaktır. Bu indekslerin sabit esaslı indekslere çevrilerek kullanılması gerekmektedir. 2007 yılı temel yıl olacağından 2008 yılı sabit ve değişken esaslı indeksleri aynı olacaktır. 2007 yılı temel yıl olduğuna göre 2009 yılı sabit esaslı indeksi ise,

$$I_{09,S} = \frac{I_{08,D}}{100} \cdot \frac{I_{09,S}}{100} 100$$

$$= \frac{157}{100} \frac{168}{100} 100 = 263,76$$

olarak bulunur. Hesaplanan sabit esaslı indeksler ile gerçek yıllık gelirler,

$$X_{i,S} = \frac{X_{i,C}}{I_i} 100$$

$$X_{03} = \frac{34}{157} 100 = 21,65$$

$$X_{09} = \frac{63}{263,76} 100 = 23,88$$

olarak bulunur.

Cari fiyatlar ile gelir artışları ise,

$$I_{08} = \frac{34}{25} 100 = 136,0$$

$$I_{09} = \frac{63}{25} 100 = 252,0$$

olacaktır. Cari fiyatlar ile gelir 2007 yılına göre 2008 yılında %36; 2009 yılında % 152 artmış ve 2007 yılı fiyatları ile indeksler,

$$I_{08} = \frac{21,65}{25} 100 = 86,6$$

$$I_{09} = \frac{23,88}{25} 100 = 95,52$$

olacaktır. Gelir 2007 yılına göre 2008 yılında % 13,4 ve 2009 yılında % 4,48 azalmıştır.

BÖLÜM 6

OLASILIK

Olasılık kavramının günlük yaşantımızda oldukça büyük yeri vardır. Karşılaştığımız olayların hepsi belirli değildir ve belirsiz olaylarda olabilir, yapılabilir, belki gibi kelimelerle olasılıktan sözedebiliriz. Bu nedenle olasılık kavramı hakkında hemen herkesin birbirinden farklı da olsa bir fikri vardır.

İstatistikte olasılık teorisi farklı şekillerde açıklanabilmektedir. Oldukça geniş olarak ele alınabilecek olasılık teorisi bu bölümde özet olarak açıklanacaktır.

6.1. ÖRNEK UZAYI, PERMÜTASYON VE KOMBİNASYON KAVRAMLARI

Olasılık teorisinde bilinmesi gereken önemli kavramlardan biri "örnek uzayı"dır. Bir deneme de gerçekleşmesi mümkün olan tüm sonuçların oluşturacağı küme, örnek uzayı olarak tanımlanır. Örnek uzayının elemanları denemede oluşabilecek belirlenmiş sonuçlardır. Örneğin, 6 yüzlü bir zar atıldığında zarın 1,2,3,4,5 ve 6 noktalı yüzlerinden biri gelebilir. Bu örnek uzayının 6 elemanı vardır.

Deneme kavramı zar atma, kart çekme gibi denemelerin yanında doğal olayları da içermektedir.

Olasılık teorisinin açıklanması, bir kümenin belirli bir sayıda elemandan oluşan kaç adet farklı alt kümesi olabileceğinin hesaplanması, alt kümelerin elemanlarının kaç farklı şekilde sıralanabileceğinin belirlenmesi gibi durumlarda permütasyon ve kombinasyon hesaplarına başvurulmasını gerektirmektedir. Bu nedenle burada kısaca permütasyon ve kombinasyon kavramlarına yer verilecektir.

6.1.1. Permütasyon

n sayıda farklı elemanın kaç değişik sayıda sıralanabileceğinin belirlenmesi için başvurulan işlem permütasyondur.

n farklı sayıda elemanın permütasyonu, P_n ile gösterilir ve

$P_n = n! = 1.2.3.4.....(n-2).(n-1).n$

dir. n!, n faktöriyel olarak tanımlanır.

ÖRNEK: a,b,c gibi 3 eleman kaç farklı şekilde sıralanabilir, hesaplayınız.

$P_n = n!$

$P_3 = 3! = 1.2.3 = 6$ farklı şekilde sıralanabilir.

Elemanları sıralayalım.

$$\left.\begin{array}{l} a,b,c \\ a,c,b \\ b,a,c \\ b,c,a \\ c,a,b \\ c,b,a \end{array}\right\} 6$$

Görüldüğü gibi 3 farklı eleman 6 farklı şekilde sıralanabilmektedir.

ÖRNEK: 9 kişiden oluşan bir grup öğrenci kaç farklı şekilde sıralanabilir?

$P_n = n!$
$P_9 = 9! = 1.2.3.4.5.6.7.8.9 = 362.880$ farklı şekilde sıralanabilir.

Tekrarlı Permütasyon

n elemanın farklı diziliş sayısı araştırılırken, bu elemanlardan bazıları birbirinin aynı olabilir. Farklı elemanların sıralanmasında iki elemanın yer değiştirmesi bile sıralamayı değiştirirken, aynı elemanların yer değiştirmesi sıralamayı değiştirmeyecektir. Bu nedenle sıralama sayısı da değişeceğinden tekrarlı permütasyon işlemine başvurulacaktır.

n permütasyon hesabına konu olacak eleman sayısını n_1, n_2, n_3, ..., n_k aynı elemanların sayılarını gösteriyorsa tekrarlı permütasyon,

$$P_n^{n_1,n_2,.....,n_k} = \frac{n!}{n_1!n_2!.....n_k!}$$

olarak hesaplanır.

ÖRNEK: a, b, c, c gibi 4 eleman kaç farklı şekilde sıralanabilir? (2 tekrarlı)

$$P_n^{n_1,n_2} = \frac{n!}{n_1!n_2!}$$

$$P_4^2 = \frac{4!}{2!} = \frac{1.2.3.4}{1.2} = 12$$ farklı şekilde sıralanabilir.

Elemanları sıralayalım.

$$\left.\begin{matrix} a,b,c,c \\ a,c,b,c \\ a,c,c,b \\ b,a,c,c \\ b,c,a,c \\ b,c,c,a \end{matrix}\right\} 6 \quad + \quad \left.\begin{matrix} c,a,b,c \\ c,a,c,b \\ c,c,a,b \\ c,c,b,a \\ c,b,a,c \\ c,b,c,a \end{matrix}\right\} 6 \quad = \quad 12$$

Görüldüğü gibi 2'si tekrarlı 4 eleman 12 farklı şekilde sıralanabilmektedir.

6.1.2. Kombinasyon

n sayıda farklı elemandan, elemanların diziliş sırasına bakılmaksızın k sayıda elemandan oluşan kaç farklı grup düzenlenebileceğinin belirlenmesinde başvurulan işlem kombinasyondur.

Tanımdan da anlaşılacağı gibi k sayıda elemandan oluşacak grubun eleman sıralaması önemli değildir. Diziliş sırası önemli olduğunda varyasyon işlemi yapılması gerekmektedir.

n sayıda farklı elemandan düzenlenecek, k sayıda elemandan oluşacak grup sayısı $\binom{k}{n}$ veya C_n^k şeklinde gösterilebilir ve

$$C_n^k = \frac{n!}{k!(n-k)!} = \frac{1.2.3..... ...n}{1.2.3..... k.\ 1.2.3..... .(n-k-2)(n-k-1)(n-k)}$$

olarak hesaplanır.

ÖRNEK: a,b,c,d elemanlarından elemanların sırasına bakılmaksızın 2'şer elemandan oluşan kaç farklı grup düzenlenebilir?

$$C_n^k = \frac{n!}{k!(n-k)!}$$

$$C_4^2 = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{1.2.3.4}{1.2.1.2} = 6$$

4 farklı elemandan, 2'şerli 6 farklı grup oluşturulabilir. Grupları yazalım.

$$\left.\begin{array}{l} a,b \\ a,c \\ a,d \\ b,c \\ b,d \\ c,d \end{array}\right\} 6$$

Görüldüğü gibi sıra dikkate alınmadan 6 farklı grup düzenlenebilir.

Tekrarlı Kombinasyon

k sayıda elemandan oluşacak grupların düzenleneceği n sayıdaki elemanın bazıları birbirinin aynı ise, permütasyonda olduğu gibi sonuç değişecektir. Bu durumda, tekrarlı kombinasyon işlemine başvurulacaktır. Tekrarlı kombinasyonda grupların sayısı,

$$tC_n^k = \frac{(n+k-1)!}{k!(n-1)!}$$

olacaktır.

ÖRNEK: a, b, c elemanlarından, 2'şer elemandan tekrarlı olarak düzenlenecek grup sayısı hesaplayınız. (Gruplarda elemanların sırası önemli değildir.)

$$tC_n^k = \frac{(n+k-1)!}{k!(n-1)!}$$

$$tC_3^2 = \frac{(3+2-1)!}{2!(3-1)!} = \frac{4!}{2!2!} = \frac{1.2.3.4}{1.2.1.2} = 6$$

farklı grup düzenlenecektir.

Grupları düzenlersek,

$$\left.\begin{array}{ll} a & a \\ a & b \\ a & c \\ b & b \\ b & c \\ c & c \end{array}\right\} \quad 6$$

olur.

6.2. OLASILIĞIN TANIMI VE TEMEL ÖZELLİKLERİ

Farklı yaklaşımlarla farklı olasılık tanımları yapılabilmektedir. Burada sadece objektif olasılıktan sözedilecektir.

Tanım için A olarak adlandıracağımız bir olay olduğunu varsayalım. A olayı ile ilgili örnek uzayında N sayıda sonuç (eleman) varsa ve bunlardan n sayıda sonuç A olayını gerçekleştirecek ise A olayının gerçekleşmesi olasılığı,

$$P(A) = \frac{n}{N}$$

dir.

A olayının gerçekleşmesi olasılığı, olayı gerçekleştirecek sonuçların (n), meydana gelebilecek tüm sonuçlara (N) oranlanması ile hesaplanır.

Örneğin, A olayı 6 yüzlü düzgün bir zarın atılışında 5 noktalı yüzün gelmesi ile gerçekleşecekse, zar atıldığında 1'den 6 noktalıya kadar tüm yüzler gelebileceğinden, örnek uzayı toplam 6 sonuçtan oluşacaktır (N=6). A olayı 5 noktalı yüz gelirse gerçekleşeceğinden, örnek uzayında A olayını gerçekleştirecek sonuç bir tanedir (n=1). Bu durumda olayın gerçekleşmesi olasılığı,

$$P(A) = \frac{n}{N} = \frac{1}{6} = 0.1666$$

dır.

A olayını gerçekleştirecek veya gerçekleştirmeyecek tüm sonuçların oluşturacağı örnek uzayını S ile gösterirsek deneme ile oluşacak sonuç mutlaka S kümesinin elemanı olacaktır. Bu nedenle

$$P(S) = 1$$

olacaktır.

A olayını gerçekleştirecek sonuçların oluşturacağı kümeyi A kümesi olarak tanımlarsak, A kümesinin eleman sayısı (n); örnek uzayı S'in eleman sayısı N'le 0 arasında olabilir. Diğer bir deyişle, A alt kümesi S kümesinin tam alt kümesi veya boş küme olabilir.

$$0 \leq n \leq N$$

tarafları N'le bölersek,

$$\underset{\downarrow}{\frac{0}{N}} \leq \underset{\downarrow}{\frac{n}{N}} \leq \underset{\downarrow}{\frac{N}{N}}$$

$$0 \qquad P \qquad 1$$

sonuç olarak

$$0 \leq P \leq 1$$

bulunur. Buna göre olasılık değeri sıfırdan küçük, birden büyük olamaz.

P=0 ise, olayın gerçekleşmesi mümkün değildir, olay gerçekleşmeyecektir. P=1 ise olay mutlaka gerçekleşecektir. Her iki durumda da kesinlik vardır.

P'nin sıfıra yaklaşması, olasılığın azaldığını, 1'e yaklaşması ise arttığı anlamını taşımaktadır.

Teorem: $\varnothing$ boş küme ise, $P(\varnothing)=0$ dır.

İspat: Herhangi bir A olayı için, kümelerin özelliklerinden yararlanarak,

$$A = A \cup \varnothing$$

yazabiliriz. Bu durumda,

$$P(A) = P(A \cup \varnothing)$$

ve

$$P(A) = P(A) + P(\varnothing)$$

yazılabilir. Taraflardan P(A) çıkartılırsa

$$P(\varnothing) = 0$$

bulunur.

Teorem: $\overline{A}$ kümesi, A kümesinin tamamlayıcısı ise,

$P(A) = 1 - P(\overline{A})$ dır.

İspat: Örnek uzayı S ise, $S = A \cup \overline{A}$ yazabiliriz.

Bu durumda,

$$P(S) = P(A) + P(\overline{A})$$

$$1 = P(A) + P(\overline{A})$$

olacaktır. Böylece,

$P(A) = 1 - P(\overline{A})$ olduğu görülmektedir.

Burada şu noktayı açıklamak gerekir. A olayının gerçekleşmesi olasılığı P(A) ise, $P(\overline{A})$ A olayının gerçekleşmemesi olasılığıdır. A olayını gerçekleştirecek eleman sayısı (n) ise, gerçekleştirmeyecek eleman sayısı (N-n) olacak ve A olayının gerçekleşmemesi olasılığı,

$$P(\overline{A}) = \frac{N - n}{N}$$

olacaktır. Bu eşitliği şöyle yazabiliriz.

$$P(\overline{A}) = \underset{\downarrow}{\frac{N}{N}} - \underset{\downarrow}{\frac{n}{N}}$$
$$1 \qquad P(A)$$

Buradan,

$$P(\overline{A}) = 1 - P(A)$$
$$P(A) = 1 - P(\overline{A})$$
$$P(A) + P(\overline{A}) = 1$$

sonuçları daha önce olduğu gibi bulunur.

Daha sonra yapılacak açıklamalarda bir olayın gerçekleşmesi olasılığını p ve gerçekleşmemesi olasılığını q harfleri ile ifade ederek p+q=1 bağıntısından yararlanacağız.

6.3. BİRDEN FAZLA OLAYIN OLASILIĞI

Buraya kadar yaptığımız açıklamalar A gibi bir tek olaya aitti. Bu bölümde birden fazla olayla ilgili olasılıklardan sözedeceğiz. Konuyla ilgili teoremleri vermeden önce bazı tanımları vermemiz, konunun daha kolay anlaşılmasına yardımcı olacaktır.

6.3.1. Bağdaşan Olaylar

İki veya daha fazla olay birlikte gerçekleşebiliyorsa bağdaşan, gerçekleşmiyorlarsa bağdaşmayan olaylar olarak tanımlanırlar.

A ve B gibi iki olay düşünelim. A olayı 6 yüzlü bir zar atıldığında çift noktalı yüzlerin gelmesi ve B olayı 2 noktalı yüz veya daha az noktalı yüzlerin gelmesi ile gerçekleşecekse, A ve B olayları bağdaşan olaylardır. Çünkü, A olayının gerçekleşmesi için zarın 2,4,6 ve B olayının gerçekleşmesi için 1 ve 2 noktalı yüzlerinin gelmesi gerekmektedir. Zar atıldığında 2 noktalı yüz gelirse, 2 noktalı yüz hem A hem de B olaylarında ortak olduğundan iki olayı birlikte gerçekleştirecektir. Bu nedenle A ve B olayları bağdaşan olaylardır.

A olayı aynı kalmak şartı ile B olayının zarın 3 ve 5 noktalı yüzlerinin gelmesi ile gerçekleşeceğini düşünelim. Bu durumda zarın hangi yüzü gelirse gelsin A ve B olayları birlikte gerçekleşmeyecektir. Bu durumda A ve B bağdaşmayan olaylardır.

A ve B aynı örnek uzayının alt kümeleri ise, A ve B kümelerinin kesişimleri varsa bağdaşan, kesişimleri yoksa bağdaşmayan olaylardır.

6.3.2. Bağlı Olaylar

Bağlı olaylar, kümelerin kesişimleri ile ilgili olarak tanımlandığından, sadece bağdaşan olaylar için söz konusudur. Kısaca bir olayın gerçekleşip gerçekleşmemesi diğer olayı ve olayları etkiliyorsa olaylar bağlı olaylar, etkilemiyorsa olaylar bağlı olmayan olaylar olarak tanımlanır.

Bir torbada 3 kırmızı ve 5 mavi top olduğunu varsayalım. Torbadan ard arda iki top çekilmesini ve topların kırmızı olması olasılığını araştıralım. Bu durumda iadeli (yerine koyarak) ve iadesiz (yerine konmadan) çekiliş yapılabilir. Birinci çekiliş A, ikinci çekiliş B olayı ise, iadeli çekilişlerle topun kırmızı olması olasılığı,

$$P(A) = \frac{n}{N} = \frac{3}{8}$$

dır. Çekilen top torbaya iade edileceğinden 2. çekilişte kırmızı top çekilmesi olasılığı

$$P(B) = \frac{n}{N} = \frac{3}{8}$$

dır.

A olayı B olayını etkilemediğinden, A ve B olayları bağlı olamayan olaylardır. Çekilişlerin iadesiz yapıldığını düşünelim. 1. topun kırmızı olması olasılığı,

$$P(A) = \frac{n}{N} = \frac{3}{8}$$

dır. Çekilen top torbaya iade edilemeyeceğinden P(B) olasılığı değişecektir. 1.topun kırmızı olduğunu kabul edersek 2. topun kırmızı olması olasılığı,

$$P(B) = \frac{n-1}{N-1} = \frac{2}{7}$$

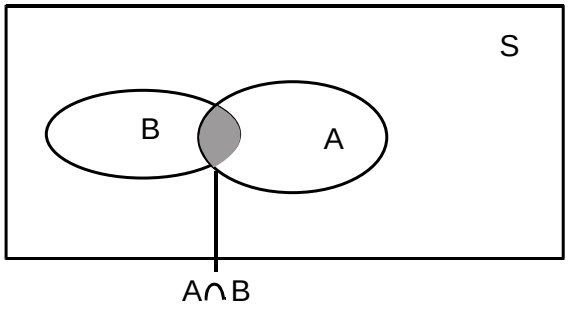
olması için ilk topun kırmızı olduğunu kabul ediyoruz. İlk top mavi ise ikinci topun kırmızı olması olasılığı 3/ 7 olacaktır. Bu durumda A olayı B olayını etkilediğinden, A ve B olayları bağlı olaylardır.

Teorem: A ve B herhangi iki olayı ise,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

dır.

Teoremi ispat etmeden gerekli açıklamaları yapalım.

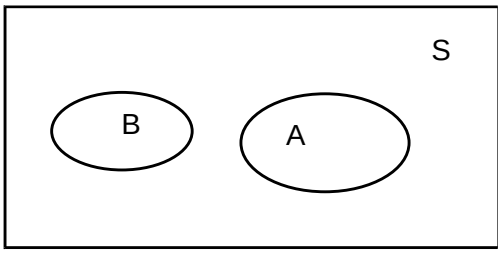


Şekil 6.1: Bağdaşan Olaylar

A ve B olayları bağdaşan olaylarsa, $A \cap B$ kümesinin elemanları hem A hem de B kümesi ile $A \cup B$ kümesinde yer almaktadır. Bu nedenle P(A) ve P(B) olasılıklarının toplamından $P(A \cap B)$ olasılığı bir defa çıkarılmaktadır. Bağdaşan olaylar için,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

dir.



Şekil 6.2: Bağdaşmayan Olaylar

Bağdaşmayan olaylar için A ve B kümelerinin kesişimi yoktur.

$A \cap B = \varnothing$ olduğundan,

$P(A \cap B) = 0$'dır.

Bu nedenle bağdaşmayan olaylar için,

$$P(A \cup B) = P(A) + P(B)$$

dir.

Teorem: A, B ve C herhangi üç olay ise

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

dir. Teoremi ispatlamadan şunları söyleyebiliriz.

A, B ve C olayları bağdaşan olaylar ise yukarıdaki formül geçerli olacaktır. Olaylar bağdaşmayan olaylarsa,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

dir.

Teorem: Eğer $A \subset B$ ise,

$$P(A) \leq P(B)$$

dir.

ÖRNEK: Bir torbada birbirinin benzeri 1'den 10'a kadar numaralanmış 10 beyaz, 1'den 7'ye kadar numaralanmış 7 siyah ve 1'den 5'e kadar numaralanmış 5 mavi top vardır. Torbadan bir top çekilecektir. Çekilişle ilgili,

A olayı; çekilen topun beyaz olması,
B olayı; çekilen topun 4 numaralı olması,
C olayı; çekilen topun herhangi bir tek sayılı olması,
D olayı; çekilen topun 5 ve daha büyük sayılı siyah olması

olarak tanımlandığına göre,

1- A, B, C ve D olaylarının gerçekleşme olasılıklarını hesaplayınız.
2- a) $P(A \cap B)$;d) $P(B \cap C)$
b) $P(A \cap C)$;e) $P(B \cap D)$
c) $P(A \cap D)$;f) $P(C \cap D)$

olasılıklarını hesaplayınız.

3- a) $P(A \cup B)$; d) $P(B \cup C)$
b) $P(A \cup C)$; e) $P(B \cup D)$
c) $P(A \cup D)$; f) $P(C \cup D)$

olasılıklarını hesaplayınız.

1- 10 Beyaz
7 Siyah
5 Mavi
_____
22 Toplam

A olayında N=22 ve beyaz top sayısı 10 olduğundan n=10'dur.

$$P(A) = \frac{n}{N} = \frac{10}{22} = 0.4545$$

B olayında N=22; Beyaz, Siyah ve Mavi 4 numaralı top bulunduğundan n=3'tür.

$$P(B) = \frac{n}{N} = \frac{3}{22} = 0.1363$$

C olayında N=22; Tek sayılar: 1,3,5,7,9 numaralı beyazlar, 1,3,5,7, numaralı siyahlar ve 1,3,5 numaralı maviler olduğundan, n= 5+4+3=12'dir.

$$P(C) = \frac{n}{N} = \frac{12}{22} = 0.5454$$

D olayında N=22; olayı gerçekleştirecek toplar 5,6 ve 7 nolu siyahlar olduğundan n=3'tür.

$$P(D) = \frac{n}{N} = \frac{3}{22} = 0.1363$$

2-

a) A ve B olayları bağdaşan olaylardır. Çünkü 4 numaralı beyaz top çekilirse hem A hem B olayları gerçekleşir. (N=22 ; n=1)

$$P(A \cap B) = \frac{1}{22} = 0.0454$$

b) A ve C olayları bağdaşan olaylardır. Çekilen tek sayılı beyaz toplar A ve C olaylarını birlikte gerçekleştirir. (N=22, Tek sayılı beyaz top sayısı n=5)

$$P(A \cap C) = \frac{5}{22} = 0.2272$$

c) A ve D olayları bağdaşmayan olaylardır. Hangi top çekilirse çekilsin iki olay birlikte gerçekleşmez.

$P(A \cap D) = 0$

d) B ve C olayları bağdaşmayan olaylardır.

$P(B \cap C) = 0$

e) B ve D olayları bağdaşmayan olaylardır.

$P(B \cap D) = 0$

f) C ve D olayları bağdaşan olaylardır. 5 ve 7 numaralı siyah toplar çekilirse her iki olayda gerçekleşecektir. (N=22, n=2)

$$P(C \cap D) = \frac{2}{22} = 0.0909$$

3-

a) A ve B olayları bağdaşan olaylar olduğundan,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A \cup B) = \frac{10}{22} + \frac{3}{22} - \frac{1}{22} = \frac{12}{22} = 0.5454$$

b) A ve C olayları bağdaşan olaylar olduğundan,
$P(A \cup C) = P(A) + P(C) - P(A \cap C)$

$$P(A \cup C) = \frac{10}{22} + \frac{12}{22} - \frac{5}{22} = \frac{17}{22} = 0.7727$$

c) A ve D olayları bağdaşmayan olaylar olduğundan,
$P(A \cup D) = P(A) + P(D)$

$$P(A \cup D) = \frac{10}{22} + \frac{3}{22} = \frac{13}{22} = 0.5909$$

d) B ve C olayları bağdaşmayan olaylar olduğundan,
$P(B \cup C) = P(B) + P(C)$

$$P(B \cup C) = \frac{3}{22} + \frac{12}{22} = \frac{15}{22} = 0.6818$$

e) B ve D olayları bağdaşmayan olaylar olduğundan,
$P(B \cup D) = P(B) + P(D)$

$$P(B \cup D) = \frac{3}{22} + \frac{3}{22} = \frac{6}{22} = 0.2727$$

f) C ve D olayları bağdaşan olaylar olduğundan,
$P(C \cup D) = P(C) + P(D) - P(C \cap D)$

$$P(C \cup D) = \frac{12}{22} + \frac{3}{22} - \frac{2}{22} = \frac{13}{22} = 0.5909$$

6.3.3. Şartlı Olasılık

Bağdaşan olaylar için $P(A \cup B)$ olasılığı,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

olarak açıklanmıştı.

$P(A \cap B)$ olasılığının hesaplanması olayların bağlı olup olmamalarına göre değişecektir. A ve B olayları bağlı değil ise,
$P(A \cap B) = P(A) \cdot P(B)$

dir. A ve B olayları bağlı olaylar ise,

$P(A \cap B) = P(A) \cdot P(B/A)$

dır. Burada P(B/A), A olayı gerçekleştiği varsayımı ile B olayının gerçekleşmesi olasılığıdır ve şartlı olasılık olarak tanımlanır. Çünkü, A olayının gerçekleşip gerçekleşmemesi B olayını etkileyecektir. Buradan P(B/A)'yı çekersek,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

olarak şartlı olasılık formülü elde edilir. Aynı olasılığı, A ve B olayları için,

$P(A \cap B) = P(B) \cdot P(A/B)$

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

şeklinde yazarsak P(A/B), B olayının gerçekleştiği varsayımı ile A olayının gerçekleşmesi olasılığıdır.

Bağlı ve bağlı olmayan olaylarla ilgili formüller ikiden fazla olay için (örneğin, A,B,C gibi 3 olay için) şöyle yazılabilir.

A, B, C bağlı olmayan olaylar ise,

$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

Bağlı olaylar ise,

$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/AB)$

olacaktır. Burada P(C/AB); A ve B olaylarının gerçekleştiği bilindiğine göre C olayının gerçekleşmesi olasılığıdır.

ÖRNEK: İki olayla ilgili P(A) =0.2 ve $P(A \cap B)$ =0.07'dir. A ve B bağlı olaylardır. A olayının gerçekleştiği kabul edildiğine göre B olayının gerçekleşmesi olasılığını hesaplayınız.

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.07}{0.2} = 0.35$$

6.4. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Hilesiz bir paranın 5 kez atılması olayı için,

a-İlk iki atışta yazı, diğer üç atışta tura gelmesi,
b-Beş atışta iki yazı, üç tura gelmesi olasılıklarını hesaplayınız.

Çözüm:

a- $P(Y) = \frac{1}{2}$, $P(T) = \frac{1}{2}$

P = P(Y).P(Y).P(T).P(T).P(T)

$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{32}$

b- Sıra sayısı = $\frac{5!}{2!3!} = 10$

c-

P = [P(Y).P(Y).P(T).P(T).P(T)] .10

$= \frac{10}{32}$

ÖRNEK 2: Bir işletmede çalışan 7 erkek, 3 kadın personelden 3 kişilik komisyon tesadüfi olarak seçilecektir.

a-Komisyona 2 kadın, 1 erkek personel seçilmesi,

b-Seçilen ilk personelin kadın diğerlerinin erkek olması olasılıklarını hesaplayınız.

Çözüm:

a- $P(K) = \frac{3}{10}$, $P(E) = \frac{7}{10}$

Sıra sayısı = $\frac{3!}{2!.1!} = 3$

P = [P(K).P(K).P(E)] .3

$= \left(\frac{3}{10} \cdot \frac{2}{9} \cdot \frac{7}{8}\right).3 = \frac{126}{720}$

b- P = P(K).P(K).P(E)

$= \frac{3}{10} \cdot \frac{7}{9} \cdot \frac{6}{8} = \frac{126}{720}$

ÖRNEK 3: Kanser hastalığının var olup olmadığını ortaya koyacak bir test geliştirilmiştir. Bu test 0,05 olasılıkla yanlış sonuç vermektedir. İçlerinde 0,005 oranında

kanserli hasta bulunan çok sayıda kişiye bu test uygulanmıştır. Test sonucu kanser hastası olarak belirlenen bir kişinin gerçekte kanserli olması olasılığı nedir?

Çözüm:

K : Bir hastanın kanserli olması

T : Testin pozitif sonuç vermesi

$P(K) = 0{,}005$

$P(T / \overline{K}) = 0{,}05$

$$P(K/T) = \frac{P(T/K)P(K)}{P(T/K)P(K) + P(T/\overline{K})P(\overline{K})}$$

$$= \frac{(0{,}95)(0{,}005)}{(0{,}95)(0{,}005) + (0{,}05)(0{,}995)}$$

$$= 0{,}08716$$

ÖRNEK 4: 50 öğrenciden 15'i fizik, 20'si kimya dersinden başarısızdır. Fizik ve kimya derslerinden başarısız olma olayları bağımsız olaylar mıdır?

$P(F) = 0{,}30 \quad , \quad P(K) = 0{,}40$

$P(F \cap K) = 0{,}50$

$P(F \cap K) \neq P(F).P(K)$

$0{,}50 \neq 0{,}12$

Bağımsız değillerdir.

ÖRNEK 5: Hilesiz bir zarın 3 kere atılması deneyi gerçekleştirilmiştir.

a- Bu deneyde 5'li yüzeyin üste gelmesi sonucunun beklenen frekansı nedir?
b- Bu deneyde kaç çeşit mümkün sonuç vardır?
c- Bu deneyde iki çift bir tek noktalı yüz kaç şekilde elde edilir?

Çözüm:

a- $P(5) = \frac{1}{6} = P(X)$

$$E(X) = n.p = 3.\frac{1}{6} = \frac{1}{2}$$

b- n: deneme sayısı
k : bir denemede mümkün olay sayısı

$$N = k^n = 6^3 = 216$$

c- Çift noktalı yüz 2,4,6
Tek noktalı yüz 1,3,5

$$N = \binom{3}{2}\binom{3}{1} = \frac{3!}{2!1!} = 9$$

ÖRNEK 6: 100 öğrencili bir dershanede arabası olan 25 öğrenci ile yapılan araştırmada 10'unun Avrupa araba sahibi olduğu görülmüştür. Aşağıdakilerden hangisi doğrudur? Nüfusun .....

a-%25'i Yerli,
b-%10'u Japon,
c-%40'ı Avrupa,
d-%10'u Avrupa,
e-%60'ı Yerli arabaya sahiptir.

Çözüm:

P(Avrupa) = $\frac{n}{N} = \frac{10}{25} = 0{,}4$

ÖRNEK 7: Bir torbada 3 kırmızı, 3 beyaz ve 4 mavi top vardır. Seçimler tesadüfi yapılmaktadır.

a-Seçilen bir topun kırmızı veya beyaz olma olasılığı nedir?
b-İadesiz seçilen birinci topun kırmızı, ikincisinin beyaz olma olasılığı nedir?
c-İadeli seçilen iki toptan birinin kırmızı diğerinin mavi olma olasılığı nedir?

Çözüm:

a- $P(K) = \frac{3}{10}$, $P(B) = \frac{3}{10}$

$$P(K \cup B) = P(K) + P(B) = \frac{3}{10} + \frac{3}{10} = \frac{6}{10}$$

b- $$P(X_1 = K \cap X_2 = B) = P(K).P(B/K) = \frac{3}{10}.\frac{3}{9} = \frac{9}{90} = \frac{1}{10}$$

c-

$$P(K \cap M) = P\left[\left(X_1 = K \cap X_2 = M\right) \cup \left(X_1 = M \cap X_2 = K\right)\right]$$

$$= \binom{2}{1} P(K)P(M) = 2.\frac{3}{10}.\frac{4}{10} = \frac{24}{100}$$

ÖRNEK 8: Bir torbada eşit ve çok sayıda sarı ve siyah toplar vardır. 3 top tesadüfi olarak seçilmiştir.

a-Kaç farklı durumla karşılaşmamız mümkün olur?
b-Hepsinin sarı gelmesinin bileşim sayısı nedir?
c-Hepsinin sarı gelmesinin olasılığı nedir?
d-İkisinin siyah olması olasılığı nedir?

Çözüm:

a- $N = k^n = 2^3 = 8$

b- $n_u = C_3^3 = \binom{3}{3} = 1$

c- $p = \frac{n_u}{N} = \frac{1}{8} = 0{,}125$

d- $n_u = C_3^2 = \binom{3}{2} = \frac{3!}{2!1!} = 3$

$p = \frac{n_u}{N} = \frac{3}{8}$

BÖLÜM 7

TESADÜFİ DEĞİŞKENLER

Tesadüfi değişkenler istatistik karar problemlerinin temeli olan olasılık teorisini tamamlayıcı niteliktedir. Daha önce klasik formüllerle hesaplanan olasılıkların, tesadüfi değişken adı verilen değişkenler ile nasıl hesaplandığı bu bölümde açıklanmaya çalışılacaktır. Konunun boyutları burada geniş olarak incelenmesini engellemektedir. Bu nedenle burada konu mümkün olduğunca sınırlandırarak, sadece temel kavramlara değinilecektir.

Bir olayla ilgili bir tek karakter sözkonusu olduğunda tek-boyutlu, karakter sayısı arttığında ise çok-boyutlu tesadüfi değişkenler sözkonusu olacaktır. Örneğin, bir makine üretiminin ağırlık karakteri aranıyorsa bu karakteri bir tek tesadüfi değişken ile ifade ederken, aynı olayla ilgili uzunluk karakteri de aranacaksa, iki farklı tesadüfi değişken ile bu iki farklı karakter ifade edilecektir. Aynı olayla ilgili çok sayıda tesadüfi değişkenden de sözedebiliriz. Buna bağlı olarak tesadüfi değişkenleri tek-boyutlu ve çok-boyutlu tesadüfi değişkenlere yerverilecektir.

7.1. TESADÜFİ DEĞİŞKEN KAVRAMI

Kesin sonucu önceden belli olmayan bir deneme (E) ile ilgili mümkün tüm sonuçların oluşturduğu bir küme düşünelim. Mümkün sonuçları X ile ifade edersek, X bir tesadüfi değişkendir. Bu değişen kümesinde $X_1,X_2,X_3,.....,X_n$ değerleri, f $(X_1),f(X_2),f(X_3),.....,f(Xn)$ olasılıkları ile alacaktır. Burada X bir tesadüfi değişken ise olasılıklar toplamının 1 olması gerekmektedir. Aksi taktirde tesadüfi değişkenden sözedilmesi mümkün değildir.

X tesadüfi değişkeninin alacağı sayısal değerlerin alt ve üst sınırları a ve b gibi iki sayısal değer olarak ifade edilebilir. X tesadüfi değişkeninin tanımlanan bu aralık

içinde alacağı değerlerin olasılıkların toplamı 1'e eşit olacaktır. X tesadüfi değişkenin a-b aralığında alacağı değerler a ve b ye eşit olabileceği gibi, bunlar arasında bir değerde olabilir. Tesadüfi değişkenin alabileceği değerlerin özelliklerine dayanarak, süreksiz ve sürekli tesadüfi değişkenler ayrımı yapabiliriz.

7.2. SÜREKSİZ TESADÜFİ DEĞİŞKENLER

X tesadüfi değişkeni a-b aralığında belirli (belirli sayıda) değerler alıyorsa "süreksiz tesadüfi değişken"den sözedilir. Örneğin, X tesadüfi değişkeni 0-10 aralığında sadece tam sayı değerleri alıyorsa süreksiz tesadüfi değişkendir.

X; $X_1,X_2,.....,X_n$ değerlerini, $f(X_1)$, $f(X_2),.....,f(X_n),.....$ veya $P_1,P_2,.....,P_n$ olasılıkları ile olan süreksiz bir tesadüfi değişken ise ve

a- $f(X_i) \geq 0$; i=1,2,.....n için

b- $\sum_{i=1}^{n} f(X_i) = 1$

şartlarını sağlıyorsa, olasılıkların $f(X_i)$ dağılımı "olasılık fonksiyonu" olarak adlandırılır.

İki paranın birlikte atıldığını düşünelim. Bu durumda X tesadüfi değişkeni tura sayısını gösteriyorsa mümkün olan haller ve X tesadüfi değişkenlerinin alacağı değerler şöyle olacaktır. (Y:yazı, T:tura)

1.para	2.para	X_i (Tura Sayısı)
Y	Y	0
T	Y	1
Y	T	1
T	T	2

P(T) =1/2 ve $P(Y) = P(\overline{T}) = 1/2$ olduğuna göre, yukarıda belirlenen durumların olasılıklarını hesaplayalım.

$P(Y \cap Y) = P(Y) \cdot P(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(T \cap Y) = P(T) \cdot P(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

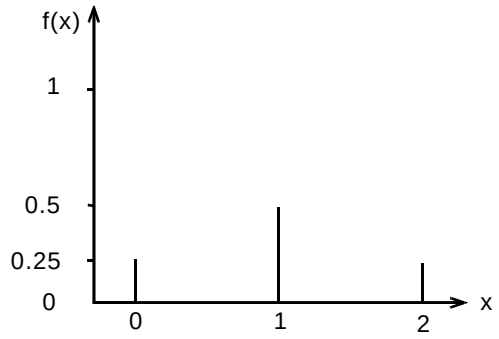
$P(Y \cap T) = P(Y) \cdot P(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(T \cap T) = P(T) \cdot P(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Sonuçları şöyle gösterebiliriz.

X_i	f(X)
0	¼
1	2/4
2	¼
	4/4 = 1

f(X) dağılımı X tesadüfi değişkeninin olasılık fonksiyonudur. Bu dağılım grafikle şöyle gösterebiliriz.



Şekil 7.1: Olasılık Fonksiyonu

X tesadüfi değişkeninin birikimli olasılık fonksiyonu

$X_1 < X_2 < ..... < X_{n-1} < X_n$ için

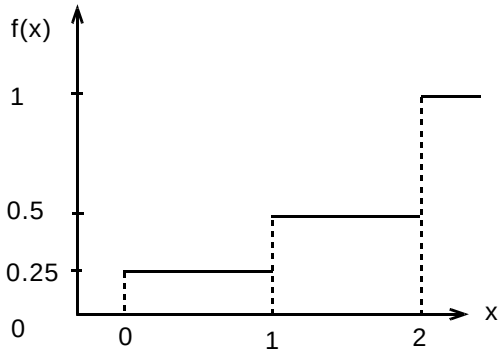
$$F(X_k) = f(X_1) + f(X_2) + ... + f(X_k) = \sum_{i=1}^{k} f(X_i) \quad ; \quad X_k \le X_n$$

şeklinde belirlenerek, artan ve azalan birikimli olasılık fonksiyonlarının bulunması sözkonusu olacaktır.

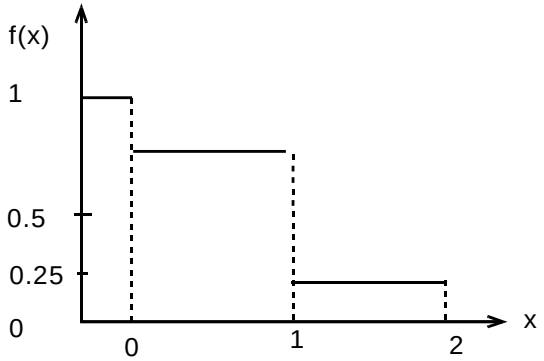
Yukarıda verilen örneğe dönelim.

X_i	f(X)	$F(X_i)$ (Artan)	$F'(X_i)$ (Azalan)
0	0.25	0.25	1.00
1	0.50	0.75	0.75
2	0.25	1.00	0.25
	1.00		

Birikimli olasılık fonksiyonları şekil 7.2 ve şekil 7.3'te görülmektedir.



Şekil 7.2: Artan Birikimli Olasılık Fonksiyonu



Şekil 7.3: Azalan Birikimli Olasılık Fonksiyonu

Hatırlanacağı gibi, istatistikte serileri oluşturan birimlerin değerlerinin büyüklüğü hakkında bilgi edinmek, seriyi bir tek değerle ifade edebilmek ve farklı serileri karşılaştırabilmek amacı ile ortalamalar hesaplanıyordu ve en yaygın kullanılan ortalama aritmetik ortalamaydı. Frekans serilerinde aritmetik ortalama; seri birimlerinin değerlerinin, bu değerlere ait frekanslar ile çarpımlarının toplamlarının frekanslar toplamına bölünmesi ile hesaplanıyordu. Bir olasılık fonksiyonu sözkonusu olduğunda, olasılık değerleri oransal frekanslar olarak düşünülürse, aritmetik ortalama benzer şekilde hesaplanabilir. X tesadüfi değişkeninin olasılıklarla hesaplanan aritmetik ortalaması "beklenen değer" olarak adlandırılır.

Süreksiz bir olasılık fonksiyonu için olasılık değerleri P_i ile gösteriliyorsa [$f(X_i)=p_i$], X tesadüfi değişkeninin beklenen değeri,

$$E(X) = \sum_{i=1}^{n} p_i X_i$$

olarak hesaplanacaktır. Bu açıklamalardan da anlaşılacağı gibi beklenen değer, tesadüfi değişkenin matematiksel olarak alması beklenen değerdir.

Süreksiz X tesadüfi değişkeninin varyansı ise,

$$V(X) = \sum_{i=1}^{n} [X_i - E(X)]^2 . P_i$$

olacaktır. Bu formül açılırsa,

$$V(X) = \sum_{i=1}^{n} p_i X_i^2 - [E(X)]^2$$

elde edilir.

Süreksiz X tesadüfi değişkeninin momentleri şöyle hesaplanır.

1- Orjine Göre Momentler:

$$m_1 = \sum_{i=1}^{n} p_i X_i$$

$$m_2 = \sum_{i=1}^{n} p_i X_i^2$$

$$m_3 = \sum_{i=1}^{n} p_i X_i^3$$

$$m_4 = \sum_{i=1}^{n} p_i X_i^4$$

2- Ortalamaya Göre Momentler:

$$\mu_1 = \sum_{i=1}^{n} p_i [X_i - E(X)] = 0$$

$$\mu_2 = \sum_{i=1}^{n} p_i [X_i - E(X)]^2 = \sum_{i=1}^{n} p_i X_i^2 - [E(X)]^2 = V(X)$$

$$\mu_3 = \sum_{i=1}^{n} p_i [X_i - E(X)]^3 = 0$$

$$\mu_4 = \sum_{i=1}^{n} p_i [X_i - E(X)]^4 = 0$$

3- Orjine göre momentler belli ise, ortalamaya göre momentler,

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

şeklinde de hesaplanabilir.

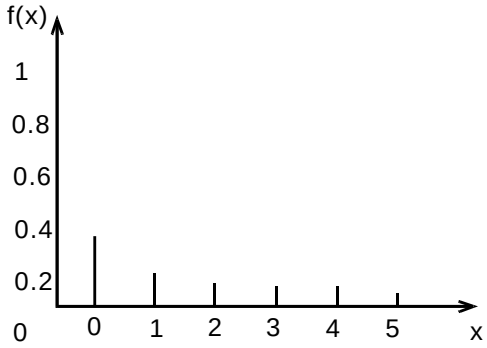
ÖRNEK: Bir atölyede A mamülleri üretilmektedir. Atölyenin bir günlük üretimi olan 500 mamül üzerindeki hata sayıları incelenerek, sonuçlar aşağıda verilmiştir. (X tesadüfi değişkeni mamülleri üzerindeki hata sayısını göstermektedir.)

X_i	f_i	P_i	$F(X)$	$F'(X)$	P_iX_i	$P_iX_i^2$
0	180	180/500=0.36	0.36	1.00	0	0
1	110	110/500=0.22	0.58	0.64	0.22	0.22
2	70	70/500=0.14	0.72	0.42	0.28	0.56
3	60	60/500=0.12	0.84	0.28	0.36	1.08
4	50	50/500=0.10	0.94	0.16	0.40	1.60
5	30	30/500=0.06	1.00	0.06	0.30	1.50
		1.00			1.56	4.96

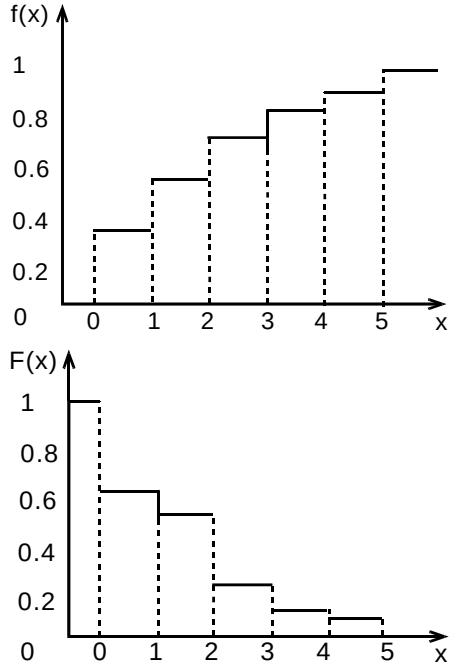
X tesadüfi değişkeninin,

1- Olasılık fonksiyonunun grafiğini çiziniz.
2- Birikimli olasılık fonksiyonunun grafiğini çiziniz.
3- Beklenen değerini hesaplayınız.
4- Varyansını hesaplayınız.
5- Aşağıdaki olasılıkları hesaplayınız.
   a- $P(X \leq 2)$
   b- $P(X>3)$
   c- $P(1< X \leq 4)$

1-



2-



3- $E(X) = \sum_{i=1}^{6} p_i X_i = 1.56$

4- $V(X) = \sum_{i=1}^{6} p_i (X_i - \overline{X})^2 = \sum_{i=1}^{6} p_i X_i^2 - [E(X)]^2 = 4.96 - (1.56)^2$

$= 4.96 - 2.4336 = 2.5264 \rightarrow \sigma = \sqrt{2{,}5264} = 1.59$

5-a) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$
$= 0.36 + 0.22 + 0.14$
$= 0.72$

veya

$P(X \leq 2) = 1 - P(X>2) = 1 - [P(X=3) + P(X=4) + P(X=5)]$
$= 1 - (0.12 + 0.10 + 0.06)$
$= 1 - 0.28$
$= 0.72 \rightarrow$ (F(X)'te X=2 karşılığı)

b) $P(X>3) = P(X=4) + P(X=5)$
$= 0.10 + 0.06$
$= 0.16 \rightarrow$ (F'(X)'te X=4 karşılığı)

c) $P(1< X \leq 4) = P(X=2) + P(X=3) + P(X=4)$
$= 0.14 + 0.12 + 0.10$
$= 0.36 \rightarrow$ (F(X)'te X=4 ile X=1 farkı)

7.3. SÜREKLİ TESADÜFİ DEĞİŞKENLER

X tesadüfi değişkeni a-b aralığında sonsuz sayıda değer alırsa "sürekli tesadüfi değişken"den sözedilir. Uygulama örneği olarak fiyat, uzunluk, ağırlık gibi karakterler gösterilebilir. Örneğin, bir makinenin ürettiği mamüllerin boy uzunlukları 0-2 milimetre arasında farklılık gösteriyorsa, hassas olarak belirlenen farkların sonsuz sayıda olabileceği kabul edilebilir.

Sürekli tesadüfi değişkenin a-b aralığında olasılık dağılımı matematiksel bir fonksiyonla ifade edileceğinden, toplam yerine integral işlemine başvurulması gerekecektir.

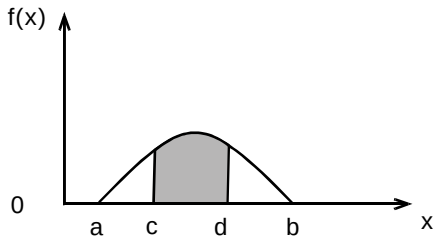
Süreksiz tesadüfi değişkenlerde açıklanan olasılık fonksiyonu ile aynı anlamda, sürekli tesadüfi değişkenler için "olasılık yoğunluk fonksiyonu" sözkonusu olacaktır. Olasılık yoğunluk fonksiyonu,

a) Tüm X değerleri için $f(X) \geq 0$

b) $\int_a^b f(X)dx = 1$

şartlarını sağlar. Aksi durumda olasılık yoğunluk fonksiyonundan sözedemeyiz.

Olasılık yoğunluk fonksiyonu şekil ile şöyle gösterilebilir.



Şekil 7.4: Olasılık Yoğunluk Fonksiyonu

Burada,

$$P(a < X < b) = \int_a^b f(X)dx = 1$$

$$P(c < X < d) = \int_c^d f(X)dx$$

olacaktır. f(X) (veya Y) sembolü ile ifade ettiğimiz olasılık yoğunluk fonksiyonundan, F(X) (veya Y) ile ifade edeceğimiz "birikimli olasılık fonksiyonu"nu,

$$F(X) = \int f(X)dx$$

şeklinde elde ederiz.

Sürekli tesadüfi değişkenlerde beklenen değer,

$$E(X) = \int_a^b X.f(X)dx$$

ve varyans,

$$V(X) = \int_a^b [X - E(X)]^2 f(X)dx$$

olacaktır. Varyans formülü açılırsa,

$$V(X) = \int_a^b X^2.f(X)dx - [E(X)]^2$$

bulunur.

Sürekli X tesadüfi değişkeni için moment formülleri şöyle olacaktır.

1- Orjine Göre Momentler:

$$m_1 = \int_a^b X \ .f(X)dx$$

$$m_2 = \int_a^b X^2.f(X)dx$$

$$m_3 = \int_a^b X^3.f(X)dx$$

$$m_4 = \int_a^b X^4.f(X)dx$$

2- Ortalamaya Göre Momentler:

$$\mu_1 = \int_a^b [X - E(X)] \ f(X)dx = 0$$

$$\mu_2 = \int_a^b [X - E(X)]^2 f(X)dx = V(X)$$

$$\mu_3 = \int_a^b [X - E(X)]^3 f(X)dx$$

$$\mu_4 = \int_a^b [X - E(X)]^4 f(X)dx$$

Orjine göre momentler biliniyorsa ortalamaya göre momentler daha önce verilen formüller ile hesaplanabilir.

ÖRNEK: 0-4 aralığında sürekli X tesadüfi değişkeninin olasılık yoğunluk fonksiyonu,

f(X) = AX

olarak verilmiştir.

1- Verilen fonksiyonun olasılık yoğunluk fonksiyonu olabilmesi için A ne olmalıdır?

2- X tesadüfi değişkeninin olasılık yoğunluk ve birikimli olasılık yoğunluk fonksiyonlarının grafiklerini çiziniz.

3- X tesadüfi değişkeninin beklenen değerini hesaplayınız.

4- X tesadüfi değişkeninin varyansını hesaplayınız.

5- a) P(X>1)

b) P(X<3)

c) P(1<X<2)

olasılıklarını hesaplayınız.

$\int_a^b f(X)dx = 1$ olmalıdır.

Verilen f(X)=AX fonksiyonunun tanımlı olduğu 0-4 aralığında belirli integrali alınırsa,

$$1 = \int_a^b f(X)dx = \int_0^4 AXdx = A\int_0^4 Xdx = A\frac{X^2}{2}\Big|_0^4$$

$$= A\left[\frac{16}{2} - \frac{0}{2}\right] = A8$$

$$A = \frac{1}{8}$$

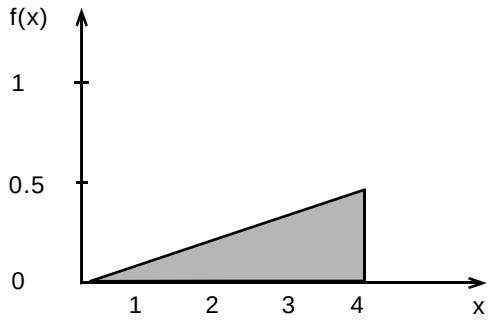
olacaktır.

2- $f(X) = \frac{X}{8}$ doğrusal olduğundan,

X=0 f(X)=0

X=4 f(X)=1/2

Olasılık yoğunluk fonksiyonunun grafiği,



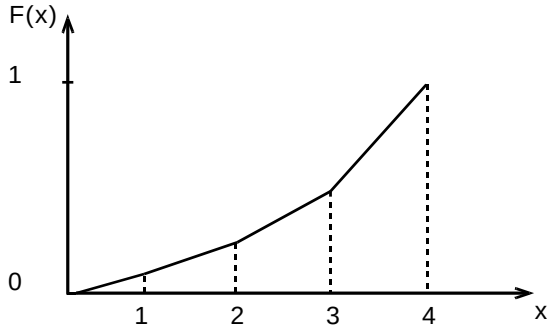
Birikimli olasılık yoğunluk fonksiyonu,

$$F(X) = \int f(X)dx = \int \frac{X}{8}dx = \frac{1}{8}.\frac{X^2}{2} = \frac{X^2}{16}$$

F(X) eğrisel fonksiyon olduğundan,

X=0	F(X)=0
X=1	F(X)=1/16
X=2	F(X)= 4/16=1/4
X=3	F(X)= 9/16
X=4	F(X)= 16/16=1

Birikimli olasılık yoğunluk fonksiyonunun grafiği,



3- $E(X)=\int_a^b X.f(X)dx$

$$=\int_0^4 X\frac{X}{8}dx=\frac{1}{8}.\frac{X^3}{3}\Big|_0^4=\frac{X^3}{24}\Big|_0^4=\frac{(4)^3}{24}-\frac{(0)^3}{24}=\frac{8}{3}$$

4-V(X)= $\int_a^b X^2.f(X)dx-[E(X)]^2$

$$=\int_0^4 X^2\frac{X}{8}dx-\left(\frac{8}{3}\right)^2=\frac{1}{8}.\frac{X^4}{4}\Big|_0^4-\frac{64}{9}=\frac{(4)^4}{32}-\frac{64}{9}=\frac{8}{9}$$

5-a) P(X>1) = $\int_1^4 f(X)dx=\frac{X^2}{16}\Big|_1^4=\frac{16}{16}-\frac{1}{16}=\frac{15}{16}$

b) P(X<3) = $\int_0^3 f(X)dx=\frac{X^2}{16}\Big|_0^3=\frac{9}{16}$

c) P(1<X<2) = $\int_1^2 f(X)dx=\frac{X^2}{16}\Big|_1^2=\frac{4}{16}-\frac{1}{16}=\frac{3}{16}$

7.4. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir sınıfta bulunan 100 öğrencinin almış oldukları notlar aşağıda görüldüğü gibidir.

Notlar	Öğrenci Sayısı
0	50
10	20
30	15
50	10
100	5
	100

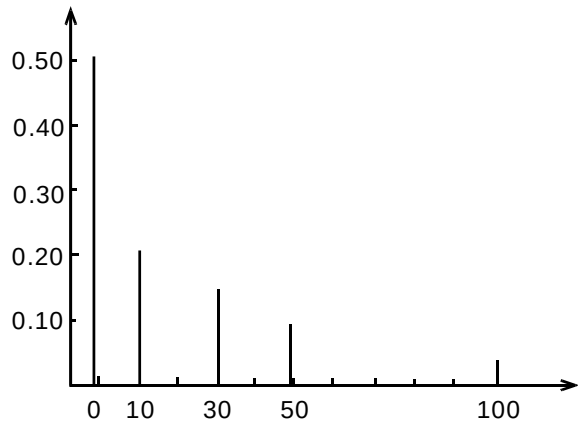
a- X tesadüfi değişkeninin olasılık fonksiyonunu oluşturunuz.
b- Olasılık fonksiyonunun grafiğini çiziniz.
c- Birikimli olasılık fonksiyonlarını oluşturunuz.
d- Beklenen değeri hesaplayınız.
e- Varyansı hesaplayınız.

Çözüm:

a-

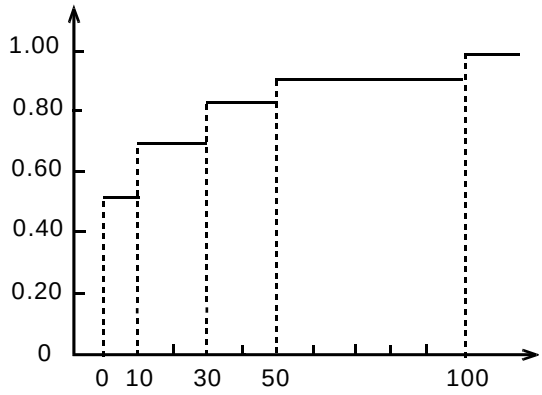
Xi	f(Xi)	
0	50/100	=0,5
10	20/100	=0,20
30	15/100	=0,15
50	10/100	=0,10
100	5/100	=0,05
		1,00

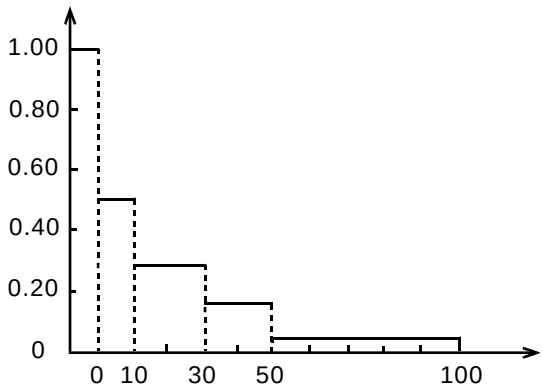
b-



c-

Xi	f(Xi)	F(X)↑	F(X)↓
0	0,50	0,50	1,00
10	0,20	0,70	0,50
30	0,15	0,85	0,30
50	0,10	0,95	0,15
100	0,05	1,00	0,05





d-

Xi	f(Xi)	Xifi(Xi)
0	0,50	0
10	0,20	2
30	0,15	4,5
50	0,10	5
100	0,05	5
		16,5

$$E(X) = \sum X_i f_i(X_i) = 16{,}5$$

e-

Xi	f(Xi)	FiXi	fiXi²
0	0,50	0	0
10	0,20	2	20
30	0,15	4,5	135
50	0,10	5	250
100	0,05	5	500
		16,5	905

$$V(X) = \sum p_i X_i^2 - \left[E(X)\right]^2$$

$$= 905 - (16{,}5)^2 = 632{,}75$$

$$\sigma = \sqrt{632{,}75} = 25{,}15$$

ÖRNEK 2: Bir işletmenin paketleme bölümünde paketleme işlemi için makineler belli bir ağırlığa göre ayarlanmıştır. Ancak bu makinelerin doldurduğu paketlerin ağırlıkları 0-3 gr. arasında sapmaktadır. Olasılık yoğunluk fonksiyonu $f(X) = A(6X - X^2)$ olarak belirlenmiştir.

a-A parametresini hesaplayınız.

b-Olasılık yoğunluk fonksiyonunun grafiğini çiziniz.

c-Birikimli olasılık yoğunluk fonksiyonunu belirleyerek grafiğini çiziniz.

d-Beklenen değeri hesaplayınız.

e-Varyansı hesaplayınız.

f-P(X< 2,4)

P(X >1,8)

P(1,3< X < 2,7) olasılıklarını hesaplayınız.

Çözüm:

a)

$$1 = \int_0^3 A(6X - X^2)dx$$

$$1 = A\left[3X^2 - \frac{X^3}{3}\right]_0^3 = A\left[27 - 9\right]$$

$$1 = A.18$$

$$A = \frac{1}{18}$$

$$f(X) = \frac{1}{18}(6X - X^2)$$

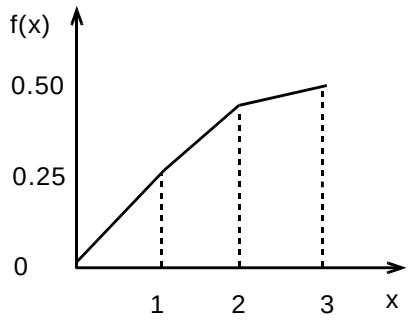
b)

$X = 0 \qquad f(X) = \frac{1}{18}\left[6(0)(0)^2\right] = 0$

$X = 1 \qquad f(X) = \frac{1}{18}\left[6(1)(1)^2\right] = \frac{5}{18}$

$X = 2 \qquad f(X) = \frac{1}{18}\left[6(2)(2)^2\right] = \frac{8}{18}$

$X = 3 \qquad f(X) = \frac{1}{18}\left[6(3)(3)^2\right] = \frac{9}{18}$



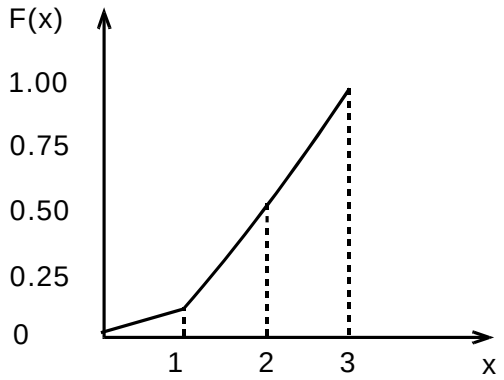
c- $F(X) = \int g(X)dx = \frac{1}{18}\int (6X - X^2)dx = \frac{1}{18}(3X^2 - \frac{X^3}{3})$

$X = 0 \qquad F(X) = \frac{1}{18}\left[3(0) - \frac{(0)^2}{3}\right] = 0$

$X = 1 \qquad F(X) = \frac{1}{18}\left[3(1) - \frac{(1)^2}{3}\right] = \frac{8}{54}$

$X = 2 \qquad F(X) = \frac{1}{18}\left[3(2) - \frac{(2)^2}{3}\right] = \frac{14}{54}$

$$X = 3 \qquad F(X) = \frac{1}{18}\left[3(3) - \frac{(3)^2}{3}\right] = \frac{18}{54}$$



d-

$$E(X) = \sum X_i p_i = \int_0^3 X f(X) dx$$

$$= \frac{1}{18}\int_0^3 (6X^2 - X^3) dx = \frac{1}{18}\left[2X^3 - \frac{X^4}{4}\right]_0^3 = 1{,}875$$

e-

$$V(X) = \frac{1}{18}\int_0^3 X^2 f(X) dx - \left[E(X)\right]^2$$

$$= \frac{1}{18}\int_0^3 (6X^3 - X^4) dx - \left[E(X)\right]^2$$

$$= \frac{1}{18}\left[\frac{6X^4}{4} - \frac{X^5}{5}\right]_0^3 - E(X)^2 = 0{,}535$$

f-

$$P(X < 2{,}4) = \frac{1}{18}\left[3X^2 - \frac{X^3}{3}\right]_0^{2,4} = 0{,}704$$

$$P(X > 1{,}8) = 1 - \frac{1}{18}\left[3X^2 - \frac{X^3}{3}\right]_0^{1,8} = 0{,}568$$

$$P(1{,}3 < X < 2{,}7) = \frac{1}{18}\left[3X^2 - \frac{X^3}{3}\right]_0^{2,7} = 0{,}608$$

BÖLÜM 8

ÇOK BOYUTLU TESADÜFİ DEĞİŞKENLER

Birden fazla olay söz konusu olduğunda tek tesadüfi değişken olayı açıklamak için yeterli olmayacağından birden fazla tesadüfi değişkene gerek olacaktır. Bu durumda olayın özelliklerine göre iki, üç,… sayıda tesadüf değişken kullanılacaktır. Kullanılan tesadüfi değişken sayısı kadar boyut söz konusu olacaktır. Birden fazla tesadüfi değişken olması genel olarak çok boyutlu tesadüfi değişkenler olarak adlandırılmaktadır.

Daha önce tek tesadüfi değişkenler için açıklandığı gibi çok boyutlu tesadüfi değişkenlerde süreksiz veya sürekli karakterli olabilirler. Buna göre ayrı ayrı incelenmeleri gereklidir. Bu bölümde sürekli ve süreksiz iki boyutlu tesadüfi değişkenlere yer verilecek, benzer şekilde ele alınabilecek ikiden fazla boyutlu tesadüfi değişkenler okuyucuya bırakılacaktır.

8.1. SÜREKSİZ TESADÜFİ DEĞİŞKENLER

İki tesadüfi değişkeni X ve Y olarak tanımlayalım. Süreksiz karakterli olan bu değişkenlerden X tesadüfi değişkeninin n sayıda, Y tesadüfi değişkeninin ise m sayıda (j=1,2, …, m) değer aldığını düşünelim. Bu X ve Y çiftlerinin gerçekleşme olasılıkları bir olasılık dağılımı oluşturur. Bu olasılık dağılımına bileşik olasılık dağılımı adı verilir. Olasılıkları $f(X_i, Y_j)$ ile ifade ederek birleşik olasılık dağılımı,

Y_j / X_i	Y_1	Y_2	.....	Y_m	$f(X_i)$
X_1	$f(X_1,Y_1)$	$f(X_1,Y_2)$	.....	$f(X_1,Y_m)$	$f(X_1)$
X_2	$f(X_2,Y_1)$	$f(X_2,Y_2)$	.....	$f(X_2,Y_m)$	$f(X_2)$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
X_n	$f(X_n,Y_1)$	$f(X_n,Y_2)$	.....	$f(X_n,Y_m)$	$f(X_n)$
$f(Y_j)$	$f(Y_1)$	$f(Y_2)$	.....	$f(Y_m)$	1

şeklinde ifade edilebilir. $f(X_i)$ olasılıkları satırların, $f(Y_j)$ olasılıkları sütunların toplam olasılıklarıdır. Alt köşede olasılıklar toplamı 1 olarak görülmektedir.

Yukarıda verilen tablodan da anlaşılacağı gibi birleşik olasılık dağılımından söz edebilmek için,

1- Olasılıklar toplamı,

$$\sum_{j=1}^{m}\sum_{i=1}^{n} f\left(X_i, Y_j\right)=1 \qquad i=1,2,3,...,n \qquad j=1,2,3,...,m$$

olmalıdır. Kısaca X ve Y çiftinin alacağı tüm değerler için olasılıklar toplanırsa, toplamın 1'e eşit olması gerekmektedir.

2- Değişken çiftinin alabileceği olasılıklar sıfır olabilir, ama sıfırdan küçük olamaz

$$f(X_i,Y_j)\geq 0 \qquad i=1,2,3,...,n \qquad j=1,2,3,...,m$$

şartlarının sağlanması gerekmektedir.

Birleşik olasılık dağılımları için birikimli olasılık dağılımları oluşturulabilir. X ve Y değişkenleri için birleşik olasılık dağılımı $F(X,Y)$ olarak ifade edilirse,

$$F(X,Y) = P\left(X \le X_i, Y \le Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} f\left(X_i, Y_j\right)$$

olarak belirlenecektir.

Yukarıda verilen birleşik olasılık dağılımı tablosunda satır ve sütun toplamları $f(X_i)$ ve $f(Y_j)$ değerleri yer almaktaydı. Tesadüfi değişkenlerin her birinin alabilecekleri farklı değerler için oluşan olasılık dağılımlarına marjinal olasılık dağılımı, kısaca marjinal dağılım denir. X tesadüfi değişkeni için marjinal dağılım,

$$f(X_i) = P\left(X = X_i\right) = \sum_{j=1}^{m} f\left(X_i, Y_j\right)$$

olacak ve marjinal olasılıklar toplamı,

$$\sum_{i=1}^{n} f(X_i) = 1$$

olacaktır. Y tesadüfi değişkeni için marjinal dağılım,

$$f(Y_j) = P\left(Y = Y_j\right) = \sum_{i=1}^{n} f\left(X_i, Y_j\right)$$

olacak ve marjinal olasılıklar toplamı,

$$\sum_{j=1}^{m} f(Y_j) = 1$$

olacaktır.

Çok boyutlu tesadüfi değişkenler için şartlı olasılıklar hesaplanabilir. Şartlı olasılık dağılımı değişkenlerden birinin bir değer aldığı varsayımı altında diğer değişkenin şartlı olasılık dağılımıdır. Şartlı olasılık ise bir tesadüfi değişkenin belirli bir değeri alması varsayımı ile diğer tesadüfi değişkenin belirli bir değeri olması olasılığıdır. X ve Y tesadüfi değişkenleri için şartlı olasılıklar, X tesadüfi değişkeni için,

$$P\left(X = X_i / Y = Y_j\right) = f\left(X_i / Y_j\right) = \frac{f\left(X_i, Y_j\right)}{f\left(Y_j\right)}$$

$$f(Y_j) > 0$$

$$\sum_{i=1}^{n} f\left(X_i / Y_j\right)=1$$

ve Y tesadüfi değişkeni için,

$$P\left(Y=Y_j / X=X_i\right)=f\left(Y_j / X_i\right)=\frac{f\left(X_i, Y_j\right)}{f\left(X_i\right)}$$

$$f(X_i)>0$$

$$\sum_{j=1}^{m} f\left(Y_j / X_i\right)=1$$

olacaktır.

Tesadüfi değişken sayısı birden fazla olduğunda değişkenler arasında ilişki olabilmesi de söz konusu olmaktadır. Diğer bir ifade ile iki veya daha fazla tesadüfi değişken söz konusu olduğunda, değişkenleri bağımsızlığı veya bağımlılığından söz etmek gerekecektir. X ve Y gibi iki tesadüfi değişken söz konusu olduğunda,

$$f(X_i, Y_j)=f\left(X_i\right).f\left(Y_j\right)$$

ise X ve Y değişkenleri bağımsızdır. Bu ifadenin $f(Y_j)$ 'ye bölümü ile,

$$f\left(X_i / Y_j\right)=f\left(X_i\right)$$

ve $f(X_i)$ 'ye bölümü ile,

$$f\left(Y_j / X_i\right)=f\left(Y_j\right)$$

sonuçları elde edilir. Bu nedenle X ve Y tesadüfi değişkenlerinin alacağı değerler için yukarıda verilen şartlar sağlanıyorsa, tesadüfi değişkenler bağımsızdır.

8.1.1. Beklenen Değer

İki veya daha fazla tesadüfi değişken söz konusu olduğunda beklenen değer hesaplanırken, tesadüfi değişkenlerin birleşimlerinin dikkate alınması gerekecektir. Örneğin, X ve Y tesadüfi değişkenleri için $X+Y, X.Y, ...$ gibi farklı ilişkiler söz konusu olabilir. Beklenen değer bu ilişkiye göre hesaplanacaktır. X ve Y tesadüfi değişkenleri arasındaki ilişkiyi $g(X,Y)$ ile ifade edersek beklenen değer genel olarak,

$$E[g(X,Y)]=\sum_{i=1}^{n}\sum_{j=1}^{m}g(X,Y)f(X_i,Y_j) \qquad i=1,2,3,...,n \qquad j=1,2,3,...,m$$

olarak ifade edilebilir. Değişkenler arasındaki ilişki $g(X,Y)=XY$ ise,

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}XYf(X_i,Y_j)$$

değişkenler arasındaki ilişki $g(X,Y)=X+Y$ ise,

$$E(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(X+Y)f(X_i,Y_j)$$

olacaktır. X ve Y tesadüfi değişkenlerinin diğer bileşimleri için de benzer şekilde beklenen değer hesaplanabilir.

İki veya daha fazla tesadüfi değişken için şartlı beklenen değer söz konusu olmaktadır. X tesadüfi değişkeninin belirli bir değer alması varsayımı ile Y değişkeninin şartlı beklenen değeri,

$$E(Y/X)=\sum_{j=1}^{m}Y_j f(Y_j/X_i)$$

ve Y tesadüfi değişkeninin belirli bir değer alması varsayımı ile X değişkeninin şartlı beklenen değeri,

$$E(X/Y)=\sum_{i=1}^{n}X_i f(X_i/Y_j)$$

olacaktır. Şartlı beklenen değer $g(X)$ veya $g(Y)$ ile ifade edeceğimiz X ve Y tesadüfi değişkenlerinin belirli bir fonksiyonu için,

$$E[g(Y)/X]=\sum_{j=1}^{m}g(Y)f(Y_j/X_i)$$

veya

$$E[g(X)/Y]=\sum_{i=1}^{n}g(X)f(X_i/Y_j)$$

olacaktır.

8.1.2. Beklenen Değerin Özellikleri

X ve Y gibi iki süreksiz tesadüfi değişken için beklenen değerin özellikleri aşağıda görüldüğü gibidir.

-Tesadüfi değişkenlerin toplamlarının beklenen değeri, X ve Y değişkenlerinin beklenen değerlerinin toplamına eşittir.

$$E(X+Y)=E(X)+E(Y)$$

Süreksiz tesadüfi değişkenler için,

$$E(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(X+Y)f(X_i,Y_j)$$

olduğundan,

$$E(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}X\,f(X_i,Y_j)+Y\,f(X_i,Y_j)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}X\,f(X_i,Y_j)+\sum_{i=1}^{n}\sum_{j=1}^{m}Y\,f(X_i,Y_j)$$

$$=\sum_{i=1}^{n}X\sum_{j=1}^{m}f(X_i,Y_j)+\sum_{j=1}^{m}Y\sum_{i=1}^{n}f(X_i,Y_j)$$

şeklinde yazılabilir. Marjinal dağılımlar,

$$f(X_i)=\sum_{j=1}^{m}f(X_i,Y_j)$$

$$f(Y_j)=\sum_{i=1}^{n}f(X_i,Y_j)$$

olduğundan,

$$E(X+Y)=\sum_{i=1}^{n}X\;f(X_i)+\sum_{j=1}^{m}Y\,f(Y_j)$$

olacaktır. X ve Y'nin beklenen değerleri,

$$E(X)=\sum_{i=1}^{n}X\;f(X_i)$$

ve

$$E(Y)=\sum_{j=1}^{m}Y\,f(Y_j)$$

olduğundan,

$$E(X+Y)=E(X)+E(Y)$$

olarak bulunur.

- Tesadüfi değişkenlerin çarpımlarının beklenen değeri, X ve Y değişkenlerinin beklenen değerlerinin çarpımına eşittir.

$$E(XY)=E(X).E(Y)$$

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}XY\,f(X_i,Y_j)$$

olduğundan bağımsızlık önem kazanacaktır. X ve Y bağımsız ise $f(X_i,Y_j)=f(X_i)f(Y_j)$ olacağından,

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}XY\,f(X_i,Y_j)$$

$$=\sum_{i=1}^{n}X\;f(X_i)+\sum_{j=1}^{m}Y\;f(Y_j)$$

olur.

$$E(X)=\sum_{i=1}^{n}X\;f(X_i)$$

ve

$$E(Y)=\sum_{j=1}^{m}Y\;f(Y_j)$$

olduğundan,

$$E(XY)=E(X).E(Y)$$

olarak elde edilir.

- X ve Y tesadüfi değişkenlerinin çarpımlarının beklenen değeri, tesadüfi değişkenlerden birinin beklenen değeri ile diğerinin şartlı beklenen değerinin çarpımına eşittir.

$$E(XY)=E(X).E(Y/X)$$

veya

$$E(XY)=E(Y).E(X/Y)$$

olacaktır.

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}XY\,f(X_i,Y_j)$$

$$f(Y_j / X_i) = \frac{f(X_i, Y_j)}{f(X_i)}$$

olduğundan,

$$f(X_i, Y_j) = f(X_i) f(Y_j / X_i)$$

olur. Buradan,

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} XY f(X_i) f(Y_j / X_i)$$

$$= \sum_{i=1}^{n} X\ f(X_i) \sum_{j=1}^{m} Y\ f(Y_j / X_i)$$

olarak elde edilir.

$$E(X) = \sum_{i=1}^{n} X\ f(X_i)$$

$$E(Y / X) = \sum_{j=1}^{m} Y\ f(Y_j / X_i)$$

olduğundan,

$$E(XY) = E(X).E(Y / X)$$

bulunur. Benzer şekilde

$$E(XY) = E(Y).E(X / Y)$$

olacaktır.

8.1.3. Varyans

X ve Y tesadüfi değişkenlerinin varyansının hesaplanmasında da X ve Y arasındaki ilişki, yani $g(X,Y)$ fonksiyonunun dikkate alınması gerekmektedir.

$$V[g(X,Y)] = E\{g(X,Y) - E[g(X,Y)]\}^2$$

olacağından bu ifadenin açılımı ile,

$$V[g(X,Y)] = E\{g(X,Y)\}^2 - E\{[g(X,Y)]\}^2$$

olarak elde edilir. Süreksiz X ve Y tesadüfi değişkenleri için varyans,

$$V[g(X,Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} \{g(X,Y) - E[g(X,Y)]\}^2 f(X_i, Y_j)$$

formülü ile hesaplanır. Bu formülün açılımı ile,

$$V[g(X,Y)]=\sum_{i=1}^{n}\sum_{j=1}^{m}[g(X,Y)]^2 f(X_i,Y_j)-\{E[g(X,Y)]\}^2$$

bulunur.

Değişkenler arasındaki ilişki $g(X,Y)=X+Y$ ise,

$$V(X+Y)=E[(X+Y)-E(X+Y)]^2$$

$$V(X+Y)=E[(X+Y)^2]-[E(X+Y)]^2$$

olacak ve süreksiz tesadüfi değişkenler için varyans,

$$V(X+Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}[(X+Y)-E(X+Y)]^2 f(X_i,Y_j)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}(X+Y)^2 f(X_i,Y_j)-E[(X+Y)]^2$$

formülleri ile hesaplanacaktır.

Değişkenler arasındaki ilişki $g(X,Y)=XY$ ise,

$$V(XY)=E[(XY)-E(XY)]^2$$

$$V(XY)=E[(XY)^2]-[E(XY)]^2$$

olacak ve süreksiz tesadüfi değişkenler için varyans,

$$V(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}[(XY)-E(XY)]^2 f(X_i,Y_j)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}(XY)^2 f(X_i,Y_j)-E[(XY)]^2$$

formülleri ile bulunacaktır.

8.1.4. Varyansın Özellikleri

X ve Y süreksiz tesadüfi değişkenlerinin varyansları ile ilgili özellikler aşağıda görüldüğü gibidir.

-Tesadüfi değişkenlerin toplamlarının varyansı,

$$V(X+Y)=V(X)+V(Y)-2[E(XY)-E(X)E(Y)]$$

olur.

$$V(X+Y)=E\left[(X+Y)-E(X+Y)\right]^2$$

olduğundan ve beklenen değerin özelliklerinden,

$$E(X+Y)=E(X)+E(Y)$$

bilindiğinden,

$$V(X+Y)=E\left[(X+Y)-\left[E(X)+E(Y)\right]\right]^2$$

$$=E\left[(X+Y)^2-2(X+Y)\left[E(X)+E(Y)\right]+\left[E(X)+E(Y)\right]^2\right]$$

$$=E\left[(X+Y)^2\right]-2E(X+Y)\left[E(X)+E(Y)\right]+\left[E(X)+E(Y)\right]^2$$

$$=E\left[(X+Y)^2\right]-2\left[E(X)+E(Y)\right]\left[E(X)+E(Y)\right]+\left[E(X)+E(Y)\right]^2$$

$$=E\left[(X+Y)^2\right]-2\left[E(X)+E(Y)\right]^2+\left[E(X)+E(Y)\right]^2$$

$$=E\left[(X+Y)^2\right]-\left[E(X)+E(Y)\right]^2$$

$$=E\left[X^2+2XY+Y^2\right]-\left[E(X)+E(Y)\right]^2$$

$$=E(X^2)+2E(XY)+E(Y^2)-\left[E(X)\right]^2-2E(X)E(Y)-\left[E(Y)\right]^2$$

$$=E(X^2)-\left[E(X)\right]^2+E(Y^2)-\left[E(Y)\right]^2+2E(XY)-2E(X)E(Y)$$

ve

$$V(X)=E\left[X-E(X)\right]^2=E(X^2)-\left[E(X)\right]^2$$

$$V(Y)=E\left[Y-E(Y)\right]^2=E(Y^2)-\left[E(Y)\right]^2$$

olarak bilindiğinden,

$$V(X+Y)=V(X)+V(Y)+2\left[E(XY)-E(X)E(Y)\right]$$

olarak elde edilir. Eşitliğin sonunda yer alan

$$E(XY)-E(X)E(Y)$$

ifadesi kovaryans olarak adlandırılır ve $Cov(X,Y)$ olarak gösterilir. Sonuç olarak,

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$$

olarak elde edilir.

-Tesadüfi değişkenler bağımsız iseler, tesadüfi değişkenlerin toplamlarının varyansı, X ve Y tesadüfi değişkenlerinin varyanslarının toplamına eşittir.

$$V(X+Y)=V(X)+V(Y)$$

X ve Y tesadüfi değişkenlerinin varyansı,

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$$

$$V(X+Y)=V(X)+V(Y)+2\left[E(XY)-E(X)E(Y)\right]$$

olarak verilmişti. Tesadüfi değişkenler bağımsız ise,

$$E(XY)=E(X)E(Y)$$

olacak ve

$$E(XY)-E(X)E(Y)=0$$

olacağından,

$$V(X+Y)=V(X)+V(Y)$$

olarak elde edilecektir. $Cov(X,Y)=0$ 'dır.

- X ve Y tesadüfi değişkenlerinin farklarının varyansı,

$$V(X-Y)=V(X)+V(Y)-2[E(XY)-E(X)E(Y)]$$

olur.

$$V(X-Y)=E[(X-Y)-E(X-Y)]^2$$

$$E(X-Y)=E(X)-E(Y)$$

olduğundan,

$$\begin{aligned}
V(X-Y)&=E[(X-Y)-[E(X)-E(Y)]]^2\\
&=E\left[(X-Y)^2-2(X-Y)[E(X)-E(Y)]+[E(X)-E(Y)]^2\right]\\
&=E\left[(X-Y)^2\right]-2E(X-Y)[E(X)-E(Y)]+[E(X)-E(Y)]^2\\
&=E\left[(X-Y)^2\right]-[E(X)-E(Y)]^2\\
&=E\left[X^2-2XY+Y^2\right]-[E(X)-E(Y)]^2\\
&=E(X^2)-2E(XY)+E(Y^2)-[E(X)]^2-2E(X)E(Y)-[E(Y)]^2\\
&=E(X^2)-[E(X)]^2+E(Y^2)-[E(Y)]^2-2E(XY)-2E(X)E(Y)
\end{aligned}$$

ve

$$V(X)=E(X^2)-[E(X)]^2$$

$$V(Y)=E(Y^2)-[E(Y)]^2$$

olduğundan,

$$\begin{aligned}
V(X-Y)&=V(X)+V(Y)-2[E(XY)-E(X)E(Y)]\\
&=V(X)+V(Y)+2Cov(X,Y)
\end{aligned}$$

bulunur.

-Bağımsız tesadüfi değişkenin farklarının varyansı, tesadüfi değişkenlerin varyansları toplamına eşittir.

$$V(X-Y)=V(X)+V(Y)$$

olacaktır.

$$V(X-Y)=V(X)+V(Y)-2[E(XY)-E(X)E(Y)]$$

olarak bilinmektedir. Tesadüfi değişkenler bağımsız ise,

$$E(XY)-E(X)E(Y)=0$$

olacağından,

$$V(X-Y)=V(X)+V(Y)$$

olur.

-Bir tesadüfi değişkenin kendisi ile toplamının varyansı, kendi varyansının 4 katına eşittir.

$$V(X+X)=4V(X)$$

olacaktır.

$$V(X+X)=V(2X)=E[2X-E(2X)]^2$$

$E(2X)=2E(X)$ olduğundan,

$$V(X+X)=E[2X-E(2X)]^2$$

$$=E[2[X-E(X)]]^2$$

$$=4E[X-E(X)]^2$$

$$=4V(X)$$

bulunur.

$V(X+X)\neq V(X)+V(X)$'dir.

ÖRNEK:

X \ Y	0	1	2	
1	0.25	0.30	0.10	
2	0	0.10	0.25	

olarak verildiğine göre,

a) P (X=1, Y>1) olasılığını,
b) Y=1 varsayıldığına göre X=2 şartlı olasılığını,
c) 9(X,Y)=XY için beklenen değeri,
d) 9(X,Y)=X+Y için varyansı hesaplayınız.

a) P(x=1, y>,1) = f (1,1)+f(1,2)
= 0,30 + 0,10
= 0,40

b) $P(X=2/Y=1) = \dfrac{f(2,1)}{fy(1)}$

$$P(Y=1) = fy(1) = \sum_{i=1}^{2} f(x_i,1)$$

$$= f(1,1) + f(2,1) = 0,30 + 0,10 = 0,4$$

$$f(x=2/Y=1) = \frac{f(2,1)}{fy(1)}$$

$$= \frac{0,10}{0,40}$$

$$= 0,25$$

c) $E(XY) = \sum_{i=1}^{2} \sum_{j=1}^{3} XY\, f(X_i, Y_i)$

$= (1)(0)\ 0,25+(1)(1).\ 0,30+(1)(2)\ 0,10+(2)(0)\ 0+(2)\ (1)0,10+(2)(2)$
$0,25$
$= 0+0,30+0,20+0+2,20+1$
$= 1,70$

d) $V(X+Y) = \sum_{i=1}^{2} \sum_{j=1}^{3} (X+Y)^2 f(X_i, Y_i) - \left[E(X+Y)\right]^2$

$= (1+0)^2\ 0,25+(1+1)^2\ 0,30+(1+2)^2\ 0,10+(2+0)^2\ 0+\ (2+1)^2$
$0,10+(2+2)^2 0,25-(1,70)^2$
$= 0,25+1,2+0,90+0+0,90+4$
$= 7,25-(1,70)^2$
$= 4,36$

8.2. SÜREKLİ TESADÜFİ DEĞİŞKENLER

Sürekli X ve Y tesadüfi değişkenlerinin birleşik olasılık yoğunluk fonksiyonu f(x,y) olarak tanımlanırsa, x tesadüfi değişkenin tanım aralığı a-b, Y tesadüfi değişkenin tanım aralığı c-d ise, f(x,y)nin birleşik olasılık yoğunluk fonksiyonu olabilmesi için;

1- X ve Y tesadüfi değişkenlerinin tanımlı olduğu aralıktaki tüm değerler için,

$f(x,y) \geq 0$

2- Söz konusu aralıklar için,

$$\int_c^d \int_a^b f(x, y)dxdy = 1$$

olmalıdır.

Sürekli X ve Y tesadüfi değişkenleri için birikimli birleşik dağılım fonksiyonu F(X,Y), X tesadüfi değişkeninin alabildiği bir X değerine eşit veya küçük ve Y tesadüfi değişkeninin alabileceği bir Y değerine eşit veya küçük değer alma olasılığını verir ve

$$F(X,Y) = \int\int f(x, y)dxdy$$

olarak belirlenir. F(X,Y)'nin X ve Y göre kısmi türevleri birleşik olasılık yoğunluk fonksiyonunu eşit olacaktır.

$$f(X, Y) = \frac{\partial^2 F(X, Y)}{\partial X \partial Y}$$

X tesadüfi değişkenini X_i gibi bir değer aldığında, X tesadüfi değişkeninin marjinal dağılımı,

$$P(X = X_i) = f(X_1) = \int_c^d f(x, y)dy \qquad a < x < b;\ c < y < d$$

olacaktır. Aynı şekilde Y tesadüfi değişkeni Y_j gibi bir değer aldığında, Y tesadüfi değişkeninin marjinal dağılımı,

$$P(X = Y_i) = f(Y_1) = \int_a^b f(x, y)dy \qquad a < x < b;\ c < y < d$$

olur. Tesadüfi X ve Y değişkenlerinin şartlı dağılım fonksiyonları,

$$f(X, Y) = \frac{f(X, Y)}{f(Y)} \qquad [f(Y) \neq 0]$$

$$f(Y/X) = \frac{f(X, Y)}{f(X)} \qquad [f(X) \neq 0]$$

şeklinde belirlenir.

$$f(X, Y) = f(X)\ f(Y)$$

olması durumunda X ve Y tesadüfi değişkenleri bağımsızdır.

8.2.1. Beklenen Değer

İki veya daha fazla süreksiz tesadüfi değişken olması durumunda olduğu gibi, iki veya daha fazla sürekli tesadüfi değişken olması durumunda da tesadüfi değişkenlerin birleşiminin dikkate alınması gerekecektir. Tesadüfi değişkenler arasındaki ilişkiyi g(x,y) olarak ifade edersek beklenen değer,

$$E\left[g(X,Y)\right]=\int_{c}^{d}\int_{a}^{b} g(X,Y)\quad f\,(X,Y)\,dxdy \qquad a<X<b$$

$$c < Y < d$$

şeklinde ifade edilir.

$g\,(X,\,Y)\;=\;X\;Y$ ise,

$$E\,(XY)=\int_{c}^{d}\int_{a}^{b} X\,Y\;\;f\,(X,\,Y)\;\;dxdy$$

$g\,(X,\,Y)\;=X+Y$ ise,

$$E\,(X+Y)=\int_{c}^{d}\int_{a}^{b}(X+Y)\;\;f\,(X,\,Y)\;\;dxdy$$

olacaktır.

X tesadüfi değişkeninin belirli bir değer alması durumunda Y tesadüfi değişkeninin şartlı beklenen değeri,

$$E\,(XY)=\int_{c}^{d}\int_{a}^{b} X\,Y\;\;f\,(X,\,Y)\;\;dxdy$$

Y tesadüfi değişkeninin belirli bir değer alması durumunda X tesadüfi değişkeninin şartlı beklenen değeri ise,

$$E\,(X/Y)=\int_{a}^{b} Xf\;(X\,/\,Y)\,dx$$

olacaktır.

X veya Y tesadüfi değişkeninin belirli fonksiyonu g(X) veya g(Y) ise, şartlı beklenen değer bunlar için de hesaplanabilir. Bu durumda,

$$E\left[g\,(Y)\,/\,(X)\right]=\int_{c}^{d} g\,(Y)\;f\;(Y/X)\,dy$$

veya

$$E\quad g\,(X)\,/\,Y)=\int_{a}^{b} g\,(X)\,f\;(X/Y)\,dx$$

olacaktır.

8.2.2. Beklenen Değerin Özellikleri

Süreksiz tesadüf değişkenlerin beklenen değerini açıklarken söz konusu olan özellikler, sürekli tesadüfi değişkenlerin beklenen değeri içinde geçerlidir. Bu özellikler aşağıda verilmiştir.

- Tesadüfi değişkenlerin toplamlarının beklenen değeri, X ve Y değişkenlerinin beklenen değerlerinin toplamına eşittir.

$$E(X+Y) = E(X) + E(Y)$$

X ve Y sürekli tesadüfi değişkenleri için,,

$$E(X+Y) = \int_c^d \int_a^b (X+Y)\ f(X,Y)\ dxdy$$

olduğundan,

$$E(X+Y) = \int_c^d \int_a^b \left[Xf(X,Y) + Yf(X,Y)\right] dxdy$$

$$= \int_c^d \int_a^b Xf(X,Y)\,dxdy + \int_c^d \int_a^b Yf(X,Y)\,dxdy$$

olacaktır. Marjinal dağılımlardan söz ederken yapılan açıklamalardan,

$$\int_c^d f(X,Y)\,dy = f(X)$$

$$\int_a^b f(X,Y)\ dx = f(Y)$$

olarak bilinmektedir. Bu nedenle,

$$E(X+Y) = \int_a^b Xf(X)\,dx + \int_c^d Yf(Y)\ dy$$

$$E(X) = \int_a^b Xf(X)\,dx$$

$$E(Y) = \int_c^d Yf(Y)\,dy$$

olduğundan,

$$E(X+Y) = E(X) + E(Y)$$

olarak bulunur.

- Tesadüfi değişkenlerin çarpımlarının beklenen değeri, X ve Y tesadüfi değişkenlerinin beklenen değerlerinin çarpımına eşittir.

$$E(XY) = E(X)\ E(Y)$$

X ve Y sürekli tesadüfi değişkenleri için,

$$E(XY) = \int_c^d \int_a^b XY\, f(X, Y)\, dxdy$$

olacaktır. X ve Y bağımsız ise $f(X, Y)=f(X)\, f(Y)$ olacağından,

$$E(XY) = \int_c^d \int_a^b XY\, f(X)\, f(Y)\, dxdy$$

$$= \int_a^b Xf(X)\, dx \int_c^d Yf(Y)\, dy$$

olur.

$$E(X) = \int_a^b Xf(X)\, dx$$

$$E(Y) = \int_c^d Yf(Y)\, dy$$

olduğundan,

$$E(X, Y) = E(X)\, E(Y)$$

elde edilir.

- Tesadüfi değişkenlerin çarpımlarının beklenen değeri, tesadüfi değişkenlerden birinin beklenen değeri ile diğerinin şartlı beklenen değerinin çarpımına eşittir.

$$E(XY) = E(X)\, E(Y/X)$$

veya

$$E(XY) = E(Y)\, E(X/Y)$$

X ve Y sürekli tesadüfi değişkenler ise,

$$E(XY) = \int_c^d \int_a^b XYf(X, Y)\, dxdy$$

$$f(Y/X) = \frac{f(X, Y)}{f(X)}$$

$$f(X,Y) = f(X)\ f(Y/X)$$

$$E(XY) = \int_c^d \int_a^b XYf(X)\ f(Y/X)\ dxdy$$

$$E(X) = \int_a^b Xf(X)\ dx$$

$$E(Y/X) = \int_c^d Yf(Y/X)\ dy$$

olduğundan,

$$E(XY) = E(X)\ E(Y/X)$$

olacaktır.

8.2.3. Varyans

X ve Y tesadüfi değişkenlerinin varyansının hesaplanmasında da, X ve Y arasındaki ilişki, yani g (X,Y) fonksiyonunun dikkate alınması gerekmektedir.

$$V[g(X,Y)] = E\{g(X,Y) - E[g(X,Y)]\}^2$$

olacağından sürekli X ve Y tesadüfi değişkenleri için,

$$V[g(X,Y)] = \int_c^d \int_a^b \{g(X,Y) - E[g(X,Y)]\}^2\ f(X,Y)\ dxdy$$

$$= \int_c^d \int_a^b [g(X,Y)]^2\ f(X,Y)\ dxdy - \{E[g(X,Y)]\}^2$$

olarak bulunur.

$g(X,Y) = X + Y$ ise,

$$V(X+Y) = E[(X+Y) - E(X+Y]^2$$

$$V(X+Y) = E[(X+Y)^2] - [E(X+Y)]^2$$

olur. Bu durumda,

$$V(X+Y) = \int_c^d \int_a^b \left[(X+Y) - E(X+Y) \right]^2 f(X,Y)dxdy$$

$$= \int_c^d \int_a^b (X+Y)^2 \; f(X,Y)\, dxdy - \left[E(X+Y) \right]^2$$

olacaktır.

$g(X,Y) = XY$ ise,

$V(XY) = E[(XY) - E(XY)]^2$

$V(XY) = E[XY)^2] - [E(XY)]^2$

olur. Bu durumda,

$$V(XY) = \int_c^d \int_a^b \left[(XY) \right] 2 \; f(X,Y)\, dxdy$$

$$= \int_c^d \int_a^b (XY)^2 \; f(X,Y)\, dxdy - \left[E(XY) \right]^2$$

olacaktır.

Çok boyutlu sürekli tesadüfi değişkenler için varyansın özellikleri, çok boyutlu tesadüfi değişkenler için açıklanan varyansın özellikleri ile aynıdır.

8.3. KOVARYANS VE ÖZELLİKLERİ

Kovaryans iki tesadüfi ilişki arasında doğrusal ilişki olup olmadığını ve ilişki varsa yönünü belirleyen ölçüdür. Kovaryans $Cov(X,Y)$ veya σ_{XY} şeklinde gösterilir. Korvaryans sıfır olduğunda değişkenler arasında ilişki yoktur. Korvaryans negatif işaretli ise değişkenler arasında ters yönlü, pozitif işaretli ise değişkenler arasında doğru yönlü ilişki vardır.

$$\sigma_{XY} = Cov(X,Y) = E\left[X - E(X)\right]\left[Y - E(Y)\right]$$

olarak tanımlanan kovaryans süreksiz değişkenler için,

$$\sigma_{XY} = Cov(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[X_i - E(X)\right]\left[Y_j - E(Y)\right] f\left(X_i, Y_j\right)$$

olur. Tanım formülünün açılımı ile,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

olarak bulunur ve,

$$Cov(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m} XY\ f(X_i,Y_j)-E(X)E(Y)$$

olarak hesaplanır.

Kovaryansın Özellikleri

Kovaryansın özellikleri aşağıda verilmiştir.:

- Değişkenlerin sırasının değişimi kovaryansı değiştirmez.

$$Cov(X,Y)=Cov(Y,X)$$

ifadesini açık olarak yazılırsa,

$$Cov(X,Y)=E[X-E(X)][Y-E(Y)]$$

$$Cov(Y,X)=E[Y-E(Y)][X-E(X)]$$

olarak elde edilir.

- Bir tesadüfi değişkenin kendisi ile ilişkisi için hesaplanan kovaryans, değişkenin varyansına eşittir.

$$Cov(X,Y)=Var(X)$$

olacaktır.

$$\begin{aligned} Cov(X,X) &= E[X-E(X)][X-E(X)] \\ &= E[X-E(X)]^2 \\ &= Var(X) \end{aligned}$$

olur.

-İki tesadüfi değişkenin sabit sayılar ile çarpımlarının kovaryansı, değişkenlerin çarpıldıkları sabit sayılar ile çarpıma eşittir.

$$Cov(aX,bY)=abCov(X,Y)$$

olacaktır.

$$\begin{aligned} Cov(aX,bY) &= E[aX-E(aX)][bY-E(bY)] \\ &= E[aX-aE(X)][bY-bE(Y)] \\ &= E\{a[X-E(X)]\}\{b[Y-E(Y)]\} \\ &= abE[X-E(X)][Y-E(Y)] \\ &= abCov(X,Y) \end{aligned}$$

olur.

-İki tesadüfi değişkenin sabit sayılar ile toplamının kovaryansı değişmez.

$$Cov(a+X),(b+Y)=Cov(X,Y)$$

olacaktır.

$$\begin{aligned} Cov(a+X),(b+Y) &= E[(a+X)-E(a+X)][(b+Y)-E(b+Y)] \\ &= E[a+X-a-E(X)][b+Y-b-E(Y)] \\ &= E[X-E(X)][Y-E(Y)] \\ &= Cov(X,Y) \end{aligned}$$

olur.

-Üç tesadüfi değişken için,

$$Cov(X+Z,Y)=Cov(X,Y)+Cov(Z,Y)$$

olacaktır.

$$\begin{aligned} Cov(X+Z,Y) &= E[(X+Z)Y]-E[X+Z]E[Y] \\ &= E[XY]-E[ZY]-\{E(X)+E(Z)\}E(Y) \\ &= E[XY]-E[X]E[Y]+E[ZY]-E[Z]E[Y] \end{aligned}$$

$$Cov(X,Y)=E[XY]-E[X]E[Y]$$

$$Cov(Z,Y)=E[ZY]-E[Z]E[Y]$$

olduğundan,

$$Cov(X+Z,Y)=Cov(X,Y)+Cov(Z,Y)$$

bulunur.

ÖRNEK: f (X,Y) = 4XY O< X < 1
O< Y <1

olarak verildiğine göre;

a) P (X>0,7, Y<0,4) olasılığını

b) E(X+Y) beklenen değerini

c) V(X+Y) varyansını hesaplayınız.

Çözüm:

a) $P(X > 0{,}7,\quad Y < 0{,}4) = \int_{0}^{0,4}\int_{0,7}^{1}(4XY)\,dxdy$

$$= 4\int_{0}^{0,4}\left(\frac{X^2Y}{2}\Big|_{0,7}^{1}\right)dy = 4\int_{0}^{0,4}\left(\frac{Y}{2}-\frac{0{,}49Y}{2}\right)dy$$

$$= 4\int_0^{0,4}\left(\frac{0,51.\ y}{2}\right)dy = \frac{0,51.4}{2}\left(\frac{Y^2}{2}\Big|_0^{0,4}\right)$$

$$= 1.02.\left(\frac{0,16}{2}\right) = 0,08$$

b) $E(X+Y) = \int_0^1\int_0^1 (X+Y)\,(4XY)\ dxdy$

$$= \int_0^1\int_0^1 (4X^2Y)\,dxdy + \int_0^1\int_0^1 (4XY)^2\,dxdy$$

$$= 4\int_0^1\left(\frac{X^3Y}{3}\Big|_0^1\right)dy + 4\int_0^1\left(\frac{X^2Y^2}{2}\Big|_0^1\right)dy$$

$$= \frac{4}{3}\left(\frac{Y^2}{2}\Big|_0^1\right) + \frac{4}{2}\left(\frac{Y^3}{3}\Big|_0^1\right)$$

$$= \frac{4}{3}.\frac{1}{2} + \frac{4}{2}.\frac{1}{3} = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

c) $V(X+Y) = \left[\int_0^1\int_0^1 (X+Y)^2\,(4XY)dxdy\right] - \left[\left(\frac{4}{3}\right)^2\right]$

$$\left[\int_0^1\int_0^1 (X^2+2XY+Y^2)\,(4XY)dxdy\right] - \frac{16}{9}$$

$$= 4\left[\int_0^1\int_0^1 (X^3Y)\,dxdy + \int_0^1\int_0^1 (2X^2Y^2)dxdy + \int_0^1\int_0^1 (XY^3)dxdy\right] - \frac{16}{9}$$

$$= 4\left[\int_0^1\left(\frac{X^4Y}{4}\Big|_0^1\right)dx + 2\int_0^1\left(\frac{X^3Y^2}{3}\Big|_0^1\right)dy + \int_0^1\left(\frac{X^2Y^3}{2}\Big|_0^1\right)dy\right] - \frac{16}{9}$$

$$= \left(\frac{Y^2}{2}\Big|_0^1 + 8.\frac{Y^3}{9}\Big|_0^1 + 4\frac{Y^4}{8}\Big|_0^1\right) - \frac{16}{9}$$

$$= \left(\frac{1}{2} + \frac{8}{9} + \frac{1}{2}\right) - \frac{16}{9} = \frac{17}{9} - \frac{16}{9} = \frac{1}{9}$$

8.4. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1:

X \ Y	1	2	3	4
8	0,1	0,2	0	0,1
9	0	0,1	0,1	0
10	0,2	0	0,1	0,1

olarak verilmiştir.

a) X'in 10'dan küçük ve Y'nin 2 veya daha büyük değer alma olasılığını hesaplayınız.

b) Y=3 için marjinal olasılığı hesaplayınız.

c) X=9 olarak varsayıldığına göre Y=2 şartlı olasılığını hesaplayınız.

d) g (X,Y) = X+Y için beklenen değeri hesaplayınız.

e) g (X,Y) = XY için varyansı hesaplayınız.

Çözüm:

i=1,2,3
j= 1,2,3,4

a) $P(X<10,\ Y\geq 2)=\sum_{j=2}\sum_{i=3}^{4} f(Xi, Yj)$

= f(8) (2) + f(8) (3) + f (8) (4) + f (9) (2) + f (9) (3) + f (9) (4)
= 0,2 + 0 + 0,1 + 0,1 - 0,1 + 0
= 0,5

b) $P(Y=3)=fy(3)\sum_{i=1}^{3}(Xi,3)$

= f (8,3) + f (9,3) + f (10,3)
= 0+0,1+0,1 = 0,2

c) $P(Y=2\ X=9)\ \frac{f(9,2)}{fx(9)}$

$P(X=9)=fx(9)\sum_{j=1}^{4} f(g, Yj)$

= f (9,1) + f (9,2) + f (9,3) + f (9,4)
= 0 + 0,1 + 0,1 + 0 = 0,2

$$P\,(Y=2/X=9) = \frac{f(9,2)}{f_x(9)}$$

$$= \frac{0,1}{0,2} = 0,5$$

d) $E(X, Y) = \sum_{i=1}^{3} \sum_{j=1}^{4} (X + Y)\, f\,(X_i, Y_j)$

$= (8+1)\,(0,1) + (8+2)\,(0,2) + (8+3)\,(0) + (8+4)\,(0,1) + (9+1)\,(0) +$
$(9+2)\,(0,1) + (9+3)\,(0,1) + (9+4)\,(0) + (10+1)\,(0,2) + (10+2)\,(0)$
$+ (10+3)\,(0,1) + (10+4)\,(0,1)$
$= 11,3$

e) $V(XY) = E\left[(XY)^2\right] - \left[E(XY)\right]^2$

$E\left[(XY)^2\right] = \left[(8).(1)\right]^2.(0,1) + \left[(8).(2)\right]^2(0,2) + \left[(8)(3)\right]^2(0) +$
$\left[(8).(4)\right]^2(0,1)$
$+ \left[(9).(1)\right]^2(0) + \left[(9).(2)\right]^2(0,1) + \left[(9).(3)\right]^2(0,1) +$

$\left[(9).(4)\right]^2(0) + \left[(10).(1)\right]^2(0,2) + \left[(10).(2)\right]^2(0) +$

$\left[(10).(3)\right]^2(0,1) + \left[(10).(4)\right]^2(0,1)$

$= 6,4 + 51,2 + 0 + 102,4 + 0 + 32,4 + 4 + 72,9 + 0 +$
$20 + 0 + 90 + 160 = 535,3$

$E(XY) = \sum_{i=1}^{3} \sum_{j=1}^{4} (XY)\, f\,(Xi, Yj)$

$= (8).(1)\,(0,1) + (8).(2)\,(0,2) + (8).(3)\,(0) + (8).(4)\,(0,1) + (9).(1)$
$(0) + (9).(2)\,(0,1) + (9).(3)\,(0,1) + (9).(4)\,(0) + (10).(1)\,(0,2) +$
$(10).(2)\,(0) + (10).(3)\,(0,1)\ \ (10).(4)\,(0,1)$
$= 0,8+3,2+0+3,2+0+1,8+2,7+0+2+0+3+4$
$= 20,7$

$V(XY) = E\left[(XY)^2\right] - \left[E(XY)^2\right]$
$= 535,3 - (20,7)^2$
$= 106,81$

ÖRNEK 2: $f(X,Y) = \frac{1}{3}(X+Y)$ $\quad 0<X<1$, $\quad 0<Y<2$

olarak tanımlanmıştır.

a) P(X<0,2, Y>0,8) olasılığını hesaplayınız.

b) Beklenen değer E(XY)'yi hesaplayınız.

c) Varyans V(XY)'yi hesaplayınız.

Çözüm:

a) $P(X<0,2,\ Y>0,8) = \int_{0,8}^{1}\int_{0}^{0,2} f(X,Y)\,dxdy$

$$= \frac{1}{3}\int_{0,8}^{1}\int_{0}^{0,2}(X+Y)\,dxdy$$

$$= \frac{1}{3}\int_{0,1}^{1}\left(\frac{x^2}{2}+xy\right)\Big|_{0}^{0,2}dy = \frac{1}{3}\int_{0,1}^{1}\left(\frac{0,04}{2}+0,2\ Y\right).dy$$

$$= \frac{1}{3}\left[\left(0,02y+\frac{0,2y^2}{2}\right)\Big|_{0,1}^{1}\right]$$

$$= \frac{1}{3}\left[(0,02+0,1) - (0,002+0,001)\right]$$

$$= \frac{1}{3}(0,117) = 0,039$$

b) $E(X,Y) = \int_{c}^{d}\int_{a}^{b} XY\ f(X_1Y)\,dxdy$

$$= \int_{0}^{1}\int_{0}^{2} XY\ f(XY)\,dxdy$$

$$= \int_{0}^{1}\int_{0}^{2} XY\ \frac{1}{3}(X+Y)\,dxdy$$

$$= \frac{1}{3}\int_0^1 \int_0^2 (X^2Y + XY^2)dxdy$$

$$= \frac{1}{3}\left[\int_0^1 \int_0^2 (X^2Y)\,dxdy + (XY^2)dxdy\right]$$

$$= \frac{1}{3}\int_0^1 \left[\frac{X^3}{3}Y\Big|_0^2 dy + \frac{X^2Y^2}{2}\Big|_0^2\right] dy$$

$$= \frac{1}{3}\left[\frac{8}{3}\int_0^1 Y.\,dy + 2.\int_0^1 Y^2.dy\right]$$

$$= \frac{1}{3}\left[\frac{8}{3}\left(\frac{Y^2}{2}\Big|_0^1\right) + 2.\left(\frac{Y3}{3}\right)\Big|_0^1\right]$$

$$= \frac{1}{3}\left[\frac{8}{3}.\frac{1}{2} + \frac{2}{3}\right] = \frac{2}{3}$$

c) $V(XY) = \int_c^d \int_a^b (XY)2\, f\,(XY)dxdy - \left[E\,(XY)\right]^2$

$$= \int_0^1 \int_0^2 (XY)^2 \frac{1}{3}(X+Y)\,dxdy - \left(\frac{2}{3}\right)^2$$

$$= \frac{1}{3}\int_0^1 \int_0^2 \left(X^3Y^2 + X^2Y^3\right)\Big|_0^2 dy - \frac{4}{9}$$

$$= \frac{1}{3}\int_0^1 \left(\frac{X^4Y^2}{4} + \frac{X^3Y^3}{3}\right)\Big|_0^2 dy - \frac{4}{9}$$

$$= \frac{1}{3} \int_0^1 \left(4Y^2 + \frac{8}{3} Y^3 \right) dy - \frac{4}{9}$$

$$= \frac{1}{3} \left[\left(\frac{4Y^3}{3} + \frac{8.Y^4}{3.4} \right) \right] \Big|_0^1 - \frac{4}{9}$$

$$= \frac{1}{3} \left(\frac{4}{3} + \frac{2}{3} \right) - \frac{4}{9} = \frac{2}{9}$$

BÖLÜM 9

OLASILIK DAĞILIMLARI

Tesadüfi değişkenler bölümünde, tesadüfi değişkenlerin olasılık veya olasılık yoğunluk fonksiyonlarından sözetmiştik. Olayın veya değişkenin taşıdığı şartlara bağlı olarak olasılık hesaplanması için teorik olasılık dağılımları veya kısaca olasılık dağılımları olarak adlandırılan matematiksel kalıplar kullanılabilmektedir. Farklı durumlar için çok sayıda olasılık dağılımı bulunmaktadır. Bu bölümde en yaygın olarak kullanılan süreksiz olasılık dağılımlarından Binom, Poisson, Hipergeometrik Dağılımlar; sürekli dağılımlardan Normal Dağılım'a yer verilecektir. Ayrıca örnekleme dağılımı olmaları nedeniyle ki-kare, t ve F dağılımları da özetlenecektir.

9.1. BİNOM DAĞILIMI

Binom dağılımı uygulamada en çok kullanılan süreksiz karakterli olasılık dağılımıdır. Bu dağılımda X tesadüfi değişken 0-n aralığında bulunan tüm tam sayı değerleri almaktadır. (X=0,1,2,....,n).

Binom dağılımının uygulanması bazı şartlara bağlıdır.

1- Olayda bir tek karakterin olumlu ve olumsuz durumu sözkonusu olmalıdır. Örneğin, bir mamül hatalıdır veya hatasızdır gibi.
2- Olayda deneme n defa, sonlu sayıda tekrarlanmalıdır.
3- Denemeler (tekrarlar) birbirinden bağımsız olmalıdır. (İadeli seçim gibi.)
4- Denemelerden sonra olasılık (p) ve ters olasılık (q) değişmemelidir.
5- Binom dağılımında olasılık genellikle 0.05 veya daha büyük olacaktır ($P \geq 0.05$).

Binom dağılımını bir tek denemenin sözkonusu olduğu Bernoulli dağılımı veya değişkeni ile açıklayabiliriz. Bernoulli değişkeni bir denemede istenen sonuç gerçekleşirse 1; gerçekleşmezse 0 değerini alan bir değişkendir.

Bir paranın atıldığını düşünelim. Bu olayda iki farklı durum ($2^n=2^1=2$) sözkonusudur; yazı veya tura gelebilir. Yazı gelmesini istiyorsak, istemin gerçekleşme olasılığı (denemenin başarılı olma olasılığı),

$$P(Y) = P = ½$$

dir. İstenenin gerçekleşmeme olasılığı (denemenin başarısız olma olasılığı)

$$P(Y) = P(T) = 1-p= q = ½$$

olacaktır.

X	Olasılık	Deneme
0	1-p=q	Başarısız (tura)
1	p	Başarılı (yazı)

Para iki kere atılırsa, $2^n=2^2=4$ değişik durum gerçekleşebilir.

$$P(Y\cap Y) = P(Y) \,.\, P(Y) = \; p \,.\, p = p^2$$

$$P(T\cap Y) = P(T) \,.\, P(Y) = \; q \,.\, p$$

$$P(Y\cap T) = P(Y) \,.\, P(T) = \; p \,.\, q$$

$$P(T\cap T) = P(T) \,.\, P(T) = \; q \,.q = q^2$$

olacaktır.

X yazı sayısını gösterdiğine göre,

X	Olasılık
0	q^2
1	pq+pq=2pq
2	p^2

Para 3 kere atlıyorsa $2^n=2^3=8$ değişik durum gerçekleşebilir. Bu durumda,

$$P(Y\cap Y\cap Y) = P(Y) \,.\, P(Y) \,.\, P(Y) = \; p \,.\, p \,.\, p \; = p^3$$

$$P(T\cap Y\cap Y) = P(T) \,.\, P(Y) \,.\, P(Y) = \; q \,.\, p \,.\, p = p^2q$$

$$P(Y\cap T\cap Y) = P(Y) \,.\, P(T) \,.\, P(Y) = p \,.\, q \,.\, p = p^2q$$

$$P(Y\cap Y\cap T) = P(Y) \,.\, P(Y) \,.\, P(T) = \; p \,.\, p \,.\, q \; = p^2q$$

$$P(Y\cap T\cap T) = P(Y) \,.\, P(T) \,.\, P(T) = \; p \,.\, q \,.\, q \; = pq^2$$

$$P(T\cap Y\cap T) = P(T) \,.\, P(Y) \,.\, P(T) = \; q \,.\, p \,.\, q \; = pq^2$$

$$P(T\cap T\cap Y) = P(T) \,.\, P(T) \,.\, P(Y) = \; q \,.\, q \,.\, p \; = pq^2$$

$$P(T\cap T\cap T) = P(T) \,.\, P(T) \,.\, P(T) = \; q \,.q \,.q = q^3$$

olacaktır.

X	Olasılık
0	q^3
1	$pq^2+pq^2+pq^2=3pq^2$
2	$p^2q+p^2q+p^2q=3p^2p$
3	p^3

Deneme sayısı arttırıldığında da olasılıklar benzer şekilde hesaplanır. k, X'in aldığı değerleri ifade ediyorsa, yukarıdaki denemelere dayanarak olasılıklar $P(X=k)=p^k q^{n-k}$ olarak yazabiliriz. Bu sıralı olasılıktır. Dikkat edilirse, sıra sayısı C_n^k 'dır ve sırasız olasılık,

$$P(X=k) = C_n^k p^k q^{n-k}$$

olarak hesaplanabilir.

Binom dağılımı iki parametreli bir dağılımdır ve parametreleri deneme sayısı (n) ve denemenin başarılı olması olasılığı (p)'dir. Binom dağılımın genel ifadesi,

$$B(n;p) = P(X=k) = C_n^k p^k q^{n-k}$$

dır. Bernoulli dağılımında,

X	0	1
Olasılık	q	p

olduğundan beklenen değer,

$$E(X)=\sum_{i=1}^{n} X_i p_i = 0(q)+1(p) = p$$

bulunur. Binom dağılımında deneme n defa tekrarlanacağından,

$$E(X) = E(X_1) + E(X_2) + E(X_3) + \ldots\ldots + E(X_n)$$

$$E(X) = \underbrace{p+p+p+\ldots\ldots+p}_{n} = np$$

olacaktır.

Bernoulli dağılımında varyans,

$$V(X)=\sum_{i=1}^{n} p_i X_i^2 - E(X)^2 = (1)^2 p + (0)^2 p - p^2 = p - p^2 = p(1-p) = pq$$

ve binom dağılımında varyans,

$$V(X) = V(X_1) + V(X_2) + V(X_3) + ... + V(X_n)$$
$$V(X) = \underbrace{pq + pq + pq + ... + pq}_{n} = npq$$

olacaktır.

Binom dağılımında ortalamaya göre 3. ve 4. momentler,

$$\mu_3 = npq(q - p)$$
$$\mu_4 = 3n^2p^2q^2 + npq(1 - 6pq)$$

olduğundan eğiklik ve basıklık ölçüleri,

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{(n-q)^2}{npq}$$
$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}$$

olacaktır.

ÖRNEK: Yazı gelmesi olasılığı 0.48 olan bir para 6 kere atılıyor.

a- 4 yazı gelmesi,
b- 1 tura gelmesi,
c- Yazı gelmemesi olasılıklarını hesaplayınız.

a- n =6
p =0.48
q = 1-p = 1-0.48 =0.52
$P(X=k) = C_n^k p^k q^{n-1}$
$P(X=4) = C_6^4 (0.48)^4 (0.52)^{6-4}$
$C_6^4 = \frac{6!}{4!(6-4)!} = \frac{4!.5.6}{4!.2!} = 15$
$P(X=4) = 15.(0.48)^4.(0.52)^2 = 0.2153$

b- Burada iki farklı hesaplama yapabiliriz.

-1 tura gelmesi için 5 yazı gelmelidir.

$P(X=5) = C_6^5 (0.48)^5 (0.52)^1 = 0.0794$

-p ve q yer değiştirirse,

$P(X=1) = C_6^1 (0.52)^1 (0.48)^5 = 0.0794$

1 tura gelmesi olasılığı her iki şekilde de hesaplanabilir.

c-P(X=0) = $C_6^0(0.48)^0(0.52)^6 = 1.1.(0.56)^6 = 0.0197$ (Bu olasılık 6 tura gelmesi olasılığına eşittir.)

9.2. HİPERGEOMETRİK DAĞILIM

Binom dağılımı gibi hipergeometrik dağılımda süreksiz karakterli dağılımlardandır. Binomdan farklı olarak, hipergeometrik dağılım iadesiz seçim için geçerlidir, p ve q ise her denemeden sonra değişmektedir.

Hipergeometrik dağılımın genel ifadesi ,

$$P(X) = \frac{C_{n_1}^{k_2} C_{n_2}^{k_2}}{C_n^k} \quad ; \quad k = k_1 + k_2 \; ve \; n = n_1 + n_2$$

dir.

İlgilendiğimiz n birimden n_1 sayıdaki birim ise, p ve q

$$p = \frac{n_1}{n} \quad ve \quad q = \frac{n - n_1}{n} = \frac{n_2}{n}$$

olarak tanımlanırsa, hipergeometrik dağılımın beklenen değeri,

E(X)=kp

ve varyansı,

$$\sigma^2 = kpq\frac{n-k}{n-1}$$

olacaktır.

Hipergeometrik dağılımda eğiklik ve basıklık ölçüleri,

$$\alpha_3 = \sqrt{kpq\frac{n-1}{n-k}}.\frac{(q-p)(n-2k)}{(n-2)}$$

$$\alpha_4 = \frac{n-1}{(n-2)(n-3)}[n(n+1) - \frac{6}{pq} + \frac{3}{k(n-k)}(n^2(k-2) - nk^2 + 6k(n-k)]$$

olarak hesaplanır.

ÖRNEK: Bir kutudaki 25 mamülden 5'i hatalıdır. Bu mamüllerden 4 adedi tesadüfi ve iadesiz olarak seçilmektedir. Seçilen mamüllerden,

a- 4'ünün de hatasız olması
b- 1'inin hatalı olması olasılıklarını hesaplayınız

a-

$n = 25 \qquad k = 4$

$n_1 = 5 \qquad k_1 = 4$

$n_2 = 20 \qquad k_2 = 0$

$$P = \frac{C_5^0 C_{20}^4}{C_{25}^4} = \frac{1.(4845)}{12650} = 0.383$$

$C_{25}^4 = 12650$

$C_{20}^4 = 4845$

b- $k_1=1$

$k_2=3$

$C_{20}^3 = 1140$

$$P = \frac{C_5^1 C_{20}^3}{C_{25}^4} = \frac{5.(1140)}{12650} = \frac{5700}{12650} = 0.45$$

9.3. POİSSON DAĞILIMI

Poisson dağılımı da süreksiz bir dağılımdır. Bu dağılımda binom dağılımından farklı olarak X tesadüfi değişkeni 0 ile $+\infty$ arasında tam sayı değerler almaktadır.

Poisson dağılımı, bir işletmenin çok sayıdaki muhasebe kayıtlarında olabilecek hatalar, hatalı veya kalitesiz mal üretme olasılığı çok az olan bir makinenin üretimindeki hatalı mamüller, her gün veya belirli bir zaman biriminde pek çok uçağın inip kalktığı bir hava limanına belirli bir zaman biriminde uçak inip kalkmaması gibi olasılık değerinin çok küçük ve deneme sayısının çok büyük olduğu olaylara ait olasılıkların hesaplanmasında kullanılır. Olasılık değeri çok küçük olduğundan küçük olasılıklar dağılımı olarak adlandırılır. Binom dağılımının, P'nin küçüldüğü n'nin büyüdüğü durumlardaki yaklaşımıdır.

$$P(\lambda) = P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad ; \quad e = 2.718$$

λ olayla ilgili ortalama değerdir. Binom dağılımının beklenen değeri, poisson dağılımının parametresi (λ) karşılığıdır.

$\lambda = np$

Poisson dağılımının beklenen değeri ve varyansı birbirine eşittir.

$E(X) = V(X) = \lambda$

Dağılımın eğiklik ve basıklık ölçüleri.

$$\alpha_3 = \frac{1}{\sqrt{\lambda}}$$

$$\alpha_4 = 3 + \frac{1}{\lambda}$$

dır.

Poisson dağılımı binom dağılımının özel hali olduğundan, $n \to \infty$ ve $p<0.05$ ise $\lambda=np$ olarak hesaplanır ve binom dağılımı yerine poisson dağılımı uygulanabilir.

Uygulamada λ tam sayı olabileceği gibi kesirli de olabilir. Bu durumda $e^{-\lambda}$ logaritma yardımı ile hesaplanabileceği gibi, Mac Laurin seri açılımı ile,

$$e^{-\lambda} = 1 + \frac{(-\lambda)^1}{1!} + \frac{(-\lambda)^2}{2!} + \frac{(-\lambda)^3}{3!} + ...$$

şeklinde de hesaplanabilir.

ÖRNEK: A hava limanına her gün 11.00-12.00 arası ortalama 3 uçak inip kalkmaktadır. Bu limana 11.00-12.00 arası,

a- 5 uçak inip kalkması,
b- Uçak inip kalkmaması olasılıklarını hesaplayınız.

a- $\lambda=3$

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad ; \quad e = 2.718$$

$$P(X=5) = e^{-3}.\frac{3^5}{5!} = \frac{1}{e^3}.\frac{3^5}{1.2.3.4.5} = \frac{1}{(2.718)^3}.\frac{243}{120} = 0.1008$$

b- $$P(X=0) = e^{-3}.\frac{3^0}{0!} = \frac{1}{e^3} = \frac{1}{(2.718)^3} = 0.0498$$

9.4. NORMAL DAĞILIM

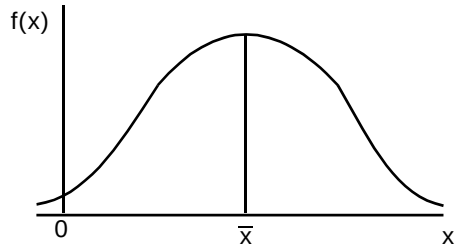
Sürekli bir dağılımdır, özellikle örnekleme ve merkez limit teoremi nedeniyle en çok kullanılan dağılımdır. Normal dağılımda olasılıkların hesaplanması diğer dağılımlara göre daha kolaydır ve belirli şartların geçerli olması ile diğer dağılımlar normal dağılıma yaklaştırılarak, çeşitli olasılık değerleri normal dağılım yardımı ile daha kolay hesaplanabilir. Bu işlemlerde sonuçların tamamen aynı olacağı beklenemez, ortaya çıkan küçük sapmalar ihmal edilebilir.

Normal dağılıma önem kazandıran bir başka nokta, daha sonra izleyeceğimiz örnekleme dağılımı, örnek parametrelerinin tahmini, hipotez testleri gibi konularda normal dağılımdan yararlanılmasıdır. Burada özellikle merkez limit teoremi önem kazanacaktır.

Normal dağılımda X tesadüfi değişkeni ± ∞ aralığında süreklidir.

$$-\infty \leq X \leq +\infty$$

X tesadüfi değişkeninin bu aralıkta göstereceği dağılımın frekans eğrisi veya olasılık yoğunluk fonksiyonu Şekil 9.1.'de görüldüğü gibi çana benzer bir şekil alır. Bu nedenle çan eğrisi olarak adlandırıldığı gibi, normal eğri, Gauss eğrisi, Laplace-Gauss eğrisi olarak da adlandırılmaktadır.



Şekil 9.1: Normal Dağılım Eğrisi

Normal dağılım fonksiyonu,

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} . e^{\frac{-(X-\bar{X})^2}{2\sigma^2}} \qquad ; \qquad -\infty \leq X \leq +\infty$$

dur. Burada X, X tesadüfi değişkeninin alacağı değeri göstermektedir. π=3.1416 ve e=2.7183 sabit değerleridir. Fonksiyondaki X ve σ ise normal dağılımın parametreleridir. İki parametreli bir dağılım olan normal dağılımın parametreleri, bilinen aritmetik ortalama ve standart sapma formülleri ile hesaplanır. Normal dağılım N($\bar{X}$, σ) şeklinde ifade edilir.

Dağılımın beklenen değeri ve varyansı,

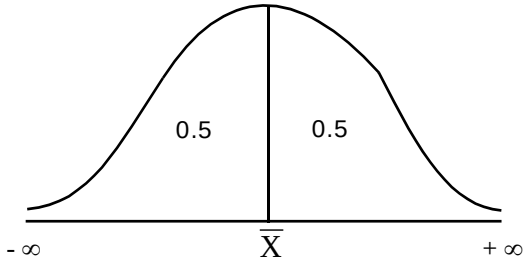
E(X) = $\bar{X}$

V(X) = σ^2 dir.

Normal dağılımın çeşitli özelliklerini şöyle özetleyebiliriz.

- Eğri ile yatay eksen (X ekseni) arasında kalan alanın toplamı 1'dir.
- Aritmetik ortalama, mod ve medyan birbirine eşittir.

  $\bar{X} = M_0 = M_e$
- Aritmetik ortalama (dolayısıyla mod ve medyan) eğrinin maksimum noktasındadır.
- Eğri aritmetik ortalamaya göre simetriktir. Aritmetik ortalamanın iki yanında kalan alanların değeri 0.5'tir



Şekil 9.2.

- Eğri yatay eksene asimtotiktir, yatay ekseni kesmez.
- Dağılımın eğiklik (asimetri) ve basıklık ölçüleri,

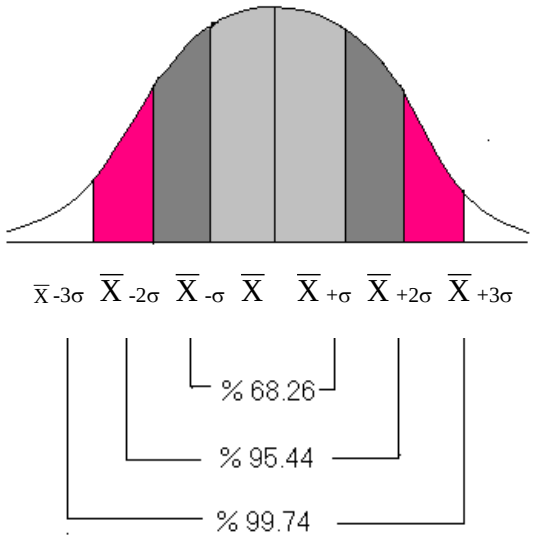
$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = 0 \qquad \text{(eğri simetrik)}$$

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = 3 \qquad \text{(orta basıklıkta)}$$

- Normal dağılımda,

$\overline{X} \pm \sigma$ aralığında terimlerin % 68.26'sı,
$\overline{X} \pm 2\sigma$ aralığında terimlerin % 95.44'ü,
$\overline{X} \pm 3\sigma$ aralığında terimlerin % 99.74'ü

bulunmaktadır veya yukarıda verilen aralıklar için olasılık değerleri verilen oranlara eşittir.



Şekil 9.3.

Olasılıkların Hesaplanması :

Normal dağılım sürekli karakterli olduğundan, X tesadüfi değişkeninin herhangi iki değer arasında olması olasılığını hesaplamak için fonksiyonun bu aralık içinde belirli integralini almak gerekir.

$$P(X_1 \le X \le X_2) = \int_{X_1}^{X_2} f(X)dx = \int_{X_1}^{X_2} \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{(X-\overline{X})^2}{2\sigma^2}} dx$$

Fonksiyonun ± ∞ aralığında belirli integrali, olasılıklar toplamını verir ve bire eşittir.

$$P(-\infty \le X \le +\infty) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{(X-\overline{X})^2}{2\sigma^2}} dx = 1$$

Normal dağılımın parametrelerinin dağılımından dağılıma değişeceği ve çeşitli olasılık değerleri için bu integralin alınmasının gerektiği düşünülürse, alınması güç olur. Bu integral olasılık hesaplamanın zorluğu anlaşılır. Bu güçlüğü gidermek için seri standartlaştırılır. Standartlaştırma işlemi için,

$$Z = \frac{X - \overline{X}}{\sigma}$$

dönüşümü yapılır. Bulunan Z serisinin aritmetik ortalaması 0 ve varyansı 1'dir.

$\overline{Z}$ = 0

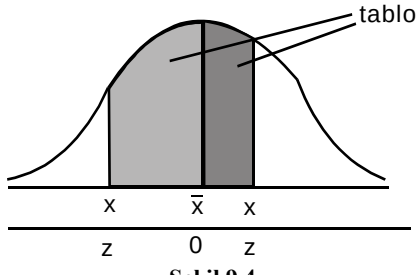
V(Z) =1

Bu durumda normal dağılım fonksiyonu,

$$f(X) = \frac{1}{\sqrt{2\pi}} . e^{-\frac{Z^2}{2}}$$

olacaktır. Bu fonksiyonun ± ∞ aralığında belirli integrali alındığında sonuç 1'e eşit olur.

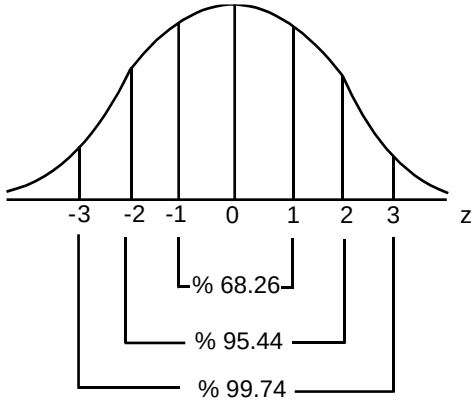
Standart Z değişkeninin değerleri için integraller hesaplanarak çeşitli tablolar düzenlenmiştir. Burada kullanacağımız tablo X ve $\overline{X}$ veya (Z ile $\overline{Z}$) arasındaki alanı olasılıkları verir.



Şekil 9.4

Hesaplanacak olasılık için tablo değerine göre işlemler yapılır. Tablo ile yapılacak işlemler örneklerle açıklanacaktır.

Bu şekilde standartlaştırılan bir dağılım içinde daha önce açıklanan standart aralık olasılıkları geçerli olacaktır.



Şekil 9.5.

ÖRNEK: Bir makinenin A parçalarının ortalama kullanım süresi 2600 saat ve kullanım sürelerinin standart sapması 50 saat olarak belirlenmiştir. Dağılım normal olduğuna göre seçilen bir parçanın,

a- 2500 saat ile 2650 saat arası,
b- 2415 saatten çok,
c- 2640 saatten çok,
d- 2580 saatten az,
e- 2676 saatten az kullanılması olasılıklarını hesaplayınız.

a- P(2500 < X < 2650)

$$Z_1 = \frac{X_1 - \overline{X}}{\sigma} = \frac{2500 - 2600}{50} = \frac{-100}{50} = -2$$

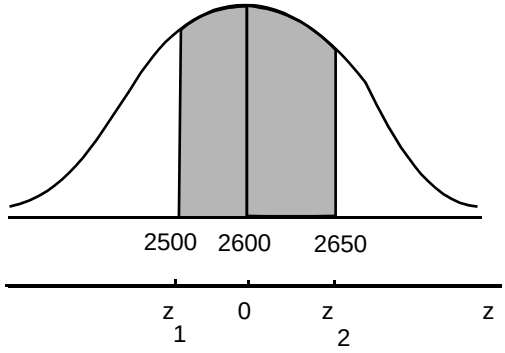
Tablodan 2'nin değeri $\rightarrow$ 0.4773

$$Z_2 = \frac{X_2 - \overline{X}}{\sigma} = \frac{2650 - 2600}{50} = \frac{50}{50} = 1$$

Tablodan 1'nin değeri $\rightarrow$ 0.3413

P(2500 < X < 2650) = 0.4773 + 0.3413 = 0.8186

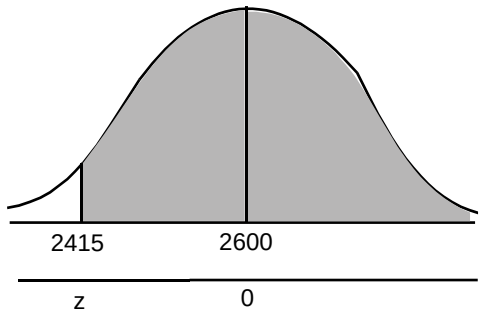
Not: Burada bulunan Z değerlerinin + veya – olması tablo değerinin değiştirmez.



b- P(X>2415)

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{2415 - 2600}{50} = -\frac{185}{50} = -3{,}7$$

Z(3,7) $\rightarrow$ 0.4999



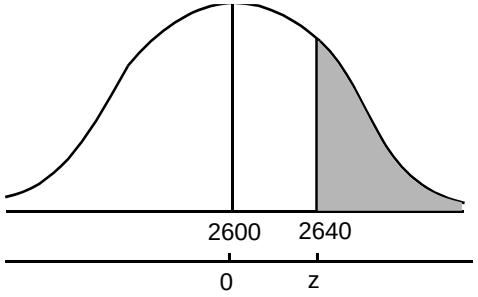
P(X>2415) = 0.5 + 0.4999 = 0.9999

c- P(X>2640)

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{2640 - 2600}{50} = \frac{40}{50} = 0.8$$

Z(0,8) $\rightarrow$ 0.2881

P(X>2640) = 0.5 - 0.2881 = 0.2119

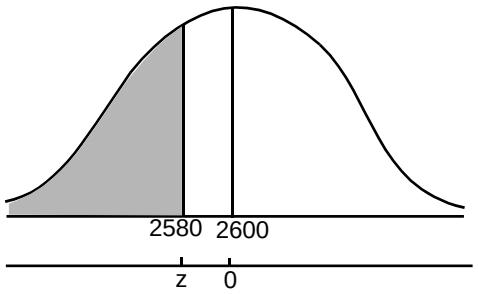


d- P(X<2580)

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{2580 - 2600}{50} = -\frac{20}{50} = -0.4$$

Z(0,4) $\rightarrow$ 0.1554

P(X<2580) = 0.5 - 0.1554 = 0.3446

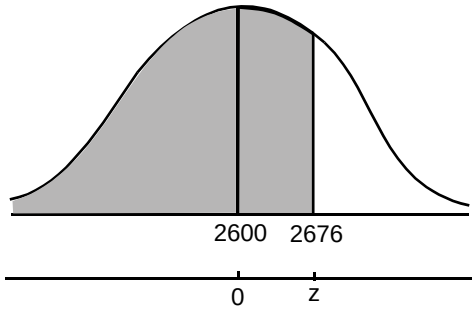


e- P(X<2676)

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{2676 - 2600}{50} = \frac{76}{50} = 1.52$$

Z(1,52) $\rightarrow$ 0.4357

P(X<2676) = 0.5 + 0.4357 = 0.9357



Not: Hesaplamada < ile ≤ ve > ile ≥ arasında fark yoktur.

9.4.1. Normalliğin İncelenmesi

Uygulamada bir dağılımın normal olup olmadığının incelenmesi gerekebilir. Ayrıca yapılan araştırmaların özelliklerine bağlı olmak şartı ile, normale yakın dağılımlar normal kabul edilebilirler. Bu durumda gerçek değerlerden sapmalar olacağından bunlar da göz önünde tutulmalıdır.

Burada normalliğin incelenmesi için, sadece incelenecek dağılımın normal dağılımın bazı özelliklerine sahip olup olmadığı araştırılacaktır.

-Normal eğri çana benzer olduğundan, incelenecek serinin grafiği çizilerek frekans eğrisinin şekline bakılabilir. Eğri, normal eğriye benzer ise kabaca dağılım normal kabul edilebilir. Grafiğin ölçeklemesinin değiştirilmesi ile eğri şekil değiştirebileceğinden verilen kararlar yanıltıcı olabilir. Dağılımın normale benzerliği frekansların artarak gidip serinin orta değerlerinde yoğunlaştıktan sonra azalmaya başlaması ile de anlaşılabilir.

-Normal dağılımın özelliklerinden biri de, aritmetik ortalama, mod ve medyanının birbirine eşit olmasıdır. Bu nedenle serinin aritmetik ortalaması, modu ve medyanı hesaplanarak birbirlerine yakın olmaları ölçüsünde dağılım normal kabul edilebilir.

-Normal dağılımda standart aralıklarda seri terimleri belirli oranlarda bulunacağından, bu aralıklar ($\overline{X} \pm Z\sigma$) yeralan terimlerin oransal paylarının hesaplanması ve standart oranlarla karşılaştırılması ile dağılımın normalliği hakkında karar verilebilir.

-Buraya kadar açıklanan yollardan daha kesin sonucu, momentlere dayanan eğiklik (asimetri) ve basıklık ölçülerinin hesaplanması verecektir. Normal dağılımda eğiklik ölçüsü $\alpha_3=0$ ve basıklık ölçüsü $\alpha_4=3$'tür. Hesaplanacak sonuçların bu değerlere yakınlığı ölçüsünde dağılım normale yakın olacaktır diyebiliriz.

ÖRNEK: Aşağıda bir seri verilmiştir.

Sınıflar	f_i	X_i	f_iX_i	Σf_i	$f_iX_i^2$
50-54	15	52	780	15	40.560
54-58	40	56	2240	55	125.440
58-62	75	60	4500	130	270.000
62-66	45	64	2880	175	184.320
66-70	25	68	1700	200	115.600
	200		12100		735.920

Bu serinin normalliğini,

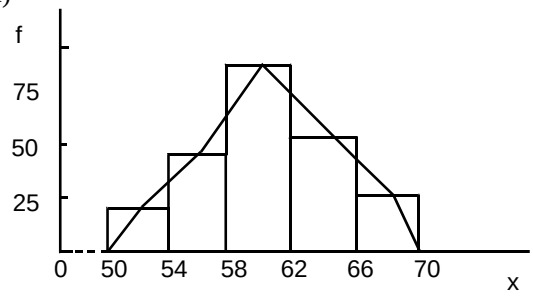
a) Grafik,

b) Aritmetik ortalama, mod ve medyan,

c) $\overline{X}$ +2σ aralığında bulunan terimler,

d) Momentlere dayanan eğiklik ve basıklık ölçüleri yardımı ile inceleyiniz

a)



Frekans eğrisinin normal eğriye benzediği görülmektedir. Dağılım normale yakındır.

b) $\overline{X} = \dfrac{\Sigma f_i X_i}{\Sigma f_i} = \dfrac{12100}{200} = 60.5$

$$M_0 = X_0 + i.\frac{\Delta_1}{\Delta_1 + \Delta_2}$$

$$M_0 = 58 + 4.\frac{35}{35+30} = 58 + \frac{140}{65} = 58 + 2{,}15 = 60{,}15$$

$$M_e = X_0 + i.\frac{N/2 - \Sigma f_{i-1}}{f_0} \quad ; \quad N/2 = 200/2 = 100$$

$$M_e = 58 + 4.\frac{100-55}{75} = 58 + \frac{180}{75} = 58 + 2.4 = 60.4$$

$\overline{X} \cong M_0 \cong M_e$

$60{,}5 \cong 60{,}15 \cong 60{,}4$

olduğundan dağılım normale yakındır.

c) $\overline{X} \pm 2\sigma$ aralığında normal dağılımda mevcut terimlerin 0.9544'ü bulunacaktır.

$$\sigma = \sqrt{\frac{\Sigma f_i X_i^2}{\Sigma f_i} - (\overline{X})^2} = \sqrt{\frac{735920}{200} - (60{,}5)^2}$$

$$= \sqrt{3679{,}6 - 3660{,}25} = 4.399 \cong 4{,}4$$

$\overline{X}$ - 2σ = 60,5 – 2. (4,4) = 51,7
$\overline{X}$ + 2σ = 60,5 + 2. (4,4) = 69,3

	Fark	Frekans
50 – 54 aralığında	4	15
51,7 – 54 aralığında	2,3	X
	X=8,625	
66 – 70 aralığında	4	25
66 – 69,3 aralığında	3,3	X
	X=20,625	

Sınıflar	fi
51,7 – 54	8,625
54 - 58	40
58 - 62	75
62 – 66	45
66 –69.3	20,625
	189,25

Toplam Oranı = 189,25/200= 0,94625

$\overline{X} \pm 2\sigma$ aralığında terimlerin %95,44'ü yerine %94,63'ü bulunmaktadır. Fark %1'den azdır ve dağılım normale yakındır, normal kabul edilebilir.

d) $\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$

$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$

Daha önce $\sigma^2 = \mu_2 = 19.35$

$f_i(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^4$
-9211.875	78300,9375
-3645	16402,5
-9.375	4,6875
1929.375	6752,8125
10546.875	79101,5625
-390.0	180562,5

$$\mu_3 = \frac{\Sigma f_i (X_i - \overline{X})^3}{\Sigma f_i} = \frac{-390}{200} = -1,95$$

$$\mu_4 = \frac{\Sigma f_i (X_i - \overline{X})^4}{\Sigma f_i} = \frac{180562,5}{200} = 902,8125$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{-1,95}{(4,39)^3} = \frac{-1,95}{84,6} = -0,02$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{902,8125}{(4,4)^4} = \frac{902,8125}{374,81} = 2,41$$

α_3 sıfıra yakın olduğundan seri simetrik kabul edilebilir. α_4 3'ten oldukça küçüktür, seri basıktır. Dağılımın normal kabul edilmesi araştırmanın niteliklerine ve araştırmacıya bağlıdır. Meydana gelecek sapmalar ihmal edilebiliyorsa, dağılım normal kabul edilebilir.

9.4.2. Binom Dağılımın Normal Dağılıma Yaklaştırılması

Binom dağılımında deneme sayısı (n) küçük ise, verilen formülle çeşitli olasılık değerlerinin bulunması kolay olacaktır. n büyüdükçe hesaplama güçleşeceğinden binom dağılımı yerine normal dağılımdan yararlanılabilir. Bu durumda normal dağılımla hesaplanan olasılık değerleri, binom dağılımı ile hesaplanan değerlerden sapabilir. Hesaplama kolaylığı yanında sapmalara katlanmak gerekir.

Binom dağılımının normal dağılıma yaklaştırılması işleminde normal dağılım parametreleri binom dağılımından hesaplanacaktır.

$$B(n,p) \rightarrow N(\overline{X}, \sigma)$$

Binom dağılımında

$$B(X)=np \text{ ve } \sigma = \sqrt{npq}$$

olduğundan,

$$B(n, p) \cong N(np, \sqrt{npq})$$

olacaktır. Normal dağılımda olasılık değerlerinin bulunmasında yararlanılan standart değişken Z ise,

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{X - np}{\sqrt{npq}}$$

şekline dönüşecektir.

Binom dağılımı süreksiz, normal dağılım ise sürekli olduğundan yaklaştırma ile olasılık hesaplanırken, X tesadüfi değişkeninin değerine dikkat etmek gerekir. Normal dağılımda > ile ≥ ve < ile ≤ arasında fark olmadığı halde binom dağılımı için aynı şeyi söyleyemeyiz. Normal dağılımda X tesadüfi değişkeninin herhangi bir değere eşit olması olasılığı suni alan oluşturarak hesaplandığından, X'in alacağı değere eşit olup olmadığa bakılarak bu değere 0,5 ilave edilir veya 0,5 çıkarılır. Örneğin, hesaplanacak herhangi bir olasılık Binom dağılımı için P(25<X ≤ 32) ise, bu normal dağılım için P(25,5<X < 32,5); binom dağılımı için P(18 ≤ X <35) ise bu normal dağılım için P(17,5<X < 34,5) olacaktır.

Binom dağılımında n>50 ise dağılım normale yaklaştırılabilir. Bu durumda p'nin alacağı değerlere bağlı olarak hesaplanacak olasılık değerlerinde farklı sapmalar olabilir. Binom dağılımını normale en uygun yaklaştırma şartı P≤0,5 olmak şartıyla np≥5'tir. Binom dağılımında olasılık 0,5 veya daha küçük ve beklenen değer 5 veya daha büyükse normal dağılıma yaklaştırma işlemi ile elde edilecek olasılıklarla gerçek olasılıklar arasındaki sapma küçük olacaktır. Yaklaştırma işlemini bir örnekle açıklayalım.

ÖRNEK: Parametreleri n=60 ve p=0.20 olan binom dağılımı ile ilgili aşağıdaki olasılıkları mümkün ise normal dağılıma yaklaştırma ile hesaplayınız.

a)P(X<8)

b)P(X ≤15)

c)P(X>12)

d)P(6≤ X <10)

e)P(10< X ≤ 14)

n>50 olduğundan bu binom dağılımı normale yaklaştırılabilir. (Ayrıca bu dağılımda P<0,5 ve np=60x0,2 =12>5 şartı da sağlanmaktadır. Bu dağılımda n=60>50 iken p=0,05 olsaydı np=0,05x60=3<5 olacağından uygun yaklaştırma şartı sağlanamayacaktı.)

Binom dağılımında,

E(X) = np= 60x0,2=12

V(X) = npq= 60x0,2x0,8=9,6

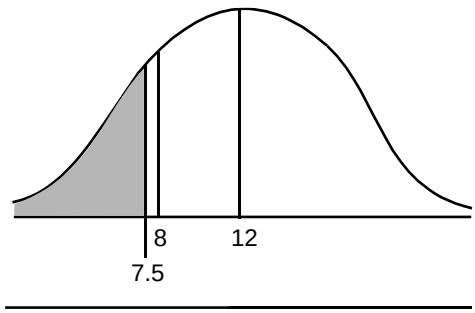
$\sigma = \sqrt{npq} = \sqrt{9,6} = 3,098$

B(n;p) = B(60;0,2) $\rightarrow$ N($\overline{X}$;σ) = N(12;3,09) şeklinde yaklaştırma işlemi yapılabilir ve standart değişken,

$$Z = \frac{X - \overline{X}}{\sigma} = \frac{X - np}{\sqrt{npq}}$$

olacaktır. Süreksiz dağılım sürekli dağılıma yaklaştırıldığından X'in alacağı değerlere 0,5 eklenecek veya çıkarılacaktır.

a-

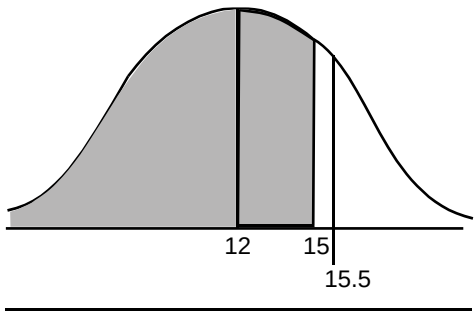


$$Z = \frac{7,5 - 12}{3,09} = -1,45 \rightarrow 0,4265$$

Binom ile hesaplanacak olasılıkları P, normale yaklaştırılarak hesaplananları P' ile gösterirsek,

P(X<8) = P'(X<7,5) = 0,5-0,4265 = 0,0735

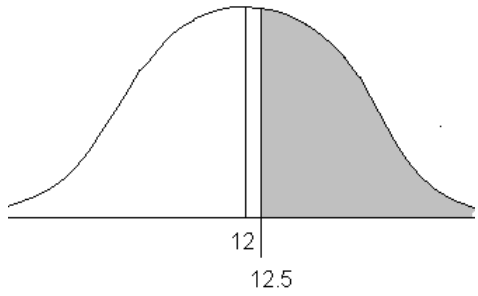
b-



$$Z = \frac{15,5 - 12}{3,09} = 1,13 \rightarrow 0,3718$$

P(X ≤ 15) = P'(X< 15.5) = 0,5 + 0,3718 = 0,8718

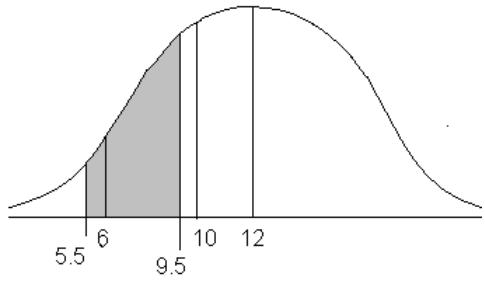
c-



$$Z = \frac{12.5 - 12}{3.09} = 0.16 \rightarrow 0.0636$$

P(X>12) = P'(X>12,5) = 0,5 – 0,0636 = 0,4364

d-

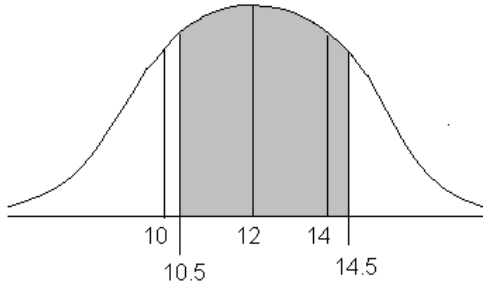


$$Z_1 = \frac{5{,}5 - 12}{3{,}09} = -2{,}10 = 0{,}4821$$

$$Z_2 = \frac{9{,}5 - 12}{3{,}09} = -0{,}80 = 0{,}2881$$

P(6≤ X>10) = P'(5,5< X<9,5) = 0,4821 – 0,2881 = 0,1940

e-



$$Z_1 = \frac{10{,}5 - 12}{3{,}09} = -0{,}48 \rightarrow 0{,}1844$$

$$Z_2 = \frac{14{,}5 - 12}{3{,}09} = 0{,}80 \rightarrow 0{,}2881$$

$$P(10 < X \leq 14) = P'(10{,}5 < X < 14{,}5) = 0{,}1844 + 0{,}2881 = 0{,}4725$$

9.5. DİĞER DAĞILIMLAR

Buraya kadar olasılıkların hesaplanması için kullanılan olasılık dağılımlarının başlıcaları açıklandı. Giriş bölümünde ise örneklemeden sözedilmişti. Örnekten elde edilen tahminler için daha sonra açıklanacak bazı işlemlerin yapılabilmesi ile ilgilidir. Burada bu dağılımların temel özellikleri açıklanacaktır.

9.5.1. Ki-kare Dağılımı

Normal dağılım açıklanırken, standart Z değişkeninden sözedilmişti. Standart Z değişkeni X tesadüfi değişkeni normal dağılıma sahip ise standart değişkenin karesinin (Z^2) dağılımı bir serbestlik dereceli Gamma dağılımının özel bir şekli olan χ^2 (ki-kare) dağılımıdır.

$X \sim N(\mu,\sigma)$

$$Z_i = \frac{X_i - \mu}{\sigma}$$

$Z_i \sim N(0,1)$

$Z_i^2 \sim \chi^2_{\nu=1}$

Serbestlik derecesi ν olan χ^2 dağılımlı bir tesadüfi değişkeninin olasılık yoğunluk fonksiyonu,

$$f(X) = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} X^{\frac{\nu-2}{2}} e^{-\frac{X}{2}} \qquad X > 0$$

dır.

Normal dağılmış bağımsız X_1, X_2,....., X_n tesadüfi değişkenleri için standart değişkenler Z_1, Z_2,....., Z_n ise, standart değişkenlerin karelerinin toplamının dağılımı,

$$Y = Z_1^2 + Z_2^2 + Z_3^2 + ... + Z_n^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

ν=n serbestlik dereceli ki-kare dağılımıdır. Standart Z_1, Z_2,....., Z_n değişkenlerinin karelerinin dağılımları sırası ile ν_1, ν_2,....., ν_n serbestlik dereceli ki-kare dağılımları ise, standart değişkenlerin karelerinin toplamı (ν_1, ν_2,....., ν_n) serbestlik dereceli ki-kare dağılımı olacaktır.

Normal dağılmış ortalaması μ ve standart sapması σ olan bir anakütleden tesadüfi olarak seçilmiş n birimli örneğin ortalaması $\overline{X}$ ve varyansı S^2 ise, ve S^2'nin bağımsız olması durumunda,

$$\frac{(n-1)S^2}{\sigma^2}$$

dağılımı, (n-1) serbestlik dereceli ki-kare dağılımıdır.

9.5.2. t - Dağılımı

Standart normal dağılmış Z değişkeni ile ν serbestlik dereceli ki-kare dağılımlı Y değişkeni bağımsız değişkenler ise,

$$T = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$$

dağılımı (n-1) serbestlik dereceli t dağılımıdır. Dağılımın olasılık yoğunluk fonksiyonu,

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad -\infty < t < +\infty$$

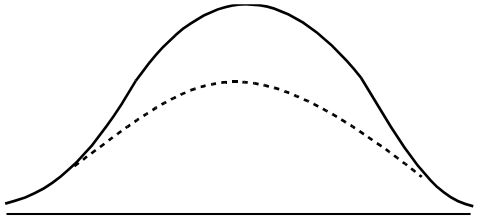
dur. Bu ν serbestlik dereceli t dağılımıdır.

Student-t dağılımı adı verilen bu dağılım anakütle varyansının bilinmediği durumlarda küçük örnekler için kullanılmaktadır. Ortalaması μ olan ve varyansı σ^2 bilinmeyen normal dağılmış bir anakütleden alınan n birimli bir örnek için,

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

dağılımı (n-1) serbestlik dereceli t dağılımıdır.

t dağılımı normal dağılım gibi $\pm\infty$ aralığında sürekli simetrik ve tek modlu bir dağılımdır.



Şekil 9.6.

t dağılımı normal dağılımdan daha basıktır, fakat serbestlik derecesinin artması ile normal dağılıma yaklaşır.

9.5.3. F Dağılımı

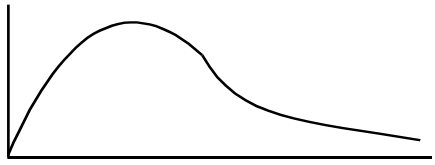
Serbestlik dereceleri ν_1 ve ν_2 olan ki-kare dağılımlı iki birbirinden bağımsız tesadüfi değişken U ve V ise bunların oranı,

$$F = \frac{U/\nu_1}{V/\nu_2}$$

ν_1 ve ν_2 serbestlik dereceli F dağılımıdır. Olasılık yoğunluk fonksiyonu,

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right).\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\frac{\nu_2}{2}-1}\left(1+\frac{\nu_1}{\nu_2}f\right)^{-\left(\frac{\nu_1+\nu_2}{2}\right)} \qquad f>0$$

dır. F dağılımı sağa eğik bir dağılımdır.



Şekil 9.7.

Serbestlik dereceleri büyüdükçe F dağılımı normal dağılıma yaklaşır. F değeri daima pozitiftir.

9.6. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir torbada birbirinin benzeri 4 beyaz, 6 siyah, 3 kırmızı top vardır. Torbadan tesadüfi ve iadeli seçimle iki top seçilmiştir.

a)İki topun da beyaz olması,
b)Bir kırmızı top çekilmesi,
c)Siyah top çekilmemesi olasılıklarını hesaplayınız.

Çözüm:

a) $P(B)=\frac{n}{N}=\frac{4}{13}$ P(X=k) = $C_n^k p^k q^{n-1}$

$P(K)=\frac{3}{13}$ n=2 (çekiliş sayısı)

$P(S)=\frac{6}{13}$

$p=\frac{4}{13}\rightarrow q=1-p=1-\frac{4}{13}=\frac{9}{13}$

$$P(X_{BB}=2)=C_2^2\left(\frac{4}{13}\right)^2\left(\frac{9}{13}\right)^0=1.\frac{16}{169}.1=\frac{16}{169}=0{,}0947$$

Aynı olasılığı daha önce hesapladığımız şekilde şöyle bulabiliriz.

$$P(B,\overline{B})=\frac{n}{N}.\frac{n}{N}=\frac{4}{13}.\frac{4}{13}=\frac{16}{169}=0{,}0947$$

$$\text{b) } P(X_K=1)=C_2^1\left(\frac{3}{13}\right)^1\left(\frac{10}{13}\right)^1=2.\frac{3}{13}.\frac{10}{13}=\frac{60}{169}=0.3550$$

veya

$$P(K,\overline{K})=\frac{n_1}{N}.\frac{n_2}{N}=\frac{3}{13}.\frac{10}{13}=\frac{30}{169}=0{,}1775$$

bu sıralı olduğundan ve 2 farklı diziliş söz konusu olduğundan,

$$P(K,\overline{K})=2.(K,\overline{K})=2.\frac{30}{169}=\frac{60}{169}=0{,}3550$$

bulunur.

$$\text{c) } P(X_S=0)=C_2^0\left(\frac{6}{13}\right)^0\left(\frac{7}{13}\right)^2=1.1.\frac{49}{169}=\frac{49}{169}=0{,}2899$$

veya

$$P(\overline{S},\overline{S})=\frac{7}{13}.\frac{7}{13}=\frac{49}{169}=0{,}2899$$

görüldüğü gibi binom dağılımı ile aynı sonuç elde edilmekteyse de, binom dağılımı ile hesaplama daha kısıtlıdır. Örneğin, bu örnekte bir siyah, bir kırmızı top çekilmesi olasılığı binom dağılımı ile hesaplanamaz. Binom dağılımından sadece bir karakterin olumlu ve olumsuz durumları için yararlanılabilir.

ÖRNEK 2: A fabrikasında üretilen yedek parçaların 0,20'si istenen niteliklere sahip değildir, kalitesizdir. Bu parçalardan tesadüfi olarak 6 adet alınmıştır.

a) Kalitesiz parçaların olasılık dağılımını belirleyin.
b) Kalitesiz parçaların olasılık dağılımının beklenen değerini ve standart sapmasını hesaplayın.
c) Bütün yedek parçaların kalitesiz olması olasılığını,
d) Bütün yedek parçaların kaliteli olması olasılığını,
e) 2 yedek parçanın kalitesiz olması olasılığını,
f) 1'le 4 arası yedek parçanın kalitesiz olması olasılığını hesaplayınız.

Çözüm:

a) n=6, p=0,20 >0,05; parçalar kaliteli veya kalitesiz; tekrarlar bağımsız; p ve q her tekrarda aynı.

Binom Dağılımı: B(n;p) = P(X=k) = $C_n^k p^k q^{n-k}$

b) $p = 0{,}2$

$n = 6$

$q = 1 - p = 0{,}8$

$E(X) = np = 6.(0{,}2) = 1{,}2$

$V(X) = npq = 6.(0{,}2).(0{,}8) = 0{,}96$

$\sigma = \sqrt{npq} = \sqrt{0{,}96} = 0{,}98$

c) $P(X = 6) = C_6^6 (0{,}2)^6 (0{,}8)^0 = (0{,}2)^6 = 0{,}000064$

d) $P(X = 0) = C_6^0 (0{,}2)^0 (0{,}8)^6 = 1.1.(0{,}8)^6 = 0{,}262144$

veya p ile q yer değiştirirse,

$P'(X = 6) = C_6^6 (0{,}8)^6 (0{,}2)^0 = (0{,}8)^6 = 0{,}262144$

e) $P(X = 2) = C_6^2 (0{,}2)^2 (0{,}8)^4 = 15.(0{,}2)^2 (0{,}8)^4 = 0{,}24576$

f) Burada farklı durumlar sözkonusu olabilir.

-Limitler dahil,

$$P(1 \le X \le 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$
$$= C_6^1(0,2)^1(0,8)^5 + C_6^2(0,2)^2(0,8)^4 + C_6^3(0,2)^3(0,8)^3 + C_6^4(0,2)^4(0,8)^2$$
$$= 0,393216 + 0,245760 + 0,08192 + 0,015360 = 0,736256$$

-1 dahil, 4 hariç

$$P(1 \le X < 4) = P(X = 1) + P(X = 2) + P(X = 3)$$
$$= 0,393216 + 0,245760 + 0,08192$$
$$= 0,720896$$

-1 hariç, 4 dahil

$$P(1 < X \le 4) = P(X = 2) + P(X = 3) + P(X = 4)$$
$$= 0,245760 + 0,08192 + 0,01536$$
$$= 0,34304$$

-Limitler hariç

$$P(1 < X < 4) = P(X = 2) + P(X = 3)$$
$$= 0,245760 + 0,08192$$
$$= 0,32768$$

ÖRNEK 3: Bir fabrikada imal edilen ampüller 0.001 olasılıkla bozuktur. Bir başka fabrikanın aydınlatılması için bu ampüllerden tesadüfi olarak 2000 adet alınmıştır.

a) Bozuk ampüllerin olasılık dağılımını belirleyerek dağılımın beklenen değerini ve varyansını hesaplayınız.
b) 4 ampülün bozuk olması,
c) 3'den az ampülün bozuk olması,
d) En az 2 ampülün bozuk olması,
e) En çok 5 ampülün bozuk olması,
f) 5'ten çok ampülün bozuk olması,
g) Bütün ampüllerin bozuk olması olasılıklarını hesaplayınız.

Çözüm:

a) $P(X = k) = e^{-\lambda}.\frac{\lambda^k}{k!}$

$\lambda = np = 2000.(0,00\ 1) = 2$

$E(X) = V(X) = \lambda = 2$

b) $P(X = 4) = e^{-2}\frac{2^4}{4!} = \frac{1}{e^2}.\frac{16}{1.2.3.4} = 0,09$

c) $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= e^{-2}.\frac{2^0}{0!} + e^{-2}.\frac{2^1}{1!} + e^{-2}.\frac{2^2}{2!}$$

$$= \frac{1}{e^2}[1+2+2] = 0.6767$$

d) $P(X \geq 2) = P(X=2) + P(X=3) + ... + P(X=1959) + P(X=2000)$

$$= 1 - \mathrm{P(X<2)}$$

$$= 1 - [\mathrm{P(X=0)} + \mathrm{P(X=1)}]$$

$$= 1 - (\mathrm{e}^{-2}.\frac{2^0}{0!} + \mathrm{e}^{-2}.\frac{2^1}{1!})$$

$$= 1 - 0{,}406 = 0{,}594$$

e) $P(X \leq 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$$= \mathrm{e}^{-2}.\frac{2^0}{0!} + \mathrm{e}^{-2}.\frac{2^1}{1!} + \mathrm{e}^{-2}.\frac{2^2}{2!} + \mathrm{e}^{-2}.\frac{2^3}{3!} + \mathrm{e}^{-2}.\frac{2^4}{4!} + \mathrm{e}^{-2}.\frac{2^5}{5!}$$

$$= \frac{1}{\mathrm{e}^2}[1+2+2+\frac{8}{6}+\frac{16}{24}+\frac{32}{120}] = 0{,}9834$$

f) $P(X > 5) = P(X=6) + ... + P(X=2000)$

$$= 1 - P(X \leq 5)$$

$$= 1 - 0{,}9834 = 0{,}0166$$

g) $P(X=2000) = e^{-2}.\frac{2^{2000}}{2000!} = 0.0000$

ÖRNEK 4: Bir fabrikada imal edilen vazolar üzerinde bulunan hata sayısı incelenmektedir. Bu vazolardan 510 adet alınmış ve toplam hata sayısı 1275 olarak bulunmuştur. Dağılımın poisson dağılımı olduğu bilinmektedir.

a) Dağılımın parametresini hesaplayınız.
b) Tesadüfi olarak seçilen bir vazo üzerinde 2 hata bulunması olasılığını hesaplayınız.
c) Tesadüfi olarak seçilen bir vazo üzerinde birden fazla hata olması olasılığını hesaplayınız.

Çözüm:

a) $\overline{\mathrm{X}} = \lambda = \frac{\Sigma \mathrm{X_i}}{\mathrm{n}} = \frac{1275}{510} = 2{,}5$

b) $P(X=2) = e^{-2.5}.\frac{2{,}5^2}{2} = \frac{1}{e^{2.5}}.\frac{6{,}25}{2} = 0{,}256$

$e^{-2.5} = 0{,}0821$

c) $P(X>1) = P(X=2) + P(X=3) + ...$

$$= 1 - P(X \le 1)$$
$$= 1 - [P(X = 0) + P(X = 1)]$$
$$= 1 - (e^{-2,5} \cdot \frac{(2,5)^0}{0!} + e^{-2,5} \cdot \frac{(2,5)^1}{1!})$$
$$= 1 - 0,2873 = 0,7127$$

ÖRNEK 5: Bir fabrikada üretilen mamüllerin ortalama ağırlıkları 50 kg. ve ağırlıkların dağılımının varyansı 49 saat olarak bulunmuştur. Ağırlıkların dağılımı normaldir ve bu mamüllerden 5000 adet alınmıştır. Bu 5000 mamülden kaç adedinin,

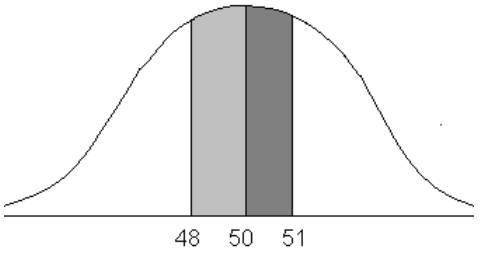
a)48 ile 51 kg. arasında,
b)55 ile 60 kg. arasında,
c)45 kg.'a eşit,
d)50 kg.'a eşit,
e)85 kg.'dan az,
f)80 kg.'dan fazla olacağını hesaplayınız.

Çözüm:

$\overline{X} = 50$

$\sigma = \sqrt{49} = 7$

a) $P(48 < X < 51)$



$$Z_1 = \frac{48 - 50}{7} = \frac{-2}{7} = -0,29 \rightarrow 0,1141$$

$$Z_2 = \frac{51 - 50}{7} = \frac{1}{7} = 0,14 \rightarrow 0,0557$$

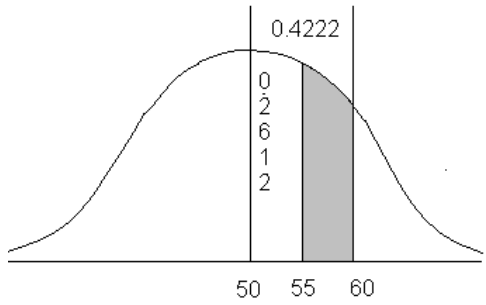
$$P(48 < X < 51) = 0,1141 + 0,0557 = 0,1698$$

Ağırlığı 48-51 kg. arasında olan mamül sayısını (teorik sayı) bulmak için hesaplanan olasılık ile verilen mamül sayısı çarpılacaktır.

n = N.p
n = 5000x0,1698 = 849

849 adet mamülün ağırlığı teorik olarak 48 kg. ile 51 kg. arasında olacaktır.

b) P(55 < X < 60)



$$Z_1 = \frac{X - \overline{X}}{\sigma} = \frac{55 - 50}{7} = \frac{5}{7} = 0{,}71 \rightarrow 0{,}2612$$

$$Z_2 = \frac{60 - 50}{7} = \frac{10}{7} = 1{,}43 \rightarrow 0{,}4236$$

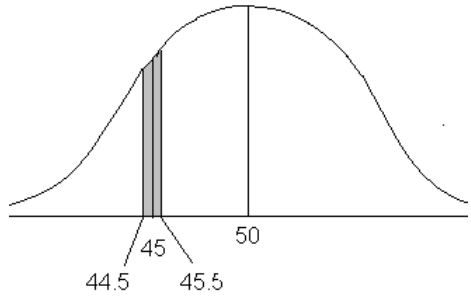
P(55 < X < 60) = 0,4236 – 0,2612 = 0,1624

n= Np = 5000x0,1624 = 812 adet

c) P(X=45)

Bu olasılık değeri iki farklı şekilde hesaplanabilirse de burada sadece suni olan oluşturularak hesaplama yapılacaktır. Diğer yol için farklı tablo gerekmektedir.

44,5 ile 45,5 arası 45'e eşit kabul edilerek hesaplama yapılır. (verilen değerin ½ birim eksiği ile ½ birim fazlası arası = verilen değer)



$$Z_1 = \frac{44{,}5 - 50}{7} = \frac{-5{,}5}{7} = -0{,}79 \rightarrow 0{,}2852$$

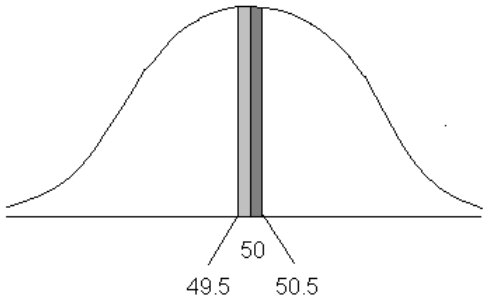
$$Z_2 = \frac{45{,}5 - 50}{7} = \frac{-4{,}5}{7} = -0{,}64 \rightarrow 0{,}2389$$

P(X=45) = 0,2852 – 0,2389 = 0,0463

n = Np = 5000x 00463

= 231,5 ≈ 232 adet

d) P(X=50)



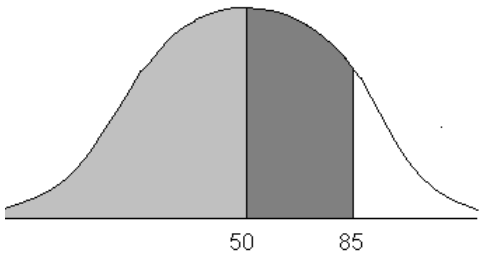
$$Z_1 = \frac{49,5 - 50}{7} = \frac{-0,5}{7} = 0,07 \rightarrow 0,0279$$

$$Z_2 = \frac{50,5 - 50}{7} = \frac{0,5}{7} = 0,07 \rightarrow 0,0279$$

P(X=50) = 0,0279 + 0,0279 = 0,0558

n = Np = 5000x 0,0558 = 279 adet

e) P(X< 85)



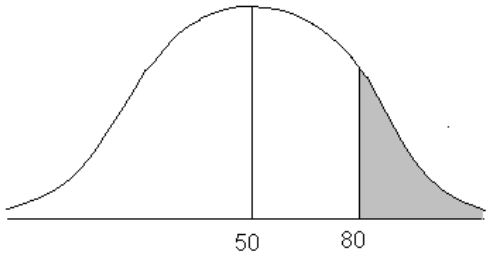
$$Z = \frac{85 - 50}{7} = \frac{35}{7} = 5 \rightarrow 0,5000$$

(Z değeri tabloda yoktur. En büyük tablo değerinden büyük olduğundan değeri 0,5'tir.)

P(X<85) = 0.5 + 0.5 =1

n = Np = 5000x 1 = 5000 adet

g) P(X>80)



$$Z = \frac{80-50}{7} = \frac{30}{7} = 4{,}29 \rightarrow 0{,}5$$

P(X>80) = 0,5 – 0,5 = 0
n = Np = 5000x 0 = 0 adet

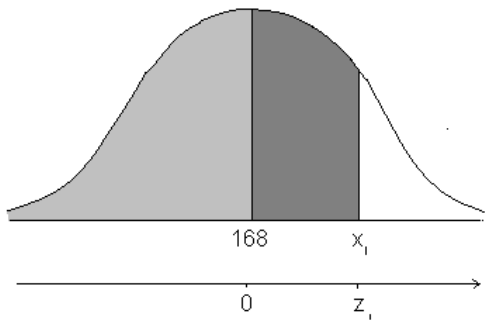
ÖRNEK 6: Bir normal dağılımın parametreleri; $\overline{X}$ = 168 ve σ=12'dir.

a) $P(X<X_1) = 0{,}8264$ olması için X_1 değerini,
b) $P(X<X_2) = 0{,}33$ olması için X_2 değerini,
c) $P(X>X_3) = 0{,}5596$ olması için X_3 değerini,
d) $P(X>X_4) = 0{,}1056$ olması için X_4 değerini,
e) $P(X_5<X<X_6) = 0{,}5$ (simetrik) olması için X_5 ve X_6

değerlerini,

f) $P(X<X_7) = 0{,}4207$ ve
$P(X_7<X<X_8) = 0{,}3979$ olması için X_7 ve X_8 değerlerini,
g) $P(160<X<X_9) = 0{,}1$ olması için X_9 değerini,
h) $P(X_{10}<X<180) = 0{,}6$ olması için X_{10} değerini hesaplayınız.

Çözüm:
a) $P(X<X_1) = 0{,}8264$
b)



$$Z_1 = \frac{X - \overline{X}}{\sigma}$$

$$Z_1 = \frac{X_1 - 168}{12}$$

Z_1 değeri bulunabilir.

0,8264 – 0,5 =0,3264

tablodan ters işlemle 0,3264 olasılık değerini (yaklaşık) veren Z=0,94 X_1>X olduğundan,

Z_1= +0,94

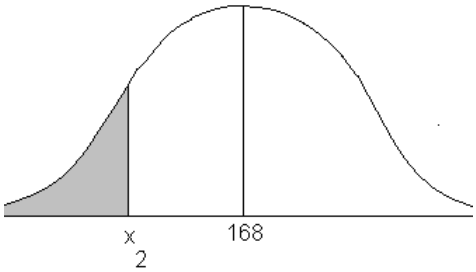
$$0{,}94 = \frac{X_1 - 168}{12}$$

$12.0{,}94 = X_1 - 168$

$X_1 = 168 + 11{,}28$

$= 179{,}28$

b) $P(X<X_2) = 0{,}33$



$$Z_2 = \frac{X_2 - 168}{12}$$

Z'nin bulunması,

0,5 – 0,33 = 0,17

tablodan ters işlemle Z=0,44

X_2<X olduğundan Z_2 = -0,44

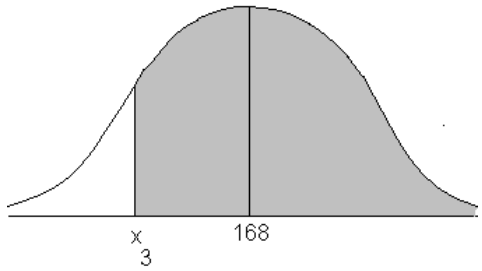
$$-0{,}44 = \frac{X_2 - 168}{12}$$

$X_2 = 168 - 12.0{,}44$

$X_2 = 168 - 5{,}28$

$= 162{,}72$

c) $P(X>X_3) = 0,5596$



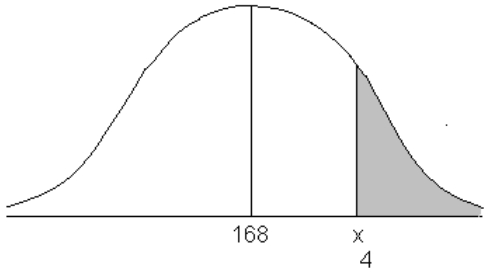
$$Z_3 = \frac{X_3 - 168}{12}$$

Z'nin bulunması,
$0,5596 - 0,5 = 0,0596$ tablodan $Z=0,15$
$X_3<X$ olduğundan $Z_3 = -0,15$

$$-0,15 = \frac{X_3 - 168}{12}$$

$X_3= 166,2$

d) $P(X>X_4) = 0,1056$



$$Z_4 = \frac{X_4 - 168}{12}$$

Z'nin bulunması,

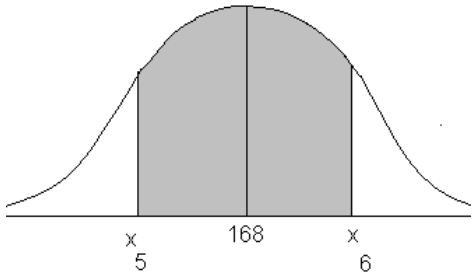
$0,5 - 0,1056 = 0,3944 \rightarrow Z = 1,25$

$X>X_4$ olduğundan $Z_4 = 1,25$

$$1,25 = \frac{X_4 - 168}{12}$$

$X_4= 183$

e) $P(X_5<X<X_6) = 0,5$ (simetrik)



$$Z_5 = \frac{X_5 - 168}{12}$$

$$Z_6 = \frac{X_6 - 168}{12}$$

0,5/2 = 0,25 $\rightarrow$ Z = 0,67

$X_5 < \overline{X}$ olduğundan Z_5 = - 0,67

$X_6 < \overline{X}$ olduğundan Z_6 = + 0,67

$$-0{,}67 = \frac{X_5 - 168}{12}$$

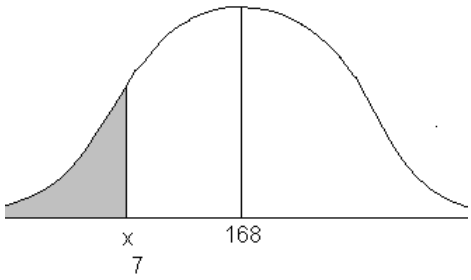
X_5=159,96

$$0{,}67 = \frac{X_6 - 168}{12}$$

X_6= 176,04

f) $P(X<X_7) = 0{,}4207$

$P(X_7<X<X_8) = 0.3979$



$$Z_7 = \frac{X_7 - 168}{12}$$

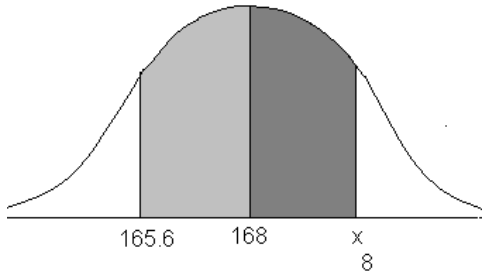
0,5 – 0,4207 = 0,0793 → Z = 0,2
$X_7 < \overline{X}$ olduğundan Z_7 = - 0,2

$$-0,2 = \frac{X_7 - 168}{12}$$

X_7 = 165,6

$P(X_7<X<X_8) = P(165,6<X<X_8) = 0,3979$
P(165,6<X<168) = 0,0793 olduğundan

$X_8 > \overline{X}$ olacaktır.



$$Z_8 = \frac{X_8 - 168}{12}$$

Z_8 değeri hesaplanabilir.
0,3979 – 0,0793 = 0,3186 → Z = 0,91
$X_8 >$ olduğundan Z_8= + 0,91

$$0,91 = \frac{X_8 - 168}{12}$$

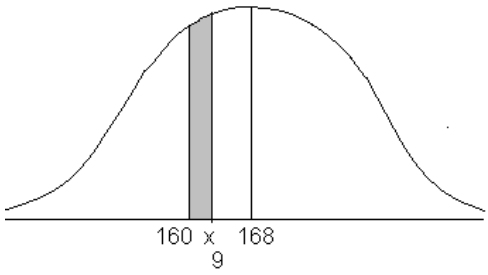
X_8= 178,92

g) $P(160<X<X_9) = 0,1$

X_9 değerinin $\overline{X}$ 'dan büyük mü küçük mü olduğunu belirleyelim.
160 – 168 arası alan,

$$Z = \frac{160 - 168}{12} = 0,67$$

0,2486, 0,1'den büyük olduğundan X_9 değeri $\overline{X}$ 'dan küçüktür.



$$Z_9 = \frac{X_9 - 168}{12}$$

0,2486 – 0,1 = 0,1486 → Z = 0.38
$X_9 < \overline{X}$ olduğundan Z_9 = - 0.38

$$-0{,}38 = \frac{X_9 - 168}{12}$$

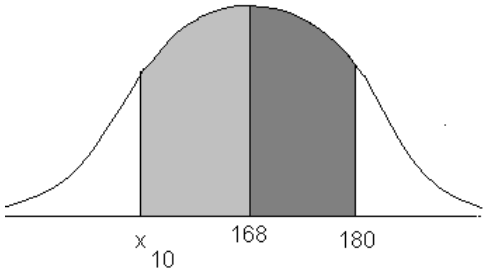
X_9 = 163.44

h) $P(X_{10}<X<180) = 0{,}6$

168 – 180 arası alan,

$$Z = \frac{180 - 168}{12} = 1 \rightarrow 0{,}3413$$

olduğundan $X_{10} < \overline{X}$ 'dır.



$$Z_{10} = \frac{X_{10} - 168}{12}$$

0,6 – 0,3413 = 0,2587 → Z = 0,7
$X_{10} < \overline{X}$ olduğundan Z_{10} = - 0,7

$$-0,7 = \frac{X_{10} - 168}{12}$$

$X_{10} = 159,6$

ÖRNEK 7: Aşağıda bir atölyede üretilen 500 mamülün çapları mm. olarak verilmiştir ve dağılım normal kabul edilmektedir.

Sınıflar	f_i	X_i	f_iX_i	$f_iX_i^2$
10-12	25	11	275	3025
12-14	95	13	1235	16055
14-16	240	15	3600	54000
16-18	110	17	1870	31790
18-20	30	19	570	10830
	500		7550	115700

a) Dağılımın parametrelerini hesaplayınız.
b) P(X>13) ve P(11<X<17,5) olasılıklarını hesaplayınız.

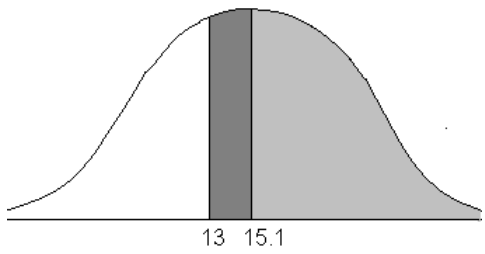
Çözüm:

a) $\overline{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{7550}{500} = 15,1$

$$\sigma = \sqrt{\frac{\Sigma f_i X_i^2}{\Sigma f_i} - (\overline{X})^2} = \sqrt{\frac{115700}{500} - (15,1)^2}$$

$$= \sqrt{231,4 - 228,01} = 1,84$$

b) P(X>13)



$$Z = \frac{X - \overline{X}}{\sigma} = \frac{13 - 15,1}{1,84} = \frac{-2,1}{1,84} = -1,14 \rightarrow 0,3729$$

P(X>13) = 0,5 + 0,3729 = 0,8729

Teorik olarak bulduğumuz olasılık değerini seriden hesaplayalım.

	Fark	Frekans
12 – 14 aralığında	2	95
13 – 14 aralığında	1	X

X=47,5

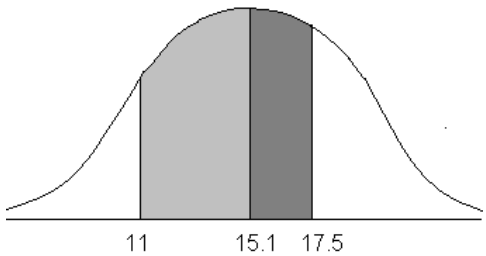
13 ve daha büyük X değerleri için frekanslar toplamı,

Sınıflar	f_i
13-14	47,5
14-16	240
16-18	110
18-20	30
	427,5

Toplam Frekans Oranı= 427,5/500=0,855

Yaklaşık sonuç elde edileceği görülmektedir.

P(11<X<17,5)



$$Z_1 = \frac{11-15,1}{1,84} = \frac{-4,.1}{1,84} = 2,23 \rightarrow 0,4871$$

$$Z_2 = \frac{17,5-15,1}{1,84} = \frac{2,4}{1,84} = 1,30 \rightarrow 0,4032$$

P(11<X<17,5) = 0,4871 + 0,4032 = 0,8903

Aynı değeri seriden bulalım.

	Fark	Frekans
10 – 12 aralığında	2	25
11 – 12 aralığında	1	X

X=12,5

16 – 18 aralığında	2	110
16 – 17,5 aralığında	1,5	X

X=82,5

11 – 17,5 aralığı toplam frekansı,

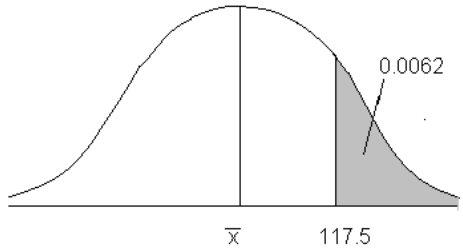
Sınıflar	f_i
11-12	12,5
12-14	95
14-16	240
16-17,5	82,5
	430,0

Toplam Frekans Oranı= 430/500=0,86
Arada yaklaşık %4 fark olduğu görülmektedir.

ÖRNEK 8: Normal olduğu bilinen bir dağılımla ilgili olarak şu bilgiler mevcuttur.

P(X>117,50) = 0,0062
P(X>94,75) = 0,7734
Dağılımın parametrelerini bulunuz.

Çözüm: P(X>117,50) = 0,0062

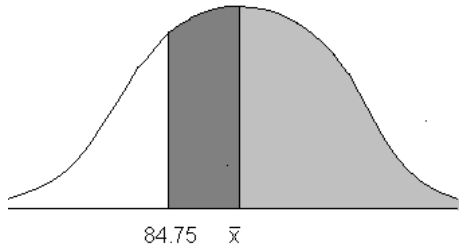


Z değeri bulunabilir.

0,5 – 0,0062 = 0,4938 → Z = 2,5
(117,5 > $\overline{X}$ olduğundan pozitif)

$2{,}5\sigma = 117{,}5 - \overline{X}$ (I)

P(X>94,75) = 0,7734



$$Z = \frac{94{,}75 - \overline{X}}{\sigma}$$

Z'yi bulalım.

$0{,}7734 - 0{,}5 = 0{,}2734 \rightarrow Z = 0{,}75$

$94{,}75 < \overline{X}$ olduğundan $Z = -0{,}75$

$$-0{,}75 = \frac{94{,}75 - \overline{X}}{\sigma}$$

$$-0{,}75\,\sigma = 94{,}75 - \overline{X} \qquad \text{(II)}$$

İki bilinmeyenli bu iki denklemden,

$$2{,}5\sigma = 117{,}5 - \overline{X}$$
$$\mp \pm 0{,}75\sigma = \mp 94{,}75 \pm \overline{X}$$

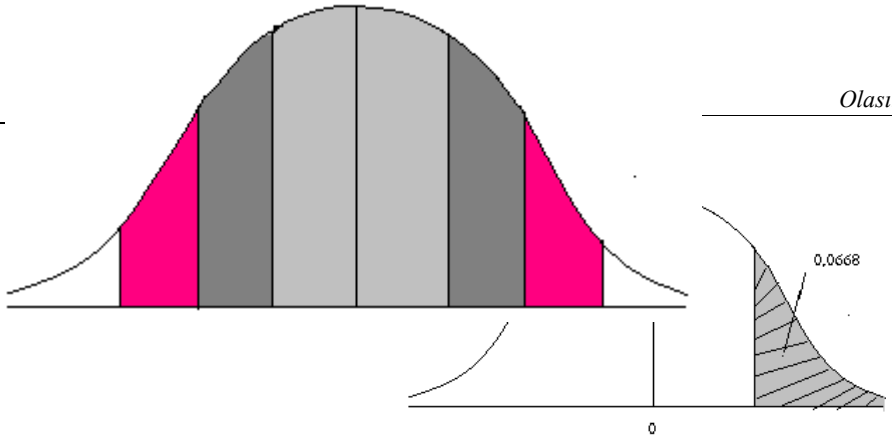
$$3{,}25\,\sigma = 22{,}75$$
$$\sigma = 7$$
$$2{,}5\sigma = 117{,}5 - \overline{X}$$
$$(2{,}5).7 = 117{,}5 - \overline{X}$$
$$\overline{X} = 117{,}5 - 17{,}5 = 100$$

bulunur.

ÖRNEK 9: Standart sapması 6 birim olan ve normal dağıldığı bilinen bir anakütleyle ilgili

$P(x>59) = 0{,}0668$

olasılığı bilinmektedir. Dağılımın aritmetik ortalamasını hesaplayınız.



Z değeri bulunabilir.

$0,5 - 0,0668 = 0,4332 \rightarrow Z = 1,5$

(59 > $\overline{X}$ olduğundan pozitif)

$1,5\ (6) = 59 - \overline{X}$

$9 = 59 - \overline{X}$

$\overline{X} = 50$

bulunur.

BÖLÜM 10

TAHMİN TEORİSİ

Daha önce tamsayım, örnek, örnekleme ve parametre kavramlarından sözedilmişti. Hem kısaca sayım olarak adlandırılan tam sayım, hem de örnekleme ile ulaşılmak istenen, anakütle parametrelerinin değerlerinin belirlenmesidir. Tam sayım sözkonusu olduğunda elde edilen değerler anakütle parametrelerinin gerçek değerleridir. Burada ancak ölçme, değerlendirme, işlem hataları gibi hatalar yapılmış olabileceği düşünülebilir. Anakütle parametrelerinin değerlerinin örnekleme ile belirlenmesi sözkonusu olduğunda tahmin olayı ile karşılaşılmaktadır.

Anakütle parametresi için örnekten bir tek değer tahmin edilmesi, nokta tahmini olarak adlandırılmaktadır. Daha sonra sözedileceği gibi bu tahmin örnekleme hatası içerebilir. Belirlenen bir olasılıkla örnekleme hatasının boyutları tahmin edilebilir ki, bu aralık tahminidir. Burada önce nokta tahmini, daha sonra ise aralık tahmini açıklanacaktır.

10.1. TAHMİN VE TAHMİNCİ

Birim sayısı N olan bir anakütleden birim sayısı n olan farklı örnekler alınabilir. Seçilebilecek farklı örnek sayısı, N ve n'in sayısal büyüklüğü yanında örneği oluşturacak birimlerin seçiminin iadeli veya iadesiz olarak yapılmasına da bağlıdır. Düzenlenebilecek farklı örnek sayısını k ile ifade edersek, iadeli seçim için,

$$k = N^n$$

ve iadesiz seçim için,

$$k = C_N^n$$

olacaktır.

Anakütle parametresinin belirlenmesi için uygulamada bu örneklerden biri seçilecektir. Seçilebilecek bu örneklerden bazılarının parametrelerinin değeri anakütle

parametresine eşit, bazılarının ki anakütle parametresinden küçük ve bazılarının ki anakütle parametresinden büyük olacaktır. Bu nedenle örnekleme söz konusu olduğunda anakütle parametresinin örnekten tahmin edilmesi gerekmektedir. Anakütlenin örnekten tahmin edilen parametresi "örnek istatistiği" veya kısaca "istatistik" olarak adlandırılmaktadır.

Olaya mantıksal olarak bakılırsa, örnek birim sayısının artması ile daha doğru tahminlerin yapılması beklenecektir. Örnek birim sayısının anakütle birim sayısına oranı "örnekleme oranı" olarak adlandırılır.

$$\text{Örnekleme Oranı} = \frac{n}{N}$$

Örnekleme oranının artması yani 1'e doğru yaklaşması durumunda yapılacak tahminlerin daha doğru olması mümkün olabilir.

Daha doğru tahminler elde edilmesi anakütle birimlerinin homojenliği ile de ilgilidir. Homojenlik birimlerin değerlerinin birbirine yakınlığını ifade etmektedir. Anakütle ne kadar homojense, anakütleyi oluşturan birimlerin değerleri birbirine ne kadar yakınsa, yapılacak tahmin gerçeğe o kadar yakın olacaktır.

Tahminin doğruluğu, tahminde kullanılacak formüle de bağlıdır. Burada cevabı araştırılması gerekli soru en doğru tahminin hangi formülle yapılması gerektiğidir. En iyi sonucu anakütle parametresini hesapladığımız formülün mü yoksa farklı bir formülün mü vereceği belirlenmelidir. Örneğin, anakütle ortalaması veya varyansı daha önce verilen ve anakütle için kullanılan formüllerle mi, yoksa farklı formüllerle mi tahmin edilirse, daha doğru sonuç elde edilir. Diğer bir ifade ile anakütle ortalaması ve varyansı örnek ortalaması ve varyansına eşit olarak mı tahmin edilecektir.

Anakütle parametrelerinin tahmini için kullanılan formüllere tahminci adı verilmektedir. Yukarıda söz ettiğimiz soruların cevapları, ancak tahmincilerin özelliklerinin belirlenmesi ve tahmincilerde bu özelliklerin aranması ile açıklık kazanacaktır. Tahmincinin özellikleri daha sonra incelenecektir.

10.2. ÖRNEKLEME DAĞILIMI VE STANDART HATA

Açıklandığı gibi N birimli bir anakütleden n birimli k sayıda örnek düzenlenebilmektedir. Alınacak örnek ile tahmin edilecek anakütle parametresi θ; kullanılacak tahminci $\hat{\theta}$ ise, seçilecek örneğe bağlı olarak k adet farklı tahmin yapılabilecektir.

Örnek No	Tahmin
1	$\hat{\theta}_1$
2	$\hat{\theta}_2$
3	$\hat{\theta}_3$
.	.
.	.
k	$\hat{\theta}_k$

$\hat{\theta}$ 'lerin beklenen değeri anakütle parametre θ'nın gerçek değerine eşit olacaktır.

$$E(\hat{\theta}) = \theta$$

$\hat{\theta}$ 'lerin standart sapması ise standart hata olarak adlandırılır. $\hat{\theta}$ 'lerin varyansı,

$$Var(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

olacağından, bunun karekökü standart hataya eşittir. Örneğin, örnek ortalaması için standart hata,

$$Var(\overline{X}) = E(\overline{X} - \mu)^2$$

ifadesinin kareköküdür.

10.3. SAPMA, ÖRNEKLEME HATASI VE ORTALAMA HATA KARE

Anakütle parametresi θ ile $\hat{\theta}$ 'lerin beklenen değeri birbirine eşit olmayabilir. Bu durumda aradaki fark sapma olarak adlandırılır.

$$\text{Sapma} = E(\hat{\theta}) - \theta$$

Örnekleme hatası ise sapmadan farklı bir kavramdır ve tahminci ile anakütle parametresinin gerçek değeri arasındaki farktır.

$$\text{Örnekleme Hatası} = \hat{\theta} - \theta$$

Daha önce bir anakütleden düzenlenebilecek k sayıdaki farklı örneklerin bazılarında $\hat{\theta} = \theta$, bazılarında $\hat{\theta} < \theta$, bazılarında $\hat{\theta} > \theta$ olabileceğinden sözetmiştik. Örnekleme θ'yı belirlemek için yapıldığından uygulamada θ bilinmeyecektir. Bu nedenle örnekleme hatası da kesin olarak belirlenemeyecektir. Ancak örnekleme hatasının büyüklüğü belirlenecek bir olasılıkla tahmin edilebilecektir.

Açıkladığımız gibi farklı örneklerden tahmin edilecek $\hat{\theta}$ 'ler bir dağılım oluştururlar. Bu dağılım için varyansa benzer bir ölçü kullanılmaktadır. Ortalama hata kare olarak adlandırılan bu ölçü, $\hat{\theta}$ 'ler ile θ arasındaki farkların karelerinin beklenen değeridir. Bu tanıma göre ortalama hata kare,

$$OHK(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

olarak ifade edilir. Bu ifadeye $\pm E(\hat{\theta})$ eklenerek açıklanırsa,

$$OHK(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]$$

bulunur. Eşitliğin sağ tarafındaki ilk ifade $\hat{\theta}$ 'in varyansına, ikinci ifade ise sapmanın karesine eşittir.

$$OHK(\hat{\theta}) = Var(\hat{\theta}) + (sapma)^2$$

Görüldüğü gibi ortalama hata kare varyansa benzer, fakat farklı bir kavramdır. Varyans aritmetik ortalama etrafındaki dağılımın, ortalama hata kare ise, anakütle parametresinin gerçek değeri etrafındaki dağılımın ölçüsüdür. Ancak sapmanın sıfır olması durumunda,

$$Var(\hat{\theta}) = OHK(\hat{\theta})$$

olacaktır.

10.4. TAHMİNCİNİN ÖZELLİKLERİ

Anakütle parametresi θ'nın tahmini için alınacak n birimli örnekten farklı tahminciler kullanılarak farklı tahminler yapılabilir. Burada önemli olan en iyi sonucu verecek tahmincinin kullanılmasıdır.

En iyi tahmini sağlayacak tahmincinin bazı özellikleri taşıması gerekmektedir. Bu özellikler şunlardır.

-Sapmasızlık

-Etkinlik

-Tutarlılık

-Yeterlilik

Sırası ile bu özellikleri açıklayalım.

10.4.1. Sapmasızlık

Sapma $\hat{\theta}$ 'in beklenen değeri ile θ arasındaki fark olarak tanımlanmıştı. Buna göre $\hat{\theta}$ 'in θ'nın sapmasız tahmincisi olabilmesi için $E(\hat{\theta})$ ile θ arasındaki fark sıfır olmalıdır. Diğer bir ifade ile,

$$E(\hat{\theta}) = \theta$$

olmalıdır. $\hat{\theta}$ 'in beklenen değeri alındığında sonuç θ, ise, $\hat{\theta}$ θ'nın sapmasız tahmincisidir.

ÖRNEK: Anakütle ortalaması μ'nün tahmini için örnek ortalaması $\overline{X}$ kullanıldığında, $\overline{X}$ 'nın μ'nün sapmasız tahmincisi olup olmadığını inceleyelim.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

olduğundan, $\overline{X}$ 'nın beklenen değeri alınacaktır. Eğer $\overline{X}$ 'nın beklenen değeri μ'ye eşit ise, örnek ortalamasının anakütle ortalamasının sapmasız tahmincisi olduğunu söyleyebiliriz. Aksi halde anakütle ortalamasını tahmin etmek için farklı tahminciler veya farklı formüller kullanılacaktır.

$$E(\overline{X}) = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n} nE(X_i) = E(X_i) = \mu$$

Örnek ortalaması anakütle ortalamasının sapmasız tahmincisidir.

10.4.2. Etkinlik

Anakütlenin bilinmeyen bir parametresinin tahmini için birden fazla sapmasız tahminci belirlenebilir. Bu durumda, bu tahmincilerden hangisinin kullanılması ile daha iyi tahmin yapılabileceğini belirlemek için tahmincilerin varyansları belirlenir. Kısaca etkinlik, sapmasız tahmincilerin varyansları ile ilgili bir kavramdır.

Bilindiği gibi, birden fazla serinin ortalamaları birbirine yakın veya eşitse dağılımların birbirinden farkını ortaya koymak için bazı ölçüler hesaplanır. Bu ölçülerden en çok kullanılanı ise varyans veya bunun karekökü olan standart sapmadır. Varyansı ve standart sapması küçük olan seride aritmetik ortalama etrafında daha yoğun bir dağılım olduğu anlaşılır. Yani, varyansı küçük olan seride seri birimleri, varyansı büyük olan seriye göre aritmetik ortalamaya daha yakın bir dağılım göstermişlerdir. Varyans ve standart sapma tek başına hesaplandığında fazla bir anlam ifade etmeyip, ancak başka varyans veya standart sapmalar ile karşılaştırıldığında anlam kazanmaktadır. Kısaca varyans nisbi bir ölçüdür.

Etkinlik kavramı ile ilgili olarak varyans ve standart sapma ile ilgili açıklamalarımıza benzer bir açıklama yapabiliriz. Burada konumuz anakütle parametresinin gerçek değerine daha yakın tahminler elde etmek olduğuna göre, bunun ölçüsü daha önce açıklanan ortalama hata karedir. Birden fazla sapmasız tahminci sözkonusu olduğunda, bunlardan ortalama hata karesi küçük olan tahminci anakütle parametresinin gerçek değeri etrafında daha yoğun bir dağılım gösterdiğinden, bu tahmincinin tercih edilmesi gerekmektedir. Ortalama hata karenin varyans ile sapmanın karesi toplamına eşit olduğu hatırlanırsa, sapmasız tahminciler sözkonusu olduğundan ortalama hata kareler varyanslara eşit olacaktır. bu nedenle varyansı küçük olan tahmin-

cinin tercih edilmesi ile ortalama hata karesi küçük olan tahminci tercih edilmiş olacaktır.

Yaptığımız bu açıklamalar etkinlik için iki şart aranması gerektiğini ortaya koymaktadır.

1) $\hat{\theta}$, θ'nın sapmasız tahmincisi ise,

2) $Var(\hat{\theta}) \leq Var(\hat{\theta})$ ise

$\hat{\theta}$, θ'nın etkin tahmincisidir. Burada $\hat{\theta}$, anakütle parametresi θ'nın $\hat{\theta}$ 'ten farklı diğer sapmasız tahmincilerini ifade etmektedir.

Etkinlik ile ilgili olarak buraya kadar yapılan açıklamalar, etkinliğin nisbi bir kavram olduğunu ortaya koymaktadır. Ancak birden fazla sapmasız tahminci olması durumunda etkinlikten sözedilebilir ve varyansı küçük olan tahmincinin daha etkin olduğu söylenir.

ÖRNEK: Anakütle ortalamasının tahmini için örnek aritmetik ortalaması ve örnek medyanının kullanıldığını düşünelim. Her ikisi de anakütle ortalamasının sapmasız tahmincisidir. Örnek ortalamalarının dağılımının varyansının,

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

olduğu daha önce gösterilmişti. Benzer şekilde örnek medyanlarının örnekleme dağılımlarının varyansı,

$$Var(M_e) = \frac{\pi\sigma^2}{2n}$$

olarak elde edilir. Bu tahmincilerden daha küçük varyansa sahip olan daha etkin olacağından varyansların karşılaştırılması gerekecektir.

$$Var(M_e) = Var(\overline{X})\frac{\pi}{2}$$

olduğundan $Var(\overline{X}) < Var(M_e)$ 'dır ve $\overline{X}$ daha etkin tahmincidir.

10.4.3. Tutarlılık

a küçük bir sayı olsun. Eğer n→ ∞ için,

$$P\left(\left|\hat{\theta} - \theta\right| \leq a\right) \rightarrow 1$$

ise $\hat{\theta}$, θ'nın tutarlı tahmincisidir. Burada ifade edilen örnek birim sayısının artması ile $\hat{\theta}$ ile θ arasında varolduğunu kabul ettiğimiz küçük farkın azalması olasılığının

bire yaklaşmasıdır. Eğer örnek birim sayısının artması ile tahminci ile anakütle parametreleri birbirine yaklaşıyorsa tahminci tutarlıdır.

Örnek birim sayısının artması durumunda sapma ile örnekleme dağılımının varyansı küçülecektir. Böylece hata kare küçülmüş olacaktır. Eğer $n \to \infty$ ise ortalama hata kare sıfıra eşit olacaktır.

$$\lim_{n\to\infty} OHK(\hat{\theta}) = 0$$

Tahmincinin ortalama hata karesinin n sonsuza giderken limiti alınırsa, sonucun sıfır çıkması tahmincinin tutarlı olduğunu gösterecektir. Fakat n sonsuza giderken ortalama hata karesinin limiti sıfıra eşit olmayan tutarlı tahminciler de bulunabilir. Bu nedenle ortalama hata karenin n sonsuza giderken limitinin sıfır olması tahmincinin tutarlı olduğunu gösterirse de, limitin sıfırdan farklı olması tahmincinin tutarsız olduğunu göstermeyecektir.

10.4.4. Yeterlilik

Anakütle parametresi θ'nın değerinin tahmini için kullanılan tahminci $\hat{\theta}$, örnekteki bilgilerin tümünü kullanıyorsa, yeterli tahmincidir. Bu özelliğe göre örnek anakütle ile ilgili bilgiler taşımaktadır. Bu bilgiler örnek birimlerinde veya bunların bütünün de bulunmaktadır. Tüm örnek birimlerini kullanmayan tahminciler yeterli değildir. Örneğin anakütle ortalaması μ'nin tahmini için tahminci olarak örnek ortalaması $\overline{X}$, örnek modu ve örnek medyanı kullanılacak olsun. Örneğin tüm birimleri kullanıldığı için örnek ortalaması $\overline{X}$ yeterli tahmincidir. Fakat aynı tahmin için örneğin tüm birimleri kullanılmadan hesaplanan mod veya medyan yeterli olmayan tahmincidir.

10.5. TAHMİN YÖNTEMLERİ

Buraya kadar iyi bir tahmincinin taşıması gerekli özellikler incelenmiştir. Ayrıca, verilen örneklerle de tahmin formülleri herhangi bir yöntemle elde edilmeyip rastgele verilmiştir. Uygulamada farklı tahmin problemleri ile karşılaşılacak ve tahminci olarak yazılabilecek pek çok formülün sözkonusu özellikleri taşıyıp taşımadığını incelemek de çok zor olacaktır.

Tahmincilerin veya tahmin formüllerinin belirlenmesi için tahmin yöntemleri adı verilen yöntemlerden yararlanılır. Bu yöntemler daha önce söz edilen özellikleri veya bu özelliklerden bazılarını taşıyan tahmincilerin belirlenmesini sağlamaktadır.

10.5.1. Momentler Yöntemi

Adından da anlaşılacağı gibi bu yöntemde tahminciler momentler yardımı ile belirlenir. Örneğin anakütle aritmetik ortalaması tahmin edilecek ise örneğin orjine göre birinci momenti tahminci olarak kullanılır.

$$m_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \mu$$

Aynı şekilde anakütle varyansı tahmin edilecekse örneğin ortalamaya göre ikinci momenti tahminci olarak kullanılır.

$$m_2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} = \sigma^2$$

Momentler yöntemi ile elde edilen tahminciler genellikle tutarlı tahmincilerdir. Fakat bazıları bazı özellikleri taşımazlar. Karşılaşılan her olayda momentler belirlenemeyeceğinden bu durumda kullanılamazlar.

10.5.2. En Küçük Kareler Yöntemi

Yöntem, farkların kareleri toplamının minimize edilmesi esasına dayanır. Örneğin, bir X tesadüfi değişkeninin r. momentinin hesaplanacağını düşünelim. Burada r = 0,1,2,3,..... değerini alacaktır.

$$E(X^r) = m_r$$

m_r'nin en küçük kareler tahmincisini elde etmek için değişkenlerin r. kuvvetleri ile m_r farklarının kareleri toplamı alınacak ve bu farkların kareleri toplamı minimize edilecektir.

$$T = \sum_{i=1}^{n} (X_i^r - m_r)^2$$

Aritmetik ortalama orjine göre birinci momenti olduğuna göre aritmetik ortalamanın tahmincisi için,

$$T = \sum_{i=1}^{n} (X_i - m_1)^2$$

minimize edilecektir. m_1 anakütle ortalaması olduğuna göre m_1 yerine μ kullanırsak, minimize edilecek toplam,

$$T = \sum_{i=1}^{n} (X_i - \mu)^2$$

olacaktır. Minimizasyon için T'nin μ'ye göre birinci türevinin alınıp sıfıra eşitlenmesi gerekecektir. Elde edilen sonucun minimum veya maksimum olduğunun belirlenmesi için ise ikinci türevin işaretine bakılacaktır.

Aritmetik ortalama için birinci türevi alıp sıfıra eşitlersek,

$$\frac{dT}{d\mu} = \sum_{i=1}^{n} 2(X_i - \mu)(-1)$$

$$= \sum_{i=1}^{n} 2(X_i - \hat{\mu})(-1) = 0$$

görüldüğü gibi tahmin olduğunu belirtmek için sıfıra eşitlendiğinde μ yerine $\hat{\mu}$ sembolü kullanılmıştır.

$$-2\sum_{i=1}^{n}(X_i - \hat{\mu}) = 0$$

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \hat{\mu} = 0$$

$$\sum_{i=1}^{n} X_i - n\hat{\mu} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

elde edilir. Yani, anakütle ortalamasının en küçük kareler yöntemi ile belirlenen tahmincisi örnek ortalamasıdır. İkinci türev alınırsa toplamın minimize edildiği görülür.

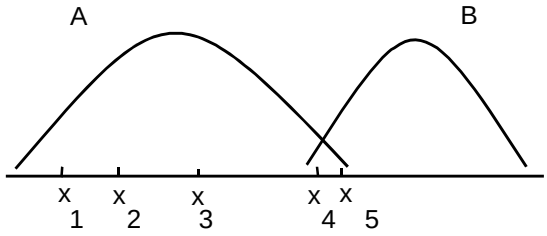
10.5.3. Maksimum Benzerlik Yöntemi

Maksimum benzerlik yönteminde herhangi bir örnek istatistiğinin ortaya çıkma olasılığı dikkate alınır. Farklı anakütlelerden alınacak örnekler de farklı olacağından, herhangi bir örneğin bazı anakütlelerden alınması olasılığı diğer bazı anakütlelerden alınması olasılığından daha yüksek olacaktır. Bir örnek diğer anakütlelerden çok alındığı anakütleye benzeyecektir.

Normal dağılmış bir anakütleyi düşünelim. Bu anakütleden alınacak n birimli örneğin birimleri ortalama etrafında dağılacaktır. Örnek ortalaması 8 ise, bu örneğin ortalaması 8 veya 8'e yakın anakütleden alınması olasılığı, ortalaması 6 veya 10 olan, yani 8'den farklı olan anakütlelerden alınması olasılığından daha fazladır. Örneğin, ortalaması 8 veya 8'e yakın anakütleye benzeme olasılığı daha fazladır.

Beş birimden oluşan (X_1, X_2, X_3, X_4, X_5) bir örnek alınmış olsun. A ve B gibi iki farklı anakütle düşünelim. Örnek birimleri şekilde görüldüğü gibi dağılıyorsa, bu

örneğin A anakütlesinden alınması veya A anakütlesine benzemesi olasılığı diğer anakütleden alınması veya diğer anakütleye benzemesi olasılığından daha fazladır.



Şekil 10.1.

Dağılım parametreleri $\theta_1, \theta_2, ..., \theta_k$ ve olasılık dağılımı f(X) olan bir anakütleden n birimden oluşan (X_1, X_2, ..., X_n) bir örnek alındığında; $\theta_1, \theta_2, ..., \theta_k$ 'nın maksimum benzerlikler yöntemi ile elde edilecek tahmincileri (X_1, X_2, ..., X_n)'in maksimizasyonu ile belirlenecektir.

Örneğin beş kişilik bir örnek alınarak kişilere soru sorulduğunu ve bunlardan ikisinin soruya evet cevap verdiğini düşünelim. Bu durumda X tesadüfi değişkeni sadece iki değer alacak ve tek parametreli bir dağılım sözkonusu olacaktır.

f(E) = p

f(H) = 1-p

Örneğimizde örnek oranı,

$$p = \frac{2}{5} = 0.4$$

tür. İki evet, üç hayır sonucunu oranı ne olan anakütleden alma olasılığının daha yüksek olduğunu belirleyelim. Bunun için anakütle oranı 0;0.1;0.2;.....;0.9;1 olarak alınacaktır. İki evet üç hayır sonucu bu olasılıklarla hesaplanacaktır.

P	f(E,E,H,H,H)
0	(0x0x1x1x1) = 0
0.1	(0.1x0.1x0.9x0.9x0.9)= 0.00729
0.2	(0.2x0.2x0.8x0.8x0.8)= 0.02048
0.3	(0.3x0.3x0.7x0.7x0.7)= 0.03087
0.4	(0.4x0.4x0.6x0.6x0.6)= 0.03456
0.5	(0.5x0.5x0.5x0.5x0.5)= 0.03125
0.6	(0.6x0.6x0.4x0.4x0.4)= 0.02304
0.7	(0.7x0.7x0.3x0.3x0.3)= 0.01323
0.8	(0.8x0.8x0.2x0.2x0.2)= 0.00512
0.9	(0.9x0.9x0.1x0.1x0.1)= 0.00081
1	(1x1x0x0x0) = 0

Yukarıda verilen olasılıklar incelenirse, en yüksek olasılığın 0.03456 olduğu ve p=0.4 için hesaplandığı görülecektir. Bu sonuç sözkonusu örneğin oranı 0.4 olan anakütleden alınma olasılığının, oranı farklı olan anakütlelerden alınma olasılığından daha yüksek olduğunu göstermektedir. f(E,E,H,H,H) bir benzerlik fonksiyonudur ve 0.4 olasılığında maksimize olmuştur.

Maksimum benzerlik fonksiyonu örneğin bileşik olasılık dağılımının formülüne verilen addır ve ℓ ile gösterilir.

$$\ell = f(X_1, X_2, ..., X_n)$$

X tesadüfi değişkeni olasılık fonksiyonu f(X) olan bir tesadüfi değişken ise tesadüfi olarak seçilen n adet örnek birimi (X_1, X_2, ..., X_n), X ile aynı dağılıma sahip bağımsız tesadüfi değişkenlerdir. X_1, X_2, ..., X_n bağımsız olduklarından benzerlik fonksiyonu,

$$\ell = f(X_1)f(X_2)f(X_3) .....f(X_n)$$

olarak yazılabilir. Örneğin birleşik olasılık formülü ile benzerlik fonksiyonu aynıdır. Birleşik olasılık dağılımında parametreler ($\theta_1, \theta_2, ..., \theta_k$) sabit, örnek birimleri (X_1, X_2, ..., X_n) değişkendir. Benzerlik fonksiyonunda ise parametreler değişken, belirli bir örnekten alındıkları için örnek birimleri sabittirler. Formüller aynı, fakat yorumları farklıdır.

Maksimum benzerlik yönteminde benzerlik fonksiyonu belirlenerek, bu fonksiyon parametreler için maksimize edilir. Maksimizasyon için benzerlik fonksiyonunun, fonksiyonda yer alan parametrelere göre birinci türevi alınarak bunlar sıfıra eşitlenir. Böylece istenen formüller elde edilmiş olur.

Elde edilen sonuçların maksimum veya minimum olduğunu belirlemek için ikinci türevin işaretine bakmak gerekir. Uygulamada benzerlik fonksiyonu ℓ yerine, bunun logaritmasının alınması kolaylık sağlayacaktır. Bir fonksiyonun kendisini veya logaritmasını maksimize etmek arasında bir fark olmayacaktır. Negatif değerler için logaritma alınması sözkonusu olamazsa da, burada birleşik olasılık fonksiyonları ile çalışıldığından, böyle bir problem ile karşılaşılmayacaktır. Logaritmik benzerlik fonksiyonunu ℓ yerine L ile ifade edeceğiz.

$$L = \ln \ell$$

ÖRNEK: Daha önce sonuçlarını sayısal olarak incelediğimiz örneği alarak tahminciyi maksimum benzerlik yöntemi ile belirleyelim. Örneğimizde X sadece iki değer alan bir tesadüfi değişkendi.

$$f(E) = p$$
$$f(H) = 1-p$$

olacaktır. Evet istenen cevap sayılarak bu durumda X=1, hayır istenmeyen cevap sayılarak bu durumda X=0 kabul edilirse, tesadüfi değişkenin olasılık fonksiyonu,

$$f(X) = p^x (1-p)^{1-x}$$

olacaktır. X'in 0 ve 1 değerleri için,

$$f(E) = (1-p)^{1-1} p^1 = p$$

$$f(H) = (1-p)^{1-0} p^0 = 1-p$$

bulunur. X tesadüfi değişkeninin beklenen değeri ise,

$$E(X) = \sum_{i=1}^{n} p_i X_i = 0(1-p) + 1p = p$$

dir. Örneğimizde 5 birimden oluşan bir örnek alınmıştı. Burada örnek birim sayısını n olarak alalım. Benzerlik fonksiyonu,

ℓ = f(X_1)f(X_2)f(X_3) .....f(X_n)

$$\ell = [p^{X_1}(1-p)^{1-X_1}][p^{X_2}(1-p)^{1-X_2}][p^{X_3}(1-p)^{1-X_3}].....[p^{X_n}(1-p)^{1-X_n}]$$

olacaktır. Çarpılacak ifadelerle tabanlar aynı olduğundan üstler toplanır.

$$\ell = p^{X_1+X_2+X_3+....+X_n}.(1-p)^{(1-X_1)+(1-X_2)+....+(1-X_n)}$$

$$\ell = p^{\sum_{i=1}^{n} X_i}.(1-p)^{n-\sum_{i=1}^{n} X_i}$$

Daha önce logaritmik fonksiyon ile çalışmanın kolaylığından sözetmiştik. Benzerlik fonksiyonunun logaritmasını alalım.

$$\mathrm{L} = \ln \ell = \left(\sum_{i=1}^{n} X_i\right) \ln p + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p)$$

p'nin tahmincisi belirleneceğine göre L'nin p'ye göre kısmi türevi alınarak sıfıra eşitlenecektir.

$$\frac{dL}{dp} = \left(\sum_{i=1}^{n} X_i\right)\left[\frac{1}{p}\right] + \left(n - \sum_{i=1}^{n} X_i\right)\left[\frac{1}{1-p}\right](-1)$$

$$\left(\sum_{i=1}^{n} X_i\right)\left[\frac{1}{p}\right] + \left(n - \sum_{i=1}^{n} X_i\right)\left[\frac{1}{1-p}\right](-1) = 0$$

Paydaları eşitleyerek içler dışlar çarpımı yapalım.

$$\frac{\left(\sum_{i=1}^{n} X_i\right)(1-\hat{p})+\left(n-\sum_{i=1}^{n} X_i\right)(-\hat{p})}{\hat{p}(1-\hat{p})}=0$$

$$\left(\sum_{i=1}^{n} X_i\right)(1-\hat{p})+\left(n-\sum_{i=1}^{n} X_i\right)(-\hat{p})=0$$

$$\sum_{i=1}^{n} X_i-\hat{p}\sum_{i=1}^{n} X_i-n\hat{p}+\hat{p}\sum_{i=1}^{n} X_i=0$$

$$\sum_{i=1}^{n} X_i-n\hat{p}=0$$

$$\hat{p}=\frac{\sum_{i=1}^{n} X_i}{n}$$

Burada n olayın tekrar sayısı olduğundan örneği n'e bölünecektir. Evet sonucu 1 ile hayır sonucu 0 ile gösterildiğinden örneğimizde $\sum_{i=1}^{n} X_i = 2$ olacaktır. Bu durumda anakütle oranı,

$$\hat{p}=\frac{2}{5}=0.4$$

olarak tahmin edilecektir.

10.5.4. En İyi Doğrusal Sapmasız Tahmin Yöntemi

Daha önce üç tahmin yönteminden söz ettik. Bu yöntemlerle elde edilen tahminciler daha önce açıklanan tahmincilerin özelliklerinden bazılarına sahipken, bazılarına sahip değildirler. Bu yöntemlerle istenen özelliklerin tümüne sahip tahminciler elde edilemez. Oysa en iyi doğrusal sapmasız tahmin yöntemi sözkonusu yöntemlerden farklıdır. Bu yöntem ile istenen özelliklere sahip olan tahminciler elde edilebilir.

Daha önce tahmincilerin özellikleri açıklanırken üçüncü küçük örnek özelliği olarak en iyi doğrusal sapmasız tahminci olma özelliği verilmiş ve tahminci $\hat{\theta}$'in anakütle parametresi θ'nın en iyi doğrusal sapmasız tahmincisi olması için,

- $\hat{\theta}$'in örnek birimlerinin doğrusal bir fonksiyonu olması,

- $\hat{\theta}$ 'in θ'nın sapmasız tahmincisi olması,

- $\hat{\theta}$ 'in varyansının diğer sapmasız tahmincilerin varyansına eşit veya onlardan küçük olması gerektiği belirtilmişti.

Burada en iyi doğrusal sapmasız tahminciler elde edilirken de tahmincilerin aynı özelliklere sahip olmaları istenecektir. Bu yöntem ile elde edilecek formüller yukarıda söz edilen özellikleri taşıyacaklardır. Bu formüller elde edilirken Lagrange Çarpanı Yöntemi kullanılmaktadır. Önce kısıtlamalar belirlenir ve daha sonra kısmi türevler alınarak tahminciler elde edilir.

10.6. ARALIK TAHMİNİ

Nokta tahmininde anakütlenin bilinmeyen parametresi θ için bir tek değer tahmin edilmektedir. Daha önce açıklandığı gibi anakütleden çok sayıda örnek alınabilir. Bu örneklerden elde edilecek tahminlerin bir kısmının değeri anakütle parametresinin değerine eşit, bir kısmınınki anakütle parametresinin değerinden büyük, bir kısmınınki ise anakütle parametresinin değerinden küçük olacaktır. Bu nedenle ele alınan örnekten ötürü tahmin ile anakütle parametresi gerçek değeri arasında bir fark olması olasılığı vardır. Örnekleme hatası olarak adlandırdığımız bu farkın belirlenen bir olasılık ile ne kadar olabileceğini tahmin etmek için aralık tahmini yapılır.

Aralık tahmininde belirlenen bir hata payı (anlam seviyesi) α veya güven olasılığı (1-α) ile anakütle parametresi θ için alt ve üst güven sınırları oluşturulur. Oluşturulacak alt sınır a ve üst sınır b ile ifade edilirse,

$$P(a < \theta < b) = 1 - \alpha$$

olacaktır. θ'nın sınırlarını belirlenmesinde tahmin $\hat{\theta}$ kullanılır. $\hat{\theta}$'lerin oluşturduğu seti A($\hat{\theta}$) ile ifade edersek,

$$P\left[\theta \in A(\hat{\theta})\right] = 1 - \alpha \qquad \text{ve}$$

$$P\left[\theta \in \overline{A}(\hat{\theta})\right] = \alpha$$

yazabiliriz. Birinci ifade θ'nın $\hat{\theta}$'lerin oluşturduğu setin bir elemanı olması olasılığının (1-α) olduğunu; ikinci ifade ise $\hat{\theta}$'nın sözkonusu setin elemanı olmaması olasılığının α olduğunu ifade etmektedir.

Aralık tahmini ile amaçlanan örneklemeden doğan sapmanın belirlenen güven olasılığı ile ne kadar olacağının tahmin edilmesidir. Örneğin bir anakütlenin ortalaması tahmin edilirken nokta tahmini 5 olarak elde edilmiş olsun. 0.05 hata payı veya 0.95 güven olasılığı ile alt ve üst güven sınırları 4 ve 6 olarak hesaplanmış olsun. Bu sonuç anakütle parametresinin değerinin 0.95 olasılıkla 4-6 arası olduğunu; anakütle parametresinin değerinin 4'ten küçük veya 6'dan büyük olması olasılığının ise 0.05 olduğunu ifade etmektedir.

Dikkat edilirse güven aralığı simetrik olarak düzenlenmektedir. Bunun sebebi sapmanın tek yönlü olmayıp, çift yönlü olmasıdır. Bazı parametrelerin aralık tahmin ile ilgili açıklamalar aşağıda yapılacaktır.

Aralık tahmininin yapılabilmesi için ilgili parametrelerin örnekleme dağılımlarının bilinmesi gerekmektedir. Bu nedenle aralık tahmininden önce örnekleme dağılımları ile ilgili açıklamalar yapılacaktır.

10.6.1. Anakütle Ortalamasının Aralık Tahmini

Anakütle ortalamasının aralık tahmininin yapılabilmesi için ortalamaların örnekleme dağılımının bilinmesi gereklidir. Bu nedenle önce ortalamaların örnekleme dağılımı açıklanacaktır.

10.6.1.1. Ortalamaların Örnekleme Dağılımı

Birim sayısı N ve ortalaması μ olan bir anakütleden alınacak n birimli örneğin ortalaması ($\overline{X}$), μ'nin istenen özellikleri taşıyan tahmincisidir. Bu nedenle,

$$\mu = \overline{X}$$

olarak tahmin edilir.

Ortalaması μ ve standart sapması σ olan bir anakütleden tesadüfi olarak seçilecek n birimli tüm örneklerin ortalamalarının dağılımının ortalaması μ ve standart sapması $\frac{\sigma}{\sqrt{n}}$ 'dir. $\frac{\sigma}{\sqrt{n}}$ tahminlerin standart hatası olarak adlandırılır ve $\sigma_{\overline{X}}$ şeklinde gösterilir.

Ortalamaların standart hatası, anakütle varyansı (σ^2) biliniyorsa şöyle tahmin edilecektir.

-iadeli seçimle örnek düzenleniyorsa,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \text{ dir.}$$

-iadesiz seçimle örnek düzenleniyorsa $\sigma_{\overline{X}}$, örnekleme oranına (n/N) bağlı olacaktır.

$$\frac{n}{N} < 0.05 \quad ise \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\frac{n}{N} \geq 0.05 \quad ise \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \text{ dir.}$$

Burada $\sqrt{\frac{N-n}{N-1}}$ sonlu anakütle düzeltme faktörü veya bassel faktörüdür. Seçim iadesiz ise örnekleme oranına bakılmaksızın formülde sonlu anakütle düzeltme faktörünün yeralması gerekmektedir. Ancak,

$$\frac{n}{N} < 0.05 \quad ise \quad \sqrt{\frac{N-n}{N-1}} = 1$$

kabul edilmekte ve bu nedenle ihmal edilmektedir.

Örnek birim sayısı (n) büyük ise, örnek ortalamalarının teorik dağılımı, anakütle dağılımı ne olursa olsun normal dağılıma yaklaşır.

Örnek büyüklüğü için kriter n ≥30'dur. n ≥30 ise örnek ortalamalarının teorik dağılımı normal kabul edilebilir. Anakütleden alınan örnek ile örnek ortalaması arasında fark olabilir. Bu fark örnek birim sayısına (n) bağlı olacak, n arttıkça örnek ortalamasının, anakütle ortalamasına yakın olma olasılığı artacaktır.

ÖRNEK: μ=50 ve σ=6 olan ve normal dağıldığı kabul edilen bir anakütleden 36 birimli bir örnek alınacaktır. Örnek ortalamasının 48-52 arasında olması olasılığını hesaplayın. (N= çok sayıda)

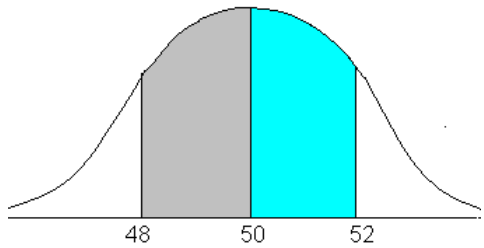
N büyük olduğundan n/N < 0.05 olacaktır. Örnek ortalamalarının dağılımı için standart değişken,

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \text{ dir.}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = \frac{6}{6} = 1$$

$$Z_1 = \frac{48-50}{1} = \frac{-2}{1} = -2 \rightarrow 0.4772$$

$$Z_2 = \frac{52-50}{1} = 2 \rightarrow 0.4772$$



$$P(48 < \overline{X} < 52) = 0.4772 + 0.4772 = 0.9544$$

n'in artarak 100 olduğunu varsayalım.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{100}} = \frac{6}{10} = 0.6$$

$$Z_1 = \frac{48-50}{0.6} = \frac{-2}{0.6} = -3.33 \rightarrow 0.4996$$

$$Z_2 = \frac{52-50}{0.6} = 3.33 \rightarrow 0.4996$$

Bu durumda

$$P(48 < \bar{X} < 52) = 0.4996 + 0.4996 = 0.9992$$

olur. n'in artması ile olasılığın arttığı görülmektedir.

Buraya kadar yaptığımız açıklamalarda anakütle varyansının (σ^2) bilindiği kabul edilmiştir. Anakütle varyansı bilinmiyorsa örnekten anakütle varyansı σ^2'nın tahmini gerekecektir. Anakütle varyansının sapmasız tahmincisi S^2'dir ve istenen diğer özellikleri de taşır.

S^2 veya S ,

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

veya

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

olarak tahmin edilir.

Örnek varyansı $\hat{\sigma}^2$ ise, $\hat{\sigma}^2$ ile S^2 arasında,

$$S^2 = \frac{n\hat{\sigma}^2}{n-1}$$

ilişkisi bulunacaktır. Buradan,

$$S = \hat{\sigma}\sqrt{\frac{n}{n-1}}$$

yazılabilir. Bu durumda,

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}$$

veya

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

olacaktır.

Ortalamaların örnekleme dağılımı ile ilgili özellikleri şöyle özetleyebiliriz.

1- Anakütle dağılımı normal ise ve anakütle varyansı biliniyorsa, örnek ortalamalarının dağılımı normaldir. Örnek birim sayısı önemli değildir. Bu durumda,

-iadeli seçim ve iadesiz seçimde n/N <0.05 ise

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

-iadesiz seçimde n/N ≥0.05 ise

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

olacaktır.

2- Anakütle dağılımı normal ise ve anakütle varyansı bilinmiyorsa, n ne olursa olsun örnek ortalamalarının dağılımı (n-1) serbestlik derecesi ile t-u dağılımıdır. n ≥30 ise t-dağılımı, normal dağılıma yaklaşacağından bu durumda dağılım normal kabul edilebilir.

Bu durumda σ bilinmediğinden S hesaplanarak birinci şıkta σ'ların yerini S alacaktır. Formüllerde bundan başka değişiklik olmayacaktır.

3- Anakütle dağılımı normal değil ise n ≥30 olmalıdır. Bu durumda n<30 için verilen formüller geçerli olmayacaktır.

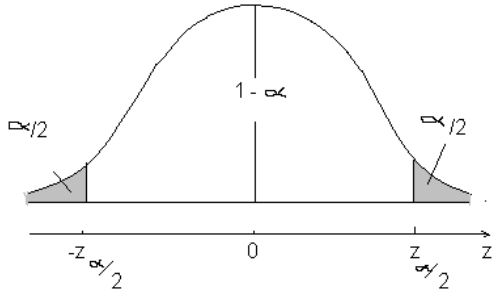
10.6.1.2. Aralık Tahmini

Örnek ortalamalarının dağılımı normal ise standart Z değişkeninin

$$Z = \frac{\overline{X} - \mu}{\sigma_{\bar{X}}}$$

değerleri normal dağılım tablosundan $Z_{\alpha/2}$ olarak elde edilecektir. Bu durumda,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$



Şekil 10.2

$$P(-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < +Z_{\alpha/2}) = 1 - \alpha$$

$$-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < +Z_{\alpha/2}$$

olacaktır. Tarafları $\sigma_{\overline{X}}$ ile çarparsak,

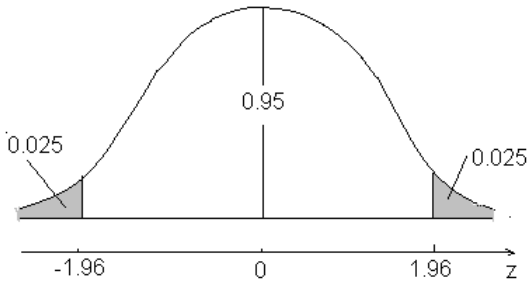
$$-Z_{\alpha/2}\sigma_{\overline{X}} < \overline{X} - \mu < +Z_{\alpha/2}\sigma_{\overline{X}}$$

$\overline{X}$ 'yı eşitsizliğin iki yanına alıp –1 ile çarparsak,

$$\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$$

olarak anakütle ortalaması μ'nün α hata payı veya 1-α güven olasılığı ile içinde bulunacağı aralığın alt ve üst sınırları bulunur. $\sigma_{\overline{X}}$ ise daha önce yapılan açıklamalara göre belirlenecektir.

ÖRNEK: Bir fabrikada üretilen A mamüllerinin boy uzunlukları araştırılmaktadır. Bu amaçla fabrikada üretilen çok sayıda mamülden 256 birimlik bir örnek alınmış ve örnek ortalaması 18 mm. bulunmuştur. Anakütle dağılımının normal ve varyansının 9 mm. olduğu bilindiğine göre, anakütle ortalamasını %95 olasılıkla (%5 hata payı ile) tahmin ediniz.



n=256 $\overline{X}$ =18

σ^2=9$\rightarrow$ σ=3 α=0.05

$$\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$$

$$Z_{\alpha/2} \rightarrow 1-\alpha = 1-0.05 = 0.95$$

$0.95/2 = 0.4750 \rightarrow$ tablodan $Z_{\alpha/2} = 1.96$

$N \rightarrow \infty$ n/N < 0.05 ve $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$18 - 1.96\frac{3}{\sqrt{256}} < \mu < 18 + 1.96\frac{3}{\sqrt{256}}$$

$$18 - \frac{5.88}{16} < \mu < 18 + \frac{5.88}{16}$$

$$18 - 0.3675 < \mu < 18 + 0.3675$$

$$17.6325 < \mu < 18.3675$$

%95 olasılıkla üretilen mamüllerin boy uzunlukları (anakütle ortalaması) 17.6375 mm. ile 18.3675 mm. arasında olacaktır.

10.6.1.3. Küçük Örneklerde Anakütle Ortalamasının Aralık Tahmini

Anakütle varyansı bilinmiyorsa ve anakütle dağılımı normal dağılım ise veya normal dağılım kabul ediliyorsa, örnekleme dağılımı t dağılımı olacağından $Z_{\alpha/2}$ değerini belirlemek için normal dağılım yerine t dağılımı kullanılır.

n ≥30 ise, t dağılımının normal dağılıma yaklaştığı görülür. Bu nedenle, örnek ortalamaları dağılımının varyansının bilinmediği ve n <30 olduğu durumda örnekleme dağılımının t dağılımı olduğunu söyleyebiliriz. Bu durumda $Z_{\alpha/2}$ değerleri yerine t- dağılımı tablosundan yararlanarak $t_{\alpha/2}$ değerleri belirlenecektir. Aralık tahmini benzer şekilde,

$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1-\alpha$$

$$-t_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < +t_{\alpha/2}$$

$$\overline{X} - t_{\alpha/2}\frac{\sigma_{\overline{X}}}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{\sigma_{\overline{X}}}{\sqrt{n}}$$

olarak yapılacaktır.

ÖRNEK: Normal dağıldığı bilinen bir anakütleden 16 birimli bir örnek alınmış ve örnek ortalaması $\overline{X}$ =35 ve S=4.6 bulunmuştur. 0.95 olasılıkla anakütle ortalamasını tahmin ediniz.

t dağılımı tablosundan,

$\alpha/2 = 0.05/2 = 0.025$ ve

serbestlik derecesi n-1 = 16-1=15 karşılığı $t_{\alpha/2}$ =2.131

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$35 - 2.131\frac{4.6}{\sqrt{16}} < \mu < 35 + 2.131\frac{4.6}{\sqrt{16}}$$

$$32.55 < \mu < 37.45$$

olarak elde edilir.

10.6.2. Anakütle Oranının Tahmini

Anakütle ortalaması gibi anakütle oranında aralık tahmini yapılabilir. Bunun için oranların örnekleme dağılımının bilinmesi gerekmektedir.

10.6.2.1. Oranların Örnekleme Dağılımı

Birim sayısı N olan anakütlenin oranını P ve bu anakütleden alınan n birimli örneğin oranını ise p ile ifade edelim. Anakütle oranı P, 0 veya 1'e yakın değilse ve anakütleden alınan örnek büyükse; örnek oranlarının dağılımının, ortalaması P ve standart sapması $\sigma_p = \sqrt{\frac{p.q}{n}}$ olan normal dağılıma yakındır.

Bu durumda standart değişken,

$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{pq}{n}}}$$

olacaktır. Anakütle oranının nokta tahmini ise P=p olarak yapılacaktır. Yani anakütle oranı P'nin tahmincisi, örnek oranı p'dir.

Anakütle oranının aralık tahmininde anakütle oranı P bilinmeyeceğinden σ_p'nin bilinmesi sözkonusu değildir. Seçimin iadeli ve iadesiz olmasına ve örnekleme oranına göre oranların standart hatası,

-iadeli seçim ve iadesiz seçimde n/N <0.05 ise

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

-iadesiz seçimde n/N ≥0.05 ise

$$\sigma_p = \sqrt{\frac{pq}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

olacaktır.

Anakütle oranı bilinmediğinden σ_p yerine örnekten tahmin edilecek S_p'nin kullanılması gerekecektir.

$$S_p^2 = \frac{n\sigma_p^2}{n-1} = \frac{n.\frac{pq}{n}}{n-1} = \frac{pq}{n-1}$$

ve

$$S_p = \sqrt{\frac{pq}{n-1}}$$

olarak tahmin edilecektir. Uygulamada n büyük ise $\sqrt{\frac{pq}{n}}$ ile $\sqrt{\frac{pq}{n-1}}$ değerleri birbirine çok yakın olacağından $\sigma_p \cong S_p$ kabul edilebilir. Verdiğimiz örneklerde de $\sigma_p \cong S_p$ kabul edilmiştir.

10.6.2.2. Aralık Tahmini

Oranların aralık tahmininde de ortalamaların aralık tahmininde olduğu gibi standart değişken için,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

olacak şekilde anakütle oranının aralık tahmini yapılacaktır. Anakütle oranının aralık tahmini,

$$P(-Z_{\alpha/2} < \frac{p-P}{\sqrt{\sigma_p}} < +Z_{\alpha/2}) = 1-\alpha$$

$$p - Z_{\alpha/2}\sqrt{\sigma_p} < P < p + Z_{\alpha/2}\sqrt{\sigma_p}$$

olarak yapılır.

ÖRNEK: Bir işletmede çalışanların iş yerine zamanında gelip gelmedikleri araştırılmaktadır. Bu nedenle işletmede görev yapan 10.000 kişiden 200 kişilik bir örnek tesadüfi olarak seçilmiş ve çalışanların 0.67'sinin son 1 ay içinde görevlerine hiç geç kalmadıkları anlaşılmıştır. 0.95 olasılıkla tüm işletmede çalışanlardan iş yerine zamanında gelenlerin oranını tahmin ediniz.

$$Z_{\varepsilon/2} = 1.96$$

$$\frac{n}{N} = \frac{200}{10000} = 0.02 < 0.05 \rightarrow \sigma_p = \sqrt{\frac{pq}{n}}$$

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < P < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$0{,}67 - 1{,}96\sqrt{\frac{0{,}67.0{,}33}{200}} < P < 0{,}67 + 1{,}96\sqrt{\frac{0{,}67.0{,}33}{200}}$$

$$0{,}67 - 1{,}96.0{,}033 < P < 0{,}67 + 1{,}67.0{,}033$$

$$0{,}67\text{-}0{,}064 < P < 0{,}67+0{,}064$$

$$0{,}606 < P < 0{,}734$$

0.95 olasılıkla tüm işletmede zamanında iş yerine gelenlerin oranı 0,606 ile 0,734 arasındadır.

10.6.3. Ortalamalar Arasındaki Farkların Aralık Tahmini

Farklı iki anakütlenin ortalamaları arasındaki farkın aralık tahmini yapılabilir. Bu nedenle ortalamalar arasındaki farkların örnekleme dağılımının bilinmesi gereklidir.

10.6.3.1. Ortalama Farklarının Örnekleme Dağılımı

Birim sayıları N_1 ve N_2 ve ortalamaları μ_1 ve μ_2 olan iki anakütleden alınacak örneklerin birim sayıları n_1 ve n_2 ; ortalamaları $\overline{X}_1$ ve $\overline{X}_2$ olsun. Anakütle ortalamaları arasındaki fark (μ_1 - μ_2), örnek ortalamaları arasındaki fark ($\overline{X}_1$ - $\overline{X}_2$)'a eşit olarak tahmin edilir.

μ_1 - μ_2 = $\overline{X}_1$ - $\overline{X}_2$

Tahmin edilen bu fark örnekleme sapması içerdiğinden aralık tahmini yapılabilir. Ortalamalar arasındaki farkların dağılımı için standart değişken

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

dır.

Ortalamalar arasındaki farkların dağılımlarının standart sapması yani standart hata $(\sigma_{\overline{X}_1 - \overline{X}_2})$ şöyle belirlenir.

-anakütlenin standart sapmaları biliniyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

-anakütlenin standart sapmaları bilinmiyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Bazı durumlarda anakütle varyansları birbirine eşit olabilir. $\sigma_1^2 = \sigma_2^2$ ise ve varyans biliniyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sigma.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

varyans bilinmiyorsa,

$$\sigma_{\bar{X}_1-\bar{X}_2} = S.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = S\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

olacaktır.

Örnek birim sayılarının ($n_1+n_2 <120$) olması durumunda anakütle varyanslarının değeri bilinmiyor, fakat anakütle dağılımının normal olduğu ve varyansların birbirine eşit olduğu biliniyorsa,

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

olarak tahmin edilir. Bu durumda n_1 veya n_2'den biri veya ikisi de 30'dan küçük olacaktır. Dağılımın serbestlik derecesi (n_1+n_2-2)'dir. 30'dan büyük serbestlik derecesi için t- dağılımı normal dağılıma yaklaşacağından, bu durumda örnekleme dağılımı olarak kabul edilebilir. Yukarıda verilen formül özellikle aynı anakütleden iki örnek alınması durumunda geçerli olacaktır.

10.6.3.2. Aralık Tahmini

Ortalamaların ve oranların aralık tahmininde olduğu gibi, ortalama farklarının aralık tahmininde de benzer işlemler tekrarlanacaktır. Ortalama farklarının aralık tahmini,

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2} < \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}} < Z_{\alpha/2}) = 1-\alpha$$

$$(\bar{X}_1-\bar{X}_2) - Z_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2} < (\mu_1-\mu_2) < (\bar{X}_1-\bar{X}_2) + Z_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2}$$

olarak yapılacaktır.

ÖRNEK: İki farklı anakütleden alınan iki örnek ile ilgili bilgiler aşağıda verilmiştir.

$n_1 = 100 \qquad \overline{X}_1 = 42 \qquad S_1 = 6$

$n_2 = 120 \qquad \overline{X}_2 = 35 \qquad S_2 = 5$

0.05 hata payı ile bu örneklerin alındığı anakütlelerin ortalamaları arasındaki farkı tahmin ediniz.

$$Z_{\varepsilon/2} = 1.96$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(\overline{X}_1 - \overline{X}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\overline{X}_1 - \overline{X}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(42-35) - 1{,}96\sqrt{\frac{(6)^2}{100} + \frac{(5)^2}{120}} < (\mu_1 - \mu_2) < (42-35) + 1{,}96\sqrt{\frac{(6)^2}{100} + \frac{(5)^2}{120}}$$

$$7 - 1{,}96(0{,}7538) < (\mu_1 - \mu_2) < 7 + 1{,}96(0{,}7538)$$

$$7 - 1{,}4774 < (\mu_1 - \mu_2) < 7 + 1{,}4774$$

$$5{,}5226 < (\mu_1 - \mu_2) < 8{,}4774$$

10.6.4. Oranlar Arasındaki Farkların Aralık Tahmini

İki anakütle ortalamaları arasındaki farkın aralık tahmini gibi iki anakütlenin oranları arasındaki farkların da aralık tahmini yapılabilir. Ancak burada önce iki oran arasındaki farkın örnekleme dağılımı açıklanacaktır.

10.6.4.1. Oran Farklarının Örnekleme Dağılımı

Birim sayıları N_1 ve N_2, oranları P_1 ve P_2 olan iki anakütleden alınacak örneklerin birim sayıları n_1 ve n_2, oranları p_1 ve p_2 olsun. Anakütle oranları arasındaki fark (P_1-P_2), örnek oranları arasındaki farka (p_1-p_2) eşit olarak tahmin edilir.

$P_1 - P_2 = p_1 - p_2$

Oranlar arasındaki farkların dağılımı için standart değişken,

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1 - p_2}}$$

olacaktır. Oran farklarının dağılımının standart hatası ise,

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

olarak belirlenir.

10.6.4.2. Aralık Tahmini

Oranların aralık tahmini formülü de standart değişkenin genel ifade de yerine konması ile elde edilecektir.

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2} < \frac{(p_1 - p_2)-(P_1 - P_2)}{\sigma_{p_1-p_2}} < Z_{\alpha/2}) = 1-\alpha$$

$$\text{P } (p_1 - p_2) - Z_{\alpha/2}\sigma_{p_1-p_2} < (P_1 - P_2) < (p_1 - p_2) + Z_{\alpha/2}\sigma_{p_1-p_2}) = 1-\alpha$$

Yukarıda verilen formül ile iki anakütle oranının farkları belirlenen hata payı veya güven olasılığı ile tahmin edilmiş olacaktır.

ÖRNEK: İki farklı anakütleden alınan iki örnek ile ilgili bilgiler aşağıda verilmiştir.

$n_1 = 1250$ $\quad p_1 = 0{,}25$

$n_2 = 1500$ $\quad p_2 = 0{,}22$

0,95 güven olasılığı ile bu örneklerin alındığı anakütlelerin oranları arasındaki farkı tahmin ediniz.

$$Z_{\varepsilon/2} = 1.96$$

$q_1 = 1-p_1 = 1-0{,}25 = 0{,}75$

$q_2 = 1-p_2 = 1-0{,}22 = 0{,}78$

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$= \sqrt{\frac{(0{,}25)(0{,}75)}{1250} + \frac{(0{,}22)(0{,}78)}{1500}} = 0{,}01626$$

$$(p_1 - p_2) - Z_{\alpha/2}\sigma_{p_1-p_2} < (P_1 - P_2) < (p_1 - p_2) + Z_{\alpha/2}\sigma_{p_1-p_2}$$

$$(0{,}25\text{-}0{,}22)\text{-}1{,}96(0{,}01626) < (P_1\text{-}P_2) < (0{,}25\text{-}0{,}22)+1{,}96(0{,}01626)$$

$$0{,}03\text{-}0{,}032 < (P_1\text{-}P_2) < 0{,}03+0{,}032$$

$$-0{,}002 < (P_1\text{-}P_2) < 0{,}062$$

10.6.5. Anakütle Varyansının Aralık Tahmini

Anakütle ortalaması ve oranı gibi anakütle varyansının da aralık tahmini yapılabilir. Bu durumda anakütle varyansının aralık tahmininde S^2'den yararlanılacaktır.

Ortalaması μ ve varyansı σ^2 olan bir anakütleden alınacak n birimli örneğin ortalaması $\overline{X}$ ve varyansı S^2 olsun. $\overline{X}$ ve S^2'nin bağımsız olması durumunda,

$$\frac{(n-1)S^2}{\sigma^2}$$

değişkeninin dağılımı (n-1) serbestlik dereceli ki-kare dağılımıdır. Bundan yararlanarak anakütle varyansının aralık tahmini,

$$P(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

olarak elde edilir.

ÖRNEK: Yeni piyasaya sürülen bir ürün nüfusları birbirine yakın 8 bölgede satışa sunulmuştur. Ürünün aylık satış ortalaması $\overline{X}$ =20500 ve S=1600 olarak hesaplanmıştır. 0,90 güven olasılığı ile anakütle varyansını ve standart sapmasını tahmin ediniz.

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

n =8

serbestlik derecesi = n-1=8-1=7

α=0,10

$\chi^2_{0,10/2,7} = \chi^2_{0,05,7} = 14,067$

$\chi^2_{1-0,10/2,7} = \chi^2_{0,95,7} = 2,167$

$$\frac{(8-1)(1600)^2}{14,067} < \sigma^2 < \frac{(8-1)(1600)^2}{2,167}$$

$$1273903,46 < \sigma^2 < 8269497$$

$$1128,67 < \sigma < 2875,67$$

Anakütle varyans ve standart sapmasının aralık tahmini büyük örnekler için örnekleme dağılımının normal dağılıma yaklaştırılması ile yapılabilir. Büyük örneklerde S'lerin dağılımının standart hatası (σ_s),

$$\sigma_s = \frac{\sigma}{\sqrt{2n}}$$

dir. Buna göre σ'ın güven aralığı ise,

$$S - Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}} < \sigma < S + Z_{\alpha/2}\frac{\sigma}{\sqrt{2n}}$$

olur. Bu eşitsizlikten σ'ın güven aralık tahmini,

$$\frac{S}{1+\frac{Z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{S}{1-\frac{Z_{\alpha/2}}{\sqrt{2n}}}$$

olarak belirlenebilir.

ÖRNEK: 100 birimli bir örnekte S=12 olarak bulunmuştur. 0,05 hata payı ile anakütle standart sapmasını tahmin edin.

$$\frac{S}{1+\frac{Z_{\alpha/2}}{\sqrt{2n}}} < \sigma < \frac{S}{1-\frac{Z_{\alpha/2}}{\sqrt{2n}}}$$

$$\frac{12}{1+\frac{1,96}{\sqrt{2,100}}} < \sigma < \frac{12}{1-\frac{1,96}{\sqrt{2,100}}}$$

$$10,539 < \sigma < 13,930$$

10.6.6. Varyans Oranlarının Aralık Tahmini

İki anakütle ortalaması veya iki anakütle oranı arasındaki fark gibi, iki anakütle varyansının birbirine oranının aralık tahmini yapılabilir. Bu durumda varyans oranlarının örnekleme dağılımı F dağılımıdır.

Varyansları σ_1^2 ve σ_2^2 olan normal dağılmış iki anakütleden alınan n_1 ve n_2 birimli örneklerin varyansları S_1^2 ve S_2^2 olsun. Bu durumda S^2'lerin σ^2'lere oranlarının, birbirine oranının dağılımı (n_1-1) ve (n_2-1) serbestlik dereceli F dağılımıdır.

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

Varyans oranlarının aralık tahmini,

$$P(F_{1-\alpha/2,n_1-1,n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < F_{\alpha/2,n_1-1,n_2-1}) = 1-\alpha$$

olarak yapılacaktır. F tablo değerlerinin bulunmasında,

$$F_{1-\alpha/2,n_1-1,n_2-1} = \frac{1}{F_{\alpha/2,n_2-1,n_1-1}}$$

olduğundan bu ilişkiden yararlanılacaktır. Bu dönüşüm ile varyans oranlarının aralık tahmini formülü,

$$\frac{S_1^2}{S_2^2}.\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}.F_{\alpha/2,n_2-1,n_1-1}$$

olarak elde edilir.

ÖRNEK: A marka mallardan tesadüfi olarak 10 adet alınarak $\overline{X}_A = 1{,}5$ ve S_A=0,4 ; B marka mallardan tesadüfi olarak 11 adet alınarak $\overline{X}_B = 2{,}0$ ve S_B=0,6 olarak hesaplanmıştır. 0,98 güven olasılığı ile anakütle varyanslarının ve standart sapmalarının oranlarını tahmin ediniz.

n_1 =10

n_2 =11 α=0,02

$F_{0,02/2,10-1,11-1} = 4{,}94$

$F_{0,02/2,11-1,10-1} = 5{,}26$

$$\frac{S_1^2}{S_2^2}.\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}.F_{\alpha/2,n_2-1,n_1-1}$$

$$\frac{(0,4)^2}{(0,6)^2}.\frac{1}{4{,}94} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{(0,4)^2}{(0,6)^2}.5{,}26$$

$$0{,}09 < \frac{\sigma_1^2}{\sigma_2^2} < 2{,}34$$

$$0{,}30 < \frac{\sigma_1}{\sigma_2} < 1{,}53$$

10.6.7. Optimum Örnek Büyüklüğünün Belirlenmesi

Örneklemenin özellikle maliyet ve zaman yönünden tasarruf sağladığını belirtmiştik. Örnek düzenlenirken örnek büyüklüğünün yeterli olması da aynı nedenlerden ötürü önemlidir. Örnek birim sayısı yetersiz ise, yapılan araştırmadan beklenen

bilgiler sağlıklı olarak elde edilemeyeceği gibi, örnek birim sayısının fazla olması maliyetin artmasına ve zaman kaybına neden olacaktır.

Örneğe alınacak en az birim sayısını bazı şartlarla, matematiksel olarak belirleyebiliriz. Belirleyeceğimiz bu birim sayısını optimum örnek büyüklüğü olarak adlandıracağız.

10.6.7.1. Anakütle Ortalamasının Tahmini İçin Optimum Örnek Büyüklüğünün Belirlenmesi

Anakütle ortalamasının aralık tahmini,

$$P(-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} < +Z_{\alpha/2}) = 1 - \alpha$$

şeklinde yapılıyordu. Buradan,

$$-Z_{\alpha/2}\sigma_{\overline{X}} < \overline{X} - \mu < +Z_{\alpha/2}\sigma_{\overline{X}}$$

eşitliğini elde ederiz. $\overline{X} - \mu$ farkını E ile gösterirsek,

$$E = \overline{X} - \mu$$

$$E < Z_{\alpha/2}\sigma_{\overline{X}}$$

yazabiliriz. Burada hata miktarı E'nin en çok belirlenen değer kadar veya daha az olması için örnek büyüklüğünü belirleyebiliriz. Hatanın en çok E değeri kadar olması için,

$$E = Z_{\alpha/2}\sigma_{\overline{X}}$$

olacaktır. $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ olduğundan,

$$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{Z_{\alpha/2}\sigma}{E}$$

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$$

olacaktır. Bu eşitlikte n'in belirlenebilmesi için anakütle varyansının bilinmesi, hata payının ve hata miktarı E'nin belirlenmesi gerekmektedir. σ^2 bilinmiyor, fakat yaklaşık olarak tahmin edilebiliyorsa örnek büyüklüğü buna dayanarak da belirlenebilir. E anakütle ortalaması ile örnek ortalaması arasında olması istenen maksimum farkı belirtecektir.

ÖRNEK: σ =8 olan bir anakütleden n birimli bir örnek alınacaktır. Örnek ortalaması ile anakütle ortalaması arasındaki farkın 0,95 olasılıkla en çok 1,5 olması için n ne olmalıdır, hesaplayınız.

$$Z_{\varepsilon/2} = 1.96$$

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E} \right)^2$$

$$n = \left(\frac{1,96.8}{1,5} \right)^2 = 109,27 \cong 109 \; birim$$

n ≥109 birim alınırsa, $\overline{X}$ *ve* μ arasındaki fark 0,95 olasılıkla ±1,5 veya daha az olacaktır.

10.6.7.2. Anakütle Oranının Tahmini İçin Optimum Örnek Büyüklüğünün Belirlenmesi

Aynı şekilde p-P = E dersek,

$E = Z_{\alpha/2}\sigma_p$ ve

$$n = \frac{Z_{\alpha/2}^2 pq}{E^2}$$

olacaktır.

Oranlarla ilgili örnek büyüklüğü belirlenirken, P bilinmediğinden tahmin edilmesi gerekecektir. p tahmin edilirken tahmin belirli bir aralıkta yapılıyorsa (%15-20 gibi) n'i belirlemek için tahmin edilen büyük oranın alınması iyi olacaktır. P tahmin edildikten sonra $Z_{\alpha/2}$ güven olasılığına ve E araştırmaya göre belirlenebilir.

ÖRNEK: Bir bölgede A malı kullanan ailelerin oranı araştırılmaktadır. Bölgede A malı kullanan ailelerin toplam ailelere oranının 0,40 olduğu tahmin edilmektedir. 0,90 olasılıkla örnek oranı ile anakütle oranı arasındaki farkın 0,04 olması için örneğe alınacak aile sayısını belirleyiniz.

$$Z_{\varepsilon/2} = 1.64$$

$$n = \frac{Z_{\alpha/2}^2 pq}{E^2} = \frac{(1,64)^2 (0,4)(0,6)}{(0,04)^2} = 403,44$$

$$n \cong 403$$

örnek en az 403 aileden oluşmalıdır.

10.7. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1 : σ=50 olan 5000 birimli bir anakütleden 400 birimden oluşan örnek iadesiz seçimle düzenlenerek, örnek ortalaması 4500 olarak bulunmuştur. 0,10 hata payı ile örnek ortalamasını tahmin ediniz.

Çözüm :

N =5000

n=400

σ=50 $\quad Z_{\alpha/2} \rightarrow 1-\alpha = 1-0{,}10 = 0{,}90$

$\overline{X} = 4500 \quad 0{,}90/2 = 0{,}4500 \rightarrow$ tablodan $Z_{\alpha/2} = 1{,}64$

α=0,10 $\quad \frac{n}{N} = \frac{400}{5000} = 0{,}08 > 0{,}05$ olduğundan

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$4500 - 1{,}64\frac{50}{\sqrt{400}}\sqrt{\frac{5000-400}{5000-1}} < \mu < 4500 + 1{,}64\frac{50}{\sqrt{400}}\sqrt{\frac{5000-400}{5000-1}}$$

$$4500 - 1{,}64.\frac{50}{20}.0{,}95 < \mu < 4500 + 1{,}64.\frac{50}{20}.0{,}95$$

$$4500 - 3{,}895 < \mu < 4500 + 3{,}895$$

$$4496{,}105 < \mu < 4503{,}895$$

0.10 hata payı ile örnek ortalaması bu aralık içinde olacaktır.

Not: Seçim iadeli veya örnekleme oranı 0,05 'ten küçük olsaydı, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ alınacaktı.

ÖRNEK 2: Bir depoda bulunan 10000 mamülün ağırlıkları incelenmektedir. Bu mamüllerden 36 birimli bir örnek düzenlenmiş (iadesiz seçimle) ve $\overline{X}$ =20 gr. ve S=4 gr. bulunmuştur. 0,70 olasılıkla ambarda bulunan mamüllerin ortalamasını tahmin ediniz.

Çözüm:

$Z_{\alpha/2} \rightarrow 1-\alpha = 0{,}70/2 = 0{,}3500 \rightarrow$ tablodan $Z_{\alpha/2} = 1{,}04$

$$\frac{n}{N} = \frac{36}{10000} = 0{,}0036 < 0{,}05 \rightarrow \quad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$$

$$\bar{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$20 - 1{,}04\frac{4}{\sqrt{36}} < \mu < 20 + 1{,}04\frac{4}{\sqrt{36}}$$

$$20 - 0{,}693 < \mu < 20 + 0{,}693$$

$$19{,}307 < \mu < 20{,}693$$

0,70 olasılıkla ambarda bulunan mamüllerin ortalaması 19,3 gr. ile 20,6 gr. arasında olacaktır.

ÖRNEK 3: 500 birimden oluşan bir topluluktan 49 birimlik bir örnek iadesiz seçimle alınmıştır. Örnek ortalaması $\bar{X}$ =98 ve örnek standart sapması $\hat{\sigma}$ =8 bulunduğuna göre, 0,01 hata payı ile anakütle ortalamasını tahmin ediniz.

Çözüm:

$Z_{\alpha/2} \rightarrow$ 1-α =1-0,01 =0,99

0,99/2 = 0,4950$\rightarrow$ tablodan $Z_{\alpha/2}$ =2,57

$\frac{n}{N} = \frac{49}{500} = 0{,}098 > 0{,}05$ olduğundan ve σ bilinmediğinden ,

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1} = \frac{n\hat{\sigma}^2}{n-1}$$

$$S = \sigma\sqrt{\frac{n}{n-1}} = 8\sqrt{\frac{49}{49-1}} = 8.1{,}01 = 8{,}08$$

$$\bar{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} < \mu < \bar{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{8{,}08}{\sqrt{49}}\sqrt{\frac{500-49}{500-1}} = 1{,}096$$

$$98 - 2{,}57.1{,}096 < \mu < 98 + 2{,}57.1{,}096$$

$$98 - 2{,}81 < \mu < 98 + 2{,}81$$

$$95{,}19 < \mu < 100{,}81$$

0,01 hata payı veya 0,99 güven olasılığı ile anakütle ortalaması 95,19 ile 100,81 birim arasında olacaktır.

Aynı örnek iadeli seçimle düzenlenseydi veya n/N<0,05 olsaydı,

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{8,08}{\sqrt{49}} = 1,15$$

$$98 - 2,57.1,15 < \mu < 98 + 2,57.1,15$$

$$95,05 < \mu < 100,95$$

olacaktı.

ÖRNEK 4: N=5000, n =49, $\sum_{i=1}^{49} X_i = 2940$ ve α=0,05 olduğuna göre μ'nün güven aralığını bulunuz. (σ =7)

Çözüm:

$Z_{\alpha/2}$ =2,57

$$\frac{n}{N} = \frac{49}{5000} < 0,05$$

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{2940}{49} = 60$$

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$60 - 1,96\frac{7}{\sqrt{49}} < \mu < 60 + 1,96\frac{7}{\sqrt{49}}$$

$$58,04 < \mu < 61,96$$

ÖRNEK 5: Bir önceki örnekte N=500 olduğuna göre anakütle ortalamasını tahmin ediniz.

$$\frac{n}{N} = \frac{49}{500} = 0,098 > 0,05$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{7}{\sqrt{49}}\sqrt{\frac{500-49}{500-1}} = 0,95$$

$$\bar{X} - Z_{\alpha/2}\sigma_{\bar{X}} < \mu < \bar{X} + Z_{\alpha/2}\sigma_{\bar{X}}$$

$$60 - 1,96.0,95 < \mu < 60 + 1,96.0,95$$

$$58,138 < \mu < 61,862$$

ÖRNEK 6: 6000 birimli bir anakütleden 64 birimli bir örnek iadesiz seçimle alınarak,

$$\sum_{i=1}^{64} X_i = 3200$$

$$\sum_{i=1}^{64}(X_i - \overline{X})^2 = 1575$$

bulunmuştur. Anakütle ortalamasını 0,99 olasılıkla tahmin ediniz.

Çözüm:

$Z_{\alpha/2} = 2{,}57$

$$\frac{n}{N} = \frac{64}{6000} < 0{,}05$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{3200}{64} = 50$$

$$S = \sqrt{\frac{(X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{1575}{64-1}} = \sqrt{25} = 5$$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$50 - 2{,}57\frac{5}{\sqrt{64}} < \mu < 50 + 2{,}57\frac{5}{\sqrt{64}}$$

$$48{,}4 < \mu < 51{,}6$$

ÖRNEK 7: Bir anakütleden alınan 200 birimli örnek için şu değerler hesaplanmıştır.

$$\sum_{i=1}^{200} X_i = 1500$$

$$\sum_{i=1}^{200} X_i^{\,2} = 12500$$

anakütle ortalamasını 0,80 güven olasılığı ile tahmin ediniz.

Çözüm:

$Z_{\alpha/2} \rightarrow$ 1-α = 0,80/2 = 0,4$\rightarrow$ tablodan $Z_{\alpha/2}$ =1,28

$$\frac{n}{N} < 0{,}05 \rightarrow \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{1500}{200} = 7{,}5$$

$$\hat{\sigma} = \sqrt{\frac{\Sigma X_i^2}{n} - (\overline{X})^2} = \sqrt{\frac{12500}{200} - (7{,}5)^2} = \sqrt{62{,}5 - 56{,}25}$$

$$= \sqrt{6{,}25} = 2{,}5$$

$$S = \hat{\sigma}\sqrt{\frac{n}{n-1}} = 2{,}5\sqrt{\frac{200}{200-1}} = 2{,}5.1{,}002 = 2{,}505$$

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$7{,}5 - 1{,}28\frac{2{,}505}{\sqrt{200}} < \mu < 7{,}5 + 1{,}28\frac{2{,}505}{\sqrt{200}}$$

$$7{,}5 - 0{,}22 < \mu < 7{,}5 + 0{,}22$$

$$7{,}28 < \mu < 7{,}72$$

ÖRNEK 8: 1000 birimli bir anakütleden alınan örnek aşağıda verilmiştir. (Örnek iadesiz seçimle düzenlenmiştir.)

X_i	f_i	f_iX_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$
2	11	22	-2	4	44
3	26	78	-1	1	26
4	45	180	0	0	0
5	30	150	1	1	30
6	9	54	2	4	36
	121	484			136

0,9544 güven olasılığı ile bu örneğin alındığı anakütle ortalamasını tahmin ediniz.

Çözüm:

$Z_{\alpha/2} \rightarrow$ 1-α = 0,9544/2 = 0,4772$\rightarrow$ tablodan $Z_{\alpha/2}$ =2

$$\frac{n}{N} = \frac{121}{1000} = 0{,}121 > 0{,}05 \rightarrow \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$\overline{X} = \frac{\Sigma f_i X_i}{\Sigma f_i} = \frac{484}{121} = 4$$

$$S = \sqrt{\frac{\Sigma f_i (X_i - \overline{X})^2}{\Sigma f_i - 1}} = \sqrt{\frac{136}{121-1}} = \sqrt{1{,}133} = 1{,}064$$

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1{,}064}{\sqrt{121}}\sqrt{\frac{1000-121}{121-1}} = 0{,}262$$

$$\overline{X} - Z_{\alpha/2}\sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2}\sigma_{\overline{X}}$$

$$4 - 0{,}262 < \mu < 4 + 0{,}262$$

$$3{,}476 < \mu < 4{,}524$$

ÖRNEK 9: Normal dağıldığı bilinen ve varyansı 81 olan bir anakütleden 25 birimli bir örnek alınmıştır. Örnek ortalaması 150 bulunduğuna göre 0,90 olasılıkla anakütle ortalamasını tahmin ediniz.

Çözüm: Anakütle dağılımı normal olduğundan n <30 olsa da normal dağılım tablosundan yararlanılacaktır.

$Z_{\alpha/2} = 1{,}64$

$$\frac{n}{N} < 0{,}05 \rightarrow \qquad \sigma^2 = 81 \rightarrow \sigma = 9 \rightarrow \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$150 - 1{,}64\frac{9}{\sqrt{81}} < \mu < 150 + 1{,}64\frac{9}{\sqrt{81}}$$

$$148{,}36 < \mu < 151{,}64$$

ÖRNEK 10: A makinesinin ürettiği hatalı mamüllerin oranı araştırılmaktadır. Bu amaçla makinenin ürettiği 2000 mamülden 180 adedi tesadüfi olarak seçilerek, bunlardan 27 adedinin hatalı olduğu anlaşılmıştır. 0,01 hata payı ile makinenin ürettiği hatalı mamül oranını tahmin ediniz. (Seçim iadesizdir.)

Çözüm:

$Z_{\alpha/2} = 2{,}57$

$$\frac{n}{N} = \frac{180}{2000} = 0{,}09 > 0{,}05 \rightarrow \sigma_p = \sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}} < P < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}}$$

$$P = \frac{27}{180} = 0{,}15$$

$$\sigma_p = \sqrt{\frac{pq}{n}}.\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0{,}15.0{,}85}{180}}.\sqrt{\frac{2000-180}{2000-1}} = 0{,}025$$

$$0{,}15\text{-}2{,}57.0{,}025 < P < 0{,}15+2{,}57.0{,}025$$

$$0{,}085 < P < 0{,}21$$

ÖRNEK 11: Varyanslarının birbirine eşit olduğu bilinen iki anakütleden alınan örnekler ile ilgili bilgiler aşağıda verilmiştir.

$n_1 = 12 \qquad \bar{X}_1 = 158 \qquad S_1 = 10$

$n_2 = 15 \qquad \bar{X}_2 = 141 \qquad S_2 = 14$

0,95 güven olasılığı ile ortalamalar arasındaki farkı tahmin ediniz.

Çözüm:

$n_1 < 30$, $n_2 < 30$ olduğundan n_1+n_2-2 serbestlik dereceli t dağılımı, $t_{0,05/2,12+15-2} = t_{0,025,25}=2,060$

$$\sigma_1^2 = \sigma_2^2$$

olduğundan,

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(12-1)(10)^2 + (15-1)(14)^2}{12+15-2}} = 12,4$$

$$\sigma_{\bar{X}_1-\bar{X}_2} = S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= 12,4\sqrt{\frac{1}{12}+\frac{1}{15}} = 4,8025$$

$$(\bar{X}_1-\bar{X}_2) - t_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2} < (\mu_1-\mu_2) < (\bar{X}_1-\bar{X}_2) + t_{\alpha/2}\sigma_{\bar{X}_1-\bar{X}_2}$$

$$(158\text{-}141)\text{-}2,060(4,8025) < (\mu_1-\mu_2) < (158\text{-}141)+2,060(4,8025)$$

$$7,11 < (\mu_1-\mu_2) < 26,89$$

BÖLÜM 11

HİPOTEZ TESTLERİ

Önceki bölümde çeşitli parametrelerin örnekleme dağılımlarını ve bunlara dayanarak aralık tahminlerini açıkladık. Bu bölümde açıklayacağımız hipotez testlerinde de aynı örnekleme dağılımlarını kullanacağız.

11.1. HİPOTEZ VE HİPOTEZ TESTİ KAVRAMLARI

Uygulamada anakütle parametreleri ile ilgili bazı varsayımların veya bilgilerin ortaya konduğu durumlarla karşılaşılabilir. Bu varsayımlar veya bilgiler anakütle ortalaması, oranı veya varyansının herhangi bir sayısal değere eşitlik veya eşitsizliğini ifade edebileceği gibi, iki farklı anakütlenin parametrelerinin birbirine eşit veya farklı olduğunun öne sürülmesi şeklinde de karşımıza çıkabilir. Bu durumda ileri sürülen varsayımın veya bilginin doğrulanması gerekecektir.

Bu soruna bir örnek yardımı ile cevap verilebilir. Eğer örnekten tahmin edilen değer ile ileri sürülen değer aynı ise bir sorun yoktur. Buradaki aynı kelimesi tam olarak sayısal eşitliği ifade etmektedir. Sayısal değerler arasında fark olması durumunda ise varsayım yanlıştır demek doğru mudur sorusunu sormak gerekir. Bilindiği gibi örneklemeden doğan bir sapma sözkonusudur. Yani, örnekleme hatası nedeniyle parametrenin tahmin edilen değeri kontrol edilebilir bir miktarda farklı olabilir. Dikkat edilirse burada söz edilen fark, aralık tahmini ile tahmin edilmeye çalışılan farka benzemektedir. Aralık tahmininde fark tahmin edilmeye, yani anakütle parametresinde oluşabilecek sapma tahmin edilmeye çalışılmaktadır. Söz ettiğimiz durumda ise varsayılan bir değer ile örnekten elde edilen bir değer vardır ve bu değerler arasındaki farkın örneklemeden kaynaklanıp kaynaklanmadığının araştırılması sözkonusudur. Diğer bir ifade ile aradaki farkın istatistiksel olarak anlamlı olup olmadığına karar verilmesi sözkonusudur. Bu amaçla hipotez testi olarak adlandırılan işlemler ile değerler arasındaki farkın istatistiksel olarak anlamlılığı araştırılmaktadır.

Hipotez testini, anakütle ile ilgili bilgiler veya varsayımlar ile bunların örnekten tahmin edilen değerleri arasındaki farkın istatistiksel olarak anlamlı olup olmadığına karar vermek için yapılan işlemler olarak tanımlanabilir. Buradaki varsayımlar veya bilgiler, kişisel görüşler olabileceği gibi yapılan başka araştırmalardan elde edilmiş veya başka kaynaklardan alınmış bilgiler de olabilir.

Hipotez testleri, örnekten elde edilecek bilgilerle anakütleye ait bilgilerin karşılaştırılması; örneklerden tahmin edilen iki anakütleye ait bilgilerin karşılaştırılması gibi amaçlarla uygulanacak ve yanlış karar alınmasını önleyici bir kriter sağlayacaktır. Hipotez testlerinin, bu nedenle karar almada önemli yeri vardır. Test sonuçlarının, tahminde olduğu gibi doğruluğunu saptamak güç ise de alınacak kararlarda yardımcı olacağı ve yanlış karar alınmasını önleyeceği unutulmamalıdır.

Bir makinenin ayarının bozulup bozulmadığının, bir fabrikada üretiminin kalitesinde değişme olup olmadığının, yeni geliştirilen bir mamülün kullanılan ile farkı olup olmadığının araştırılması gibi pek çok durumda örnek alınarak anakütlenin bilinen değerleri ile karşılaştırılması gerekebilir. Anakütle parametresinin değeri ile örnekten tahmin edilen değer arasında fark varsa, bunun anakütle parametresinin değeri ile örnekten tahmin edilen değer arasındaki farklılıktan mı kaynaklandığı, yoksa tesadüfi olarak örneklemeden mi kaynaklandığı test ile belirlenebilir. Benzer şekilde, karar almak için iki anakütle parametre tahmininin karşılaştırılması da gerekli olabilir. Örneğin, A ve B gibi iki fabrika aynı malı üretiyorsa, üretimle ilgili herhangi bir özelliği karşılaştırılmak için, her ikisinden de birer örnek alınarak, örnekten tahmin edilen değerler arasında fark varsa, bu farkın A ve B fabrikalarının üretim yapılarındaki özelliklerden mi, yoksa tesadüfi olarak mı oluştuğunu belirlemek için bir test yapılabilir.

Hipotez testlerinde yapılacak ilk işlem karar alınacak olayın yapısına göre hipotezleri ortaya koymaktır. Hipotez, anakütle ile ilgili varsayım veya ortaya konan bilgi olarak tanımlanabilir. Bir testte biri temel, diğeri karşıt veya alternatif olmak üzere iki hipotez tanımlanır. Temel veya sıfır hipotez H_0, karşıt veya alternatif hipotez ise H_1 hipotezi olarak ifade edilecektir. Hipotezler oluşturulurken temel hipotez anakütle ile ilgili varsayımın doğruluğunu ifade ederken, karşıt hipotez varsayımın doğru olmadığını ifade edecektir. Test sonucu bu hipotezlerden biri kabul edilirken diğeri reddedilecektir.

Temel hipotezin kabulü anakütle ile ilgili varsayımın doğruluğunu, diğer bir ifade ile varsayım ile örnekten tahmin edilen parametre değeri arasındaki farkın istatistiksel olarak anlamsız olduğunu ortaya koyacaktır. Karşıt hipotezin kabülü ise söylenenlerin aksini ifade edecektir.

Hipotezlerden birinin kabul edilmesi gerekirken, diğeri kabul edilirse hata yapılmış olunacaktır. H_0 hipotezi doğru iken, H_1 hipotezi kabul edilerek, H_0 hipotezi reddedilirse yapılan hata " I. tip hata"; H_0 hipotezi yanlış iken, H_1 hipotezi yerine H_0 hipotezi kabul edilirse yapılan hata " II. tip hata " olarak adlandırılır.

	Ho Kabul	**Ho Red**
Ho Doğru	Doğru Karar	I.Tip Hata
Ho Yanlış	II.Tip Hata	Doğru Karar

Birinci tip hataya α hatası, ikinci tip hataya ise β hatası adı da verilmektedir.

11.2. HİPOTEZ TESTİNİN AŞAMALARI

Sözedildiği gibi hipotez testi bir dizi işlemler sonucu karar verilmesi için yapılmaktadır. Test edilecek parametrelere göre farklı işlemler uygulansa da, test işleminin temel yapısı aynıdır ve genelleştirilebilir. Dört aşamada genelleştireceğimiz test işleminin aşamaları aşağıda açıklanmıştır.

-Hipotezlerin Oluşturulması – Tek ve Çift Taraflı Testler

Test edilecek anakütle parametresi θ ve bunun ile ilgili varsayımı θ_0 ile ifade edelim. θ anakütle ortalaması veya oranı ise θ_0 bunun ile ilgili varsayılan değeri ifade edecektir. Bu anakütleden alınacak n birimli örnek ile tahmin edilen değer ise $\hat{\theta}$ olacaktır. $\hat{\theta}$, θ'nın tahmincisidir ve burada sayısal değeri ifade ettiği varsayılmıştır. Sonuçta $\hat{\theta}=\theta_0$ ise teste gerek yoktur. Test yapılsa bile sonuçta temel hipotez kabul edilecektir. Fakat $\hat{\theta}\neq\theta_0$ olması durumunda aradaki farkın istatistiksel olarak anlamlı olup olmadığı test edilmelidir. Bu durumda temel hipotez,

$H_0 : \theta = \theta_0$

şeklinde oluşturulacaktır.

Karşıt hipotezde temel hipotez gibi farklı bir sayısal değere eşitliği ifade edecek şekilde oluşturulabilir. Bu tür hipotezlere basit hipotez adı verilmektedir. Burada açıklayacağımız gibi eşitsizliklerle oluşturulan hipotezlere ise karmaşık hipotez denilmektedir.

Karşıt hipotez iki farklı şekilde oluşturulabilir. İlk olarak karşıt hipotez eşitsizlik şeklinde kurulur ki, buna çift taraflı hipotez adı verilmektedir.

$H_1 : \theta \neq \theta_0$

Bu tür hipotez testlerine de çift taraflı hipotez testi denmektedir. Anlaşılacağı gibi "çift taraf" küçüklük ve büyüklük ayrımı yapılmadan ikisinin birlikte testinden kaynaklanmaktadır.

Karşıt hipotez büyüklüğü veya küçüklüğü test edecek şekilde de oluşturulabilir. Bu durumda alternatif hipotez tek taraflı hipotez olarak adlandırılır.

$H_1 : \theta > \theta_0$

veya

$H_1 : \theta < \theta_0$

olarak kurulur ve bu tür hipotez testlerine tek taraflı hipotez testleri denmektedir. Karşıt hipotezin büyüklüğü ifade etmesi durumuna sağ taraflı test, küçüklüğü ifade etmesi durumuna ise sol taraflı test adı da verilmektedir.

Çift taraflı testlerde $\hat{\theta}$ ile θ_0 arasındaki ilişki önemli değilken, tek taraflı testlerde önemlidir. $\hat{\theta} > \theta_0$ olması durumunda alternatif hipotezin $\hat{\theta} < \theta_0$ şeklinde oluşturulması anlamsızdır. Yani, büyüklük şeklinde hipotez ancak $\hat{\theta} > \theta_0$ ise kurulmalıdır. Küçüklük içinde aynı açıklamalar geçerlidir.

Aralık tahmininde iki ortalama veya iki oran farkının olduğu gibi birbirinden farklı iki anakütle parametrelerinin aynı olup olmadığı da test edilebilir. Birinci anakütlenin test edilecek parametresi θ_1, ikinci anakütlenin ki θ_2 ise, burada bir varsayım olmayacak, fakat ana kütlelerden alınacak n_1 ve n_2 birimli örneklerden elde edilecek $\hat{\theta}_1$ ve $\hat{\theta}_2$ tahmincilerinin sayısal değerleri ile test yapılacaktır. Nokta tahminin $\theta_1 = \hat{\theta}_1$ ve $\theta_2 = \hat{\theta}_2$ olarak yapılacağı unutulmamalıdır. Bu durumda temel hipotez,

$H_0: \theta_1 = \theta_2$ veya $\theta_1 - \theta_2 = 0$

şeklinde oluşturulur. Alternatif hipotez ise çift taraflı test için,

$H_1: \theta_1 \neq \theta_2$ veya $\theta_1 - \theta_2 \neq 0$

ve tek taraflı test için,

$H_1: \theta_1 > \theta_2$ veya $\theta_1 - \theta_2 > 0$

veya

$H_1: \theta_1 < \theta_2$ veya $\theta_1 - \theta_2 < 0$

şeklinde kurulacaktır. Bu gibi durumlarda alternatif hipotez oluşturulurken $\hat{\theta}_1$ ile $\hat{\theta}_2$ sayısal değerlerinin büyüklüklerine dikkat edilmelidir.

-Hata Payının ve Kritik Değerin Belirlenmesi

Hata payı ve güven olasılığını aralık tahmininden biliyoruz. Hipotez testleri de hata payı ve güven olasılığı ile yapılır. Bu olasılık testin sonucuna güveni ifade etmektedir. Diğer bir ifade ile testin sonucu belirlenen hata payı veya güven olasılığı ile geçerlidir.

Hipotez testlerinde de hata payını α, güven olasılığını da $(1-\alpha)$ olarak ifade edeceğiz. Hipotez testlerinde hata payı α birinci tür hata yapma olasılığıdır ve bu nedenle birinci tür hataya α hatası da denilmektedir. Uygulamada hata payı 0,05 veya 0,01 olarak alınmaktadır. Aslında teorik olarak α veya $(1-\alpha)$ olasılık olduğuna göre testi yapan kişi istediği hata payını seçebilir. Fakat 0,05'ten büyük hata payları sonuçlara güveni sağlamak için uygulamada pek kullanılmamaktadır.

Uygun hata payı belirlendikten sonra testin karar aşamasında kullanılacak tablo değeri de belirlenecektir. Tablo değerine kritik değer adı verilmektedir. Bu değerin belirlenmesi için önce örnekleme dağılımı bilinmelidir. Bu yapılacak testin türüne göre normal dağılımı t dağılımı, F dağılımı veya ki-kare dağılımı olabilir. Bu bölümde yapacağımız açıklamalarda ki-kare dağılımı yer almayacaktır. Örnekleme dağılımına göre düzenlenmiş tablodan seçilen hata payı (α) ve varsa dağılımın serbestlik derecesi de dikkate alınarak kritik değer belirlenir.

-Test İstatistiğinin Hesaplanması

Hipotez testlerinde üçüncü olarak test istatistiği hesaplanır. Test istatistiği test edilecek parametreye ve test edilecek parametre sayısına göre değişir. Örneğin, bir anakütle ortalamasının belirli bir değere eşitliği veya iki anakütle ortalamasının eşitliği test edilirken farklı test istatistikleri kullanılacaktır. Özellikle ortalama ve oranların testinde, örnekleme dağılımı ile ilgili açıklamalarda belirtilen standart hatanın hesaplanması ile ilgili özellikler önemlidir. Standart hatanın doğru olarak hesaplanmasına dikkat edilmelidir.

-Karar

Yapılan test işleminin son adımı hipotezlerden birinin kabul, diğerinin reddedildiği karar aşamasıdır. Kararın verilmesinde hesaplanan test istatistiği ile tablodan bulunan kritik değerler karşılaştırılır. Dağılımın türüne göre kabul ve red bölgeleri oluşturulabilir. Gerekli açıklama daha sonra yapılacaktır. Genelleyerek, test istatistiğinin mutlak değer olarak teorik değerden küçük olması durumunda H_0 hipotezinin, büyük olması durumunda ise H_1 hipotezinin kabul edileceğini söyleyebiliriz.

11.3. ORTALAMALARIN TESTİ

Anakütle ortalaması μ ile ilgili olarak varolan bilgi veya varsayımın doğruluğu, örnekten elde edilecek bilgi ile incelenebilir. Anakütle ortalaması ile ilgili varsayım μ_0 gibi sayısal bir değer ise seçilecek n birimli örnek ortalaması anakütle ortalamasının en iyi tahmincisi olduğundan $\mu = \overline{X}$ olarak tahmin edilecektir. Anakütle ile ilgili varsayımın doğruluğu durumunda $\mu_0 = \overline{X}$ olmalıdır. $\mu_0 \neq \overline{X}$ olması durumunda anakütle ile ilgili varsayımın doğru olmadığı söylenemez. $\overline{X}$ örnekten elde edildiğinden örnekleme hatası söz konusu olabilir. Bu nedenle yapılacak bir test ile aradaki farkın istatistiksel olarak anlamlı olup olmadığı, yani örneklemeden kaynaklanıp kaynaklanmadığı belirlenecektir.

Test için ilk olarak hipotezler belirlenecektir. Yapılacak test çift taraflı da, tek taraflı da olsa temel hipotez,

$$H_0 : \mu = \mu_0$$

şeklinde oluşturulacaktır. Karşıt hipotez ise testin çift veya tek taraflı olmasına göre kurulur. Test çift taraflı ise,

$$H_1 : \mu \neq \mu_0$$

test tek taraflı ise, büyüklük ve küçüklüğe göre,

$$H_1 : \mu < \mu_0$$

veya

$$H_1 : \mu > \mu_0$$

olacaktır.

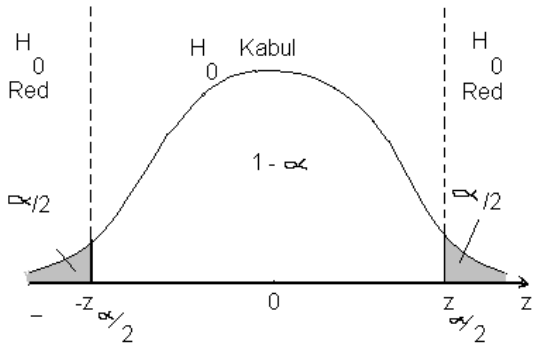
Hipotezler oluşturulduktan sonra α hata payı veya $(1-\alpha)$ güven olasılığı belirlenerek kritik değer bulunacaktır. Kritik değerin belirlenmesinde daha önce ortalamaların aralık tahmininde sözedilen örnekleme dağılımı ile ilgili özellikler dikkate alınır. Anakütle dağılımı normal ise çift taraflı test için $Z_{\alpha/2}$, tek taraflı test için Z_α değeri normal dağılım tablosundan bulunacaktır. Anakütle dağılımı normal dağılım kabul ediliyorsa ve örnek birim sayısı 30'dan küçükse örnekleme dağılımı normal dağılım yerine (n-1) serbestlik dereceli t dağılımı olacaktır. Bu nedenle $Z_{\alpha/2}$ ve Z_α değerleri yerine t dağılımı tablosundan n-1 serbestlik derecesi ile testin çift veya tek taraflı olmasına göre $t_{\alpha/2,n-1}$ veya $t_{\alpha,n-1}$ değerleri bulunacaktır.

Ortalamaların testinde test istatistiği,

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

olarak hesaplanır. Burada $\overline{X}$ örnekten hesaplanırken μ yerine varsayılan değeri μ_0 kullanılacaktır. Burada dikkat edilmesi gereken standart hatanın hesaplanmasında hata yapılmamasıdır. Standart hatanın seçimin iadeli veya iadesiz olmasına, örnekleme oranına ve anakütle varyansının bilinip bilinmemesine göre nasıl belirlendiği ortalamaların aralık tahmini bölümünde açıklanmıştı. Bu açıklamalara göre standart hata belirlenerek test istatistiği hesaplanmış olacaktır.

Test istatistiği de belirlendikten sonra hipotezlerden hangisinin kabul edileceğine karar verilecektir. Çift taraflı test için H_0 hipotezinin red ve kabul bölgeleri aşağıdaki şekilde gösterilmiştir. H_0 hipotezinin kabulü H_1 hipotezinin reddini, H_0 hipotezinin reddi ise H_1 hipotezinin kabulünü ifade etmektedir.

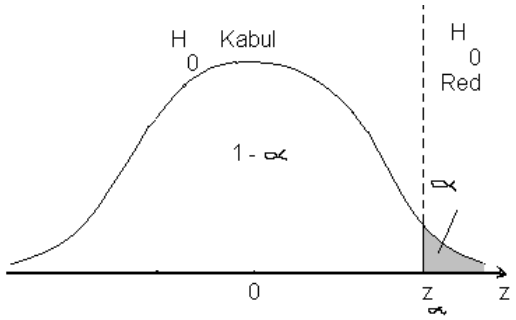


Şekil 10.1

Tek taraflı testte karşıt hipotez,

$H_1 : \mu > \mu_0$

şeklinde oluşturulmuş ise red ve kabul bölgeleri şöyle olacaktır.

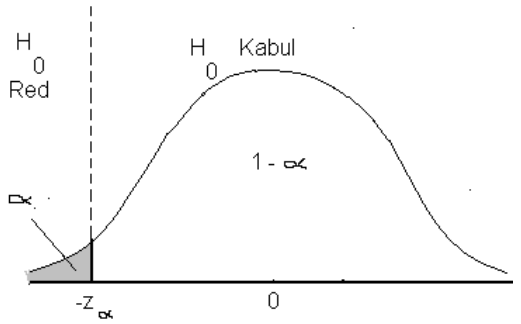


Şekil 10.2

Tek taraflı testte karşıt hipotez,

$H_1 : \mu < \mu_0$

şeklinde oluşturulmuş ise red ve kabul bölgeleri şöyle olacaktır.



Şekil 10.3

Test istatistiğinin şekillerde düştüğü bölgeye göre hipotezlerden biri kabul diğeri reddedilecektir.

ÖRNEK: Bir makinenin ürettiği A mallarının ağırlıklarının dağılımlarının ortalaması 5 kg. ve standart sapması 1,8 kg.'dır. Makine ayarını kontrol etmek amacı ile 81 birimlik bir örnek alınmış ve örnek ortalaması 5,4 kg. bulunmuştur. Üretilen mamüllerin ortalamasının 5 kg.'dan fazla olduğu söylenebilir mi, 0,05 anlam seviyesi ile inceleyiniz.

$\mu = 5$

$\sigma = 1,8$

$n = 81$

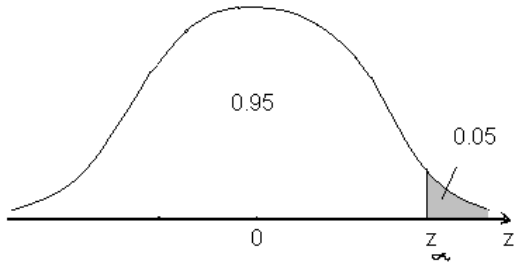
$\bar{X} = 5,4$

$\alpha = 0,05$

$H_0 : \mu = 5$

$H_1 : \mu > 5$

$Z_\alpha \rightarrow \alpha = 0,05$ için



1-0,05 =0,95

0,95 – 0,5 = 0,4500 → Tablodan $Z_\alpha = 1,64$

$$Z = \left| \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \right| = \left| \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{5,4 - 5}{\frac{1,8}{\sqrt{81}}} \right| = \left| \frac{0,4}{0,2} \right| = 2$$

$|Z| > |Z_\alpha|$; 2 >1,64 olduğundan H_0 reddedilerek, H_1 kabul edilir. 0,05 anlam seviyesi ile makinenin ürettiği mamüllerin ortalama ağırlığı 5 kg.'dan fazladır.

11.4. ORANLARIN TESTİ

Anakütle ortalamasında olduğu gibi, anakütle oranı ile ilgili varsayım veya bilgide olabilir. Anakütle oranı P ile ilgili varsayım P_0 ise n birimli örnekten hesaplanacak örnek oranı ise p ise $p=P_0$ olmaması durumunda aradaki farkın istatistiksel olarak anlamlı olup olmadığı test edilecektir. Bu durumda test tek taraflı da çift taraflı da olsa temel hipotez,

$H_0 : P = P_0$

olarak karşıt hipotez çift taraflı test için,

$$H_1 : P \neq P_0$$

tek taraflı test için,

$$H_1 : P < P_0$$

veya

$$H_1 : P > P_0$$

olarak oluşturulacaktır.

Kritik değer ortalamaların testinde açıklandığı gibi belirlenecektir. Küçük örnekler için oranların testi ile uygulamada pek sık karşılaşılmamaktadır. Bu nedenle oranların testinde t dağılımı hemen hemen hiç kullanılmaz.

Oranların testi için test istatistiği,

$$Z = \frac{p - P}{\sigma_p}$$

dir. Burada P yerine P_0 değeri kullanılır. Standart hata σ_p daha önce açıklandığı gibi hesaplanacaktır. Standart hatanın

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

olduğu düşünülürse burada oran olarak biri varsayım, diğeri örnek oranı olmak üzere iki oran bulunmaktadır. Standart hata hesaplanırken oran olarak anakütle ile ilgili varsayım alınacaktır. Hipotezlerden hangisinin kabul edileceği kararı daha önce açıklandığı gibi verilecektir.

ÖRNEK: Bir bölgede yaşayan ailelerin 0,70'inin A malı tüketicisi olduğu kabul edilmektedir. Bu bölgeden 150 aileden oluşan bir örnek alınmış ve bu ailelerin 0,65'inin A malı tüketicisi olduğu anlaşılmıştır. 0,01 anlam seviyesine göre bu bölgede yaşayan ailelerden 0,70'ten daha azının A malı tüketicisi olduğu söylenir mi?

$P = 0{,}70$

$p = 0{,}65$

$n = 150$

$\alpha = 0{,}01$

$H_0 : P = 0{,}70 \qquad (P = P_0)$

$H_1 : P < 0{,}70 \qquad (P < P_0)$

$Z_\alpha = 2{,}33$

$$Z = \left| \frac{p - P}{\sigma_p} \right| = \left| \frac{p - P}{\sqrt{\frac{pq}{n}}} \right| = \left| \frac{0{,}65 - 0{,}70}{\sqrt{\frac{0{,}70 \cdot 0{,}30}{150}}} \right| = \frac{0{,}05}{0{,}038} = 1{,}31$$

$|Z| < |Z_\alpha|$; 1,31<2,33 olduğundan H_0 hipotezi kabul edilir. Bu bölgede yaşayan ailelerden 0,70'ten azının A malı tüketicisi olduğunu söylemek doğru olmayacaktır.

11.5. ORTALAMALAR ARASINDAKİ FARKLARIN TESTİ

Ortalamalar arasındaki farkların aralık tahmini açıklanırken örneklerin bağımsızlığından sözedilmişti. Ortalamalar arasındaki farkların testinde de yapılacak işlemler örneklerin bağımsız veya bağımlı olmasına göre değişmektedir. Bu nedenle konu iki alt başlık altında açıklanacaktır.

11.5.1. Bağımsız Örnekler

İki farklı anakütleden alınan n_1 ve n_2 birimli iki bağımsız örneğin ortalamaları $\overline{X}_1$ ve $\overline{X}_2$ olsun. Örneklerin alındığı anakütlelerin ortalamaları μ_1 ve μ_2 örnek ortalamalarına eşit olarak tahmin edilecektir. Ortalamalar arasında fark olması durumunda, farkın örnekleme hatasından kaynaklanıp kaynaklanmadığının araştırılması istenebilir. Diğer bir ifade ile istatistiksel olarak anakütle ortalamaları birbirine eşit kabul edilebilir mi, ortalamalar arasındaki fark istatistiksel olarak anlamsız mıdır sorusuna yapılacak bir test ile cevap verilebilir.

Ortalamalar arasındaki farkların testinde temel hipotez,

$H_0: \mu_1 = \mu_2$ veya $H_0: \mu_1 - \mu_2 = 0$

şeklinde oluşturulur. Karşıt hipotez ise çift taraflı test için,

$H_1: \mu_1 \neq \mu_2$ veya $H_1: \mu_1 - \mu_2 \neq 0$

tek taraflı testlerde ise,

$H_1: \mu_1 > \mu_2$ veya $H_1: \mu_1 - \mu_2 > 0$

veya

$H_1: \mu_1 < \mu_2$ veya $H_1: \mu_1 - \mu_2 < 0$

olarak oluşturulur.

Kritik değerler daha önce açıklandığı gibi belirlenen hata payına ve testin tek veya çift taraflı olmasına göre normal dağılım tablolarından veya (n_1+n_2-2) serbestlik derecesi ile t dağılımı tablolarından bulunacaktır.

Bağımsız örnekler için ortalama farklarının testinde test istatistiği,

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

dır. Temel hipotez ortalamalar arasında fark olmadığını ifade ettiğinden $\mu_1 - \mu_2 = 0$ alınırsa test istatistiği,

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

olacaktır. Ortalamalar arasındaki farkların standart hatası daha önce açıklandığı gibi varyanslarla ilgili bilgiler ve örnek birim sayıları da dikkate alınarak belirlenecektir. Hipotezlerden hangisinin kabul edileceği kararı da daha önce açıklandığı gibi tablolardan bulunan değerler ile test istatistiği karşılaştırılarak verilecektir.

ÖRNEK: A ve B bölgelerinde yaşayan ailelerin aylık kişi başına beslenme harcamalarını belirlemek amacı ile, A bölgelerinde 140 ve B bölgesinden 90 aileden oluşan iki örnek alınmıştır. A bölgesinden alınan örnekten $\overline{X}_1$=76 milyon TL. ve S_1=7 milyon TL.; B bölgesinden alınan örnekten $\overline{X}_2$ =81 milyon TL. ve S_2=8 milyon TL. bulunmuştur. 0,05 anlam seviyesine göre A ve B bölgelerinde yaşayan ailelerin günlük kişi başına beslenme harcamaları arasında fark olup olmadığını inceleyiniz.

$$\overline{X}_1 = 76 \quad ; \quad S_1 = 7 \quad ; \quad n_1 = 140$$
$$\overline{X}_2 = 81 \quad ; \quad S_2 = 8 \quad ; \quad n_2 = 90$$

1- $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

2- $Z_{\alpha/2}$ =1,96

3-

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{76 - 81}{\sqrt{\frac{7^2}{140} + \frac{8^2}{90}}}$$

$$= \frac{-5}{1,030} = -4,85$$

4- $|Z| > |Z_{\alpha/2}|$; 4,85 > 1,96 olduğundan H_1 hipotezi kabul edilir. Harcamalar arasında fark vardır.

ÖRNEK: Önceki örnekte n_1=14 ve n_2=9 kabul ederek örneği tekrar çözünüz.

1- $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

2- n_1 ve n_2 küçük örnekler olduğundan n_1+n_2-2 =14+9-2 =21 serbestlik dereceli t dağılımı $t_{0,05/2,21}$ =2,080

3- $Z = \dfrac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}}$

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(14 - 1)(49) + (9 - 1)(64)}{14 + 9 - 2} = 54{,}7143$$

$$S = 7{,}3969$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 7{,}3969\sqrt{\frac{1}{14} + \frac{1}{9}} = 3{,}1603$$

$$Z = \frac{76 - 81}{3{,}1603} = -1{,}58$$

4- $|1{,}58| < 2{,}080$ olduğundan temel hipotez kabul edilir.

11.5.2. Bağımlı Örnekler

Bazı durumlarda aynı birimlerde yapılan farklı ölçümlerin ortalamalarının arasındaki farkın istatistiksel olarak anlamlılığı da test edilmek istenebilir. Örneğin bir grup öğrenciye uygulanan eğitim öncesi ve sonrası sınavların ortalamaları, hastalara uygulanan tedavi öncesi ve sonrası ölçümlerin ortalaması, makinenin üretiminin ayarlama öncesi ve sonrası ölçümlerin ortalaması gibi. Bu gibi durumlarda bağımlı örnekler sözkonusu olmaktadır.

Aynı birimde yapılan iki farklı ölçümün ortalamaları arasındaki fark, ölçümlerin birbirinden farklarının ortalamasına eşit olacaktır. Ölçümleri X_1 ve X_2 değişkenleri ile gösterelim. Ölçüm farklarını D_i ile ifade edersek,

$$D_i = X_{1i} - X_{2i}$$

olacaktır. Değişkenlerin ayrı ayrı ortalamaları $\overline{X}_1$ ve $\overline{X}_2$ arasındaki fark $\overline{D}$ 'ya eşittir.

$$\overline{X}_1 - \overline{X}_2 = \overline{D}$$

Bu durumda standart değişken

$$Z = \frac{\overline{D} - D}{\sigma_{\overline{D}}}$$

ve standart hata ise,

$$\sigma_D = \sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n^2(n-1)}}$$

olacaktır. Burada n ölçüm yapılan birim sayısını göstermektedir. İki farklı ölçüm için birim sayısının iki katı alınmamaktadır.

Yukarıda söz edildiği gibi bağımlı örnekler söz konusu olduğunda da hipotezler bağımsız örneklerin ortalamalarının farklarında olduğu gibi çift veya tek taraflı olarak oluşturulur. Kısaca hipotezler değişmemektedir.

Kritik değerin belirlenmesi örnekleme dağılımı ile ilgilidir. Burada da örnekleme dağılımı normal dağılım ve t dağılımı olmaktadır. n <30 olması durumunda (n-1) serbestlik dereceli t dağılımı sözkonusudur. α hata payı ile ilgili tablolardan kritik değerler daha önce açıklandığı gibi belirlenir.

$$Z = \frac{\overline{D}}{S_{\overline{D}}}$$

olacaktır. D'lerin ortalaması,

$$\overline{D} = \frac{\Sigma D_i}{n}$$

ve $\sigma_{\overline{D}}$ değerleri ile test istatistiği,

$$Z = \frac{\frac{\Sigma D}{n}}{\sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n^2(n-1)}}} = \frac{\Sigma D_i}{\sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n-1}}}$$

olacaktır. Karar ise daha önce olduğu gibi verilecektir.

ÖRNEK: 9 öğrenci tesadüfi olarak seçilerek 2 hafta süre ile eğitim verilmiştir. Eğitim öncesi ve sonrası yapılan sınav notları aşağıda verilmiştir.

Öğrenci No	1. Sınav	2. Sınav
1	42	75
2	61	60
3	96	90
4	38	57
5	71	60
6	17	35
7	47	63

8	50	94
9	41	88

0,05 hata payı ile öğrencilere verilen eğitim etkili olmuş mudur, inceleyiniz.

Öğrenci No	1. Sınav	2. Sınav	$D_i=X_{1i}-X_{2i}$	D_i^2
1	42	75	-33	1089
2	61	60	1	1
3	96	90	6	36
4	38	57	-19	361
5	71	60	11	121
6	17	35	-18	324
7	47	63	-16	256
8	50	94	-44	1936
9	41	88	-47	2209
			-159	6333

1- $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$

2- n <30

$\alpha = 0,05$

SD = n-1 = 9-1= 8 t dağılımı

$t_{0,05;\,8} = 1,860$

3-

$$Z = \frac{\Sigma D_i}{\sqrt{\frac{n\Sigma D_i^2 - (\Sigma D_i)^2}{n-1}}}$$

$$= \frac{-159}{\sqrt{\frac{9(6333) - (-159)^2}{9-1}}} = -2,525$$

4- $|Z| > |t_{0,05,8}|$; $|2,525| > |1,860|$ olduğundan H_0 hipotezi reddedilerek, H_1 hipotezi kabul edilir. Buna göre eğitim etkili olmuştur.

11.6. ORANLAR ARASINDAKİ FARKLARIN TESTİ

İki farklı anakütleden alınan bağımsız örneklerin ortalamaları arasındaki farkın anlamlılığı gibi, oranlar arasındaki farkın anlamlılığı da test edilebilir. Anakütle oranı P_1 ve P_2, alınacak n_1 ve n_2 birimli örneklerin oranları p_1 ve p_2 ise anakütle oranları örnek oranlarına eşit olarak tahmin edilecektir. Yapılacak test ile p_1 ve p_2 arasındaki farkın istatistiksel anlamlılığı incelenecektir.

Oranlar arasındaki farkların testinde temel hipotez,

$$H_0 : P_1 = P_2 \quad veya \quad P_1 - P_2 = 0$$

karşıt hipotez ise çift taraflı test için,

$$H_1 : P_1 \neq P_2 \quad veya \quad P_1 - P_2 \neq 0$$

tek taraflı test için ise,

$$H_1 : P_1 > P_2 \quad veya \quad P_1 - P_2 > 0$$

veya

$$H_1 : P_1 < P_2 \quad veya \quad P_1 - P_2 < 0$$

şeklinde oluşturulur.

Oran farklarının testinde kritik değer daha önce açıklandığı gibi belirlenen hata payı veya güven olasılığı ile bulunur. Burada test istatistiği,

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1 - p_2}}$$

dir. Standart hata ise oran farklarının aralık tahmininde açıklandığı gibi belirlenir. Hipotezlerden birinin kabulü kararı da aynı şekilde verilir.

ÖRNEK: İki ambarda bulunan malların bazılarının hatalı olduğu bilinmektedir. Hata oranlarının karşılaştırılması için iki örnek alınmıştır. Birinci örnek 148 birimden oluşmuştur ve bunlardan 37 adedi hatalıdır. İkinci örnek 125 birimden oluşmuştur ve 25 adedi hatalıdır. 0,05 hata payı ile örneklerin alındığı anakütle oranları arasında fark var mıdır?

1- $H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$

2- $Z_{\alpha/2} = 1{,}96$

3- $$Z = \frac{|P_1 - P_2|}{\sigma_{p_1 - p_2}} = \frac{|P_1 - P_2|}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

$$P_1 = \frac{k_1}{n_1} = \frac{37}{148} = 0{,}25$$

$$P_2 = \frac{k_2}{n_2} = \frac{25}{125} = 0{,}20$$

$$Z = \frac{|0{,}25 - 0{,}20|}{\sqrt{\frac{(0{,}25).(0{,}75)}{148} + \frac{(0{,}20).(0{,}80)}{125}}} = \frac{|0{,}05|}{0{,}05} = 1$$

4- 1<1,96 olduğundan H_0 hipotezi kabul edilir. Örneklerin alındığı anakütle oranları arasında 0,05 anlam seviyesi ile fark yoktur.

11.7. ANAKÜTLE VARYANSININ TESTİ

Önceki bölümlerde ortalamalar ve oranlar ile ilgili testler açıklanmıştı. Bu parametreler gibi varyansların testi de gerekebilir. Örneğin, belirli özellikte malları üreten üreticiler ürettikleri malların özelliklerinin değişkenliğini test etmek isteyebilirler. Bu gibi durumlarda ortalamaların ve oranların testinde olduğu gibi anakütle varyansı ile ilgili bir bilgi veya varsayım gerekecektir.

Normal dağılmış bir anakütle varyansı (σ^2) ile ilgili varsayılan değer σ_0^2 ise, bu varsayımın doğru olup olmadığını belirlemek için n birimli bir örnek alınarak, hesaplanacak S^2 ile test yapılacaktır. Anakütle varyansının testinde de çift veya tek taraflı test yapılabilir. Her iki durum için de temel hipotez anakütle varyansının varsayılan değere eşit oluğunu ifade edecek şekilde,

$$H_0 : \sigma^2 = \sigma_0^2$$

olarak oluşturulacaktır. Karşıt hipotez ise çift taraflı test için,

$$H_1 : \sigma^2 \neq \sigma_0^2$$

tek taraflı test için ise,

$$H_1 : \sigma^2 > \sigma_0^2$$

veya

$$H_1 : \sigma^2 < \sigma_0^2$$

şeklinde oluşturulacaktır.

Anakütle varyansının testinde dağılım χ^2 dağılımıdır. Bu nedenle belirlenen α hata payı ile çift taraflı test için, $\chi^2_{1-\alpha/2,n-1}$ ve $\chi^2_{\alpha/2,n-1}$ değerleri; tek taraflı test için büyüklük sözkonusu ise $\chi^2_{\alpha,n-1}$, küçüklük sözkonusu ise $\chi^2_{1-\alpha,n-1}$ değerleri ki-kare tablosundan bulunacaktır.

Anakütle varyansının testinde kullanılacak test istatistiğini χ^2 ile ifade edersek,

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

dir. Hesaplanan χ^2 test istatistiği belirlenen tablo değerleri ile karşılaştırılarak hipotezlerden biri kabul edilecektir. Çift taraflı testlerde, $\chi^2_{1-\alpha,n-1} \leq \chi^2 \leq \chi^2_{\alpha/2,n-1}$ ise H_0 hipotezi, aksi sözkonusu ise H_1 hipotezi kabul edilecektir. Tek taraflı testlerde $H_1 : \sigma^2 > \sigma_0^2$ hipotezi için

$\chi^2 \geq \chi^2_{\alpha,n-1}$ ise, $H_1 : \sigma^2 < \sigma_0^2$ hipotezi için $\chi^2 \leq \chi^2_{1-\alpha,n-1}$ ise H_0 hipotezi reddedilerek H_1 hipotezi kabul edilecektir.

ÖRNEK: Normal dağılan bir anakütle varyansının 225 olduğu ileri sürülmektedir. Bu anakütleden 25 birimli bir örnek alınarak S^2=312 bulunmuştur. 0,05 hata payı ile anakütle varyansının 225 birimden büyük olduğu söylenebilir mi, inceleyiniz.

1- $H_0 : \sigma^2 = 225$
$H_1 : \sigma^2 > 225$

2- α =0,05

SD= n-1=25-1=24

$\chi^2_{0,05;24} = 36,415$

3-

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

$$= \frac{(25-1).312}{225} = 33,28$$

4- $\chi^2 < \chi^2_{\alpha,n-1}$; 33,28<36,415 olduğundan H_0 hipotezi kabul edilir. Anakütle varyansı 225 birim olarak kabul edilir.

11.8. VARYANS ORANLARININ TESTİ

Ortalamalar ve oranlar arasındaki farkların testi gibi varyansların farklarının testi de gerekebilir. Örneğin, daha önce açıklandığı gibi ortalamalar arasındaki farklar test edilirken varyansların eşit olduğu ileri sürülebilir. Bu gibi durumlarda ortalamalar arasındaki farktan önce varyansların eşitliği test edilebilir. Burada varyansların farklarının testi yerine varyans oranlarının testinden sözedilmesi test istatistiğinin yapısından kaynaklanmaktadır.

Varyans oranlarının veya varyans farklarının testinde de çift ve tek taraflı test yapılabilir. Her iki test içinde temel hipotez,

$H_0 : \sigma_1^2 = \sigma_2^2$ *veya* $\sigma_1^2 - \sigma_2^2 = 0$

şeklinde oluşturulur. Karşıt hipotez ise çift taraflı test için,

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad veya \quad \sigma_1^2 - \sigma_2^2 \neq 0$$

tek taraflı test için ise,

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad veya \quad \sigma_1^2 - \sigma_2^2 > 0$$

veya

$$H_1 : \sigma_1^2 < \sigma_2^2 \quad veya \quad \sigma_1^2 - \sigma_2^2 < 0$$

şeklinde oluşturulur.

Varyans oranlarının testinde dağılım F dağılımıdır. Bu nedenle belirlenen α hata payı ile (n_1-1) ve (n_2-1) serbestlik dereceleri ile çift taraflı test için $F_{\alpha/2,n_1-1,n_2-1}$ ve $F_{\alpha/2,n_2-1,n_1-1}$; tek taraflı testler için büyüklük sözkonusu ise F_{α,n_1-1,n_2-1}; küçüklük sözkonusu ise F_{α,n_2-1,n_1-1} değerleri F tablosundan bulunur.

Daha önce varyans oranlarının tahmininde açıklandığı gibi,

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$

dir. Testte temel hipotez $\sigma_1^2 = \sigma_2^2$ olduğundan test istatistiği

$$F = \frac{S_1^2}{S_2^2}$$

olacaktır.

Test istatistiği (F), S_1 ve S_2'nin büyüklüğüne göre belirlenecektir. Büyük olan S değeri payda yeralacaktır. Çift taraflı test için,

$$S_1^2 \geq S_2^2 \quad ise \quad F = \frac{S_1^2}{S_2^2} \geq F_{\alpha/2,n_1-1,n_2-1}$$

$$S_2^2 \geq S_1^2 \quad ise \quad F = \frac{S_2^2}{S_1^2} \geq F_{\alpha/2,n_2-1,n_1-1}$$

olması durumunda H_1 hipotezi kabul edilecektir. Tek taraflı testlerde

$H_1 : \sigma_1^2 > \sigma_2^2$ için,

$$F = \frac{S_1^2}{S_2^2} \geq F_{\alpha,n_1-1,n_2-1}$$

$H_1 : \sigma_1^2 < \sigma_2^2$ için,

$$F = \frac{S_2^2}{S_1^2} \geq F_{\alpha, n_2-1, n_1-1}$$

olması durumunda H_1 hipotezi kabul edilecektir.

ÖRNEK: Normal dağılmış anakütleden $n_1=6$ ve $n_2=6$ birimden oluşan iki örnek alınarak $S_1^2 = 186$ ve $S_2^2 = 141$ olarak belirlenmiştir. 0,05 hata payı ile S_1^2 'nin ve S_2^2 'den büyük olduğu söylenebilir mi, inceleyiniz.

1- $H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$

2- $\alpha = 0{,}05$

SD $= n_1 - 1 = 6 - 1 = 5$

SD $= n_2 - 1 = 6 - 1 = 5$

$$F_{\alpha, n_2-1, n_1-1} = F_{0{,}05;5,5} = 5{,}05$$

3- $$F = \frac{S_1^2}{S_2^2} = \frac{186}{141} = 1{,}319$$

4- $1{,}319 < 5{,}05$ olduğundan H_0 hipotezi kabul edilir. Varyanslar birbirine eşittir.

11.9. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: A marka ampullerin ortalama kullanılma sürelerinin 1100 saat olduğu ileri sürülmektedir. Ampullerin ortalama kullanılma sürelerini belirlemek için 64 ampulden oluşan bir örnek alınarak, örnek ortalaması $\overline{X} = 1093$ saat ve $S^2 = 900$ saat bulunmuştur. 0,01 hata payı ile ampüllerin ortalama kullanılma sürelerinin 1100 saatten az olduğu söylenebilir mi?

Çözüm:

$\mu = 1100$

$S^2 = 900 \rightarrow S = 30$

$n = 64$

$\overline{X} = 1093$

$\alpha = 0{,}01$

1- $H_0 : \mu = 1100$
$H_1 : \mu < 1100$

2- $1-\alpha = 0{,}99$

$0,99 - 0,5 = 0,49 \rightarrow$ Tablodan $Z_\alpha = 2,33$

3- $Z = \left| \dfrac{\overline{X} - \mu}{\sigma_{\overline{X}}} \right| = \left| \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}} \right| = \left| \dfrac{1093 - 1100}{\dfrac{30}{\sqrt{64}}} \right| = \left| \dfrac{-7}{3,75} \right| = 1,86$

4- $|Z| < |Z_\alpha|$; 1,86<2,33 olduğundan H_0 kabul, H_1 reddedilir. Ampullerin ortalama kullanılma sürelerinin 0,01 anlam seviyesine göre 1100 saatten az olduğu söylenemez.

ÖRNEK 2: Bir fabrikada üretilen A parçalarının çapları 33 mm.'dir. Üretim kontrolü için fabrika üretiminden 200 birimlik bir örnek alınarak $\overline{X} = 34$ mm. ve S =4 mm. bulunmuştur. Buna göre parçaların çaplarının 33 mm. olduğu kabul edilebilir mi? 0,01 hata payı ile inceleyiniz.

Çözüm:

$\mu = 33$

$S = 4$

$n = 200$

$\overline{X} = 34$

$\alpha = 0,01$

1-

$H_0 : \mu = \overline{X}$

$H_1 : \mu \neq \overline{X}$

2- $1-\alpha = 0,99$

$0,99 - 0,5 = 0,495 \rightarrow$ Tablodan $Z_\alpha = 2,57$

3- $Z = \left| \dfrac{\overline{X} - \mu}{\sigma_{\overline{X}}} \right| = \left| \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}} \right| = \left| \dfrac{34 - 33}{\dfrac{4}{\sqrt{200}}} \right| = \left| \dfrac{1}{0,28} \right| = 3,57$

4- $|Z| > |Z_{\alpha/2}|$; 3,57>2,57 olduğundan H_0 reddedilir, H_1 kabul edilir. Parçaların çaplarının 33 mm. olduğu kabul edilmez.

ÖRNEK 3: Normal olduğu bilinen ve μ=220 olan bir anakütleden n=25 birimlik bir örnek alınarak $\overline{X} = 216$ ve S=12 bulunmuştur. 0,95 güven olasılığı ile anakütle ortalaması ile örnek ortalaması arasında fark olup olmadığını inceleyiniz.

Çözüm:

1-

$H_0 : \mu = 220$

$H_1 : \mu \neq 220$

2- n <30, anakütle dağılımı normal ve σ bilinmediğinden t- dağılımı tablosundan,

$\alpha/2 = 0{,}025$ ve 25-1 =24 serbestlik derecesi ile $t_{\alpha/2} = 2{,}064$

3- Z yerine t kullanalım.

$$t = \left| \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \right| = \left| \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \right| = \left| \frac{216 - 220}{\frac{12}{\sqrt{25}}} \right| = 1{,}66$$

4- $|t| < |t_{\alpha/2}|$; 1,66<2,064 olduğundan H_0 hipotezi kabul edilir. Örnek ortalaması ile anakütle ortalaması arasında fark yoktur.

ÖRNEK 4:

$\overline{X}_1 = 18 ; S_1 = 3 ; n = 49 \quad ve \quad \overline{X}_2 = 21 ; S_2 = 4 ; n = 36$

olan iki örneğin μ= 20 olan bir anakütleden alındığı ileri sürülmektedir. Bu örneklerin 0,10 hata payı ile bu anakütleden alınıp alınmadığını belirleyiniz.

Çözüm:

I. Örnek İçin;

1-

$H_0 : \mu = 20$

$H_1 : \mu \neq 20$

2- 1-α =0,90

0,90/2 = 0,45 → Tablodan $Z_{\alpha/2} = 1{,}64$

3- $$Z = \left| \frac{\overline{X}_1 - \mu}{\sigma_{\overline{X}_1}} \right| = \left| \frac{\overline{X}_1 - \mu}{\frac{S_1}{\sqrt{n_1}}} \right| = \left| \frac{18 - 20}{\frac{3}{\sqrt{49}}} \right| = 4{,}66$$

4- $|Z| > |Z_{\alpha/2}|$; 4,66>1,64 olduğundan H_1 kabul edilir. Anakütle ve örnek ortalaması arasında 0,10 anlam seviyesi ile fark vardır. Bu örnek bu anakütleyi temsil etmemektedir ve bu anakütleden alınmadığı söylenebilir. ($\mu \neq \overline{X}$)

II. Örnek İçin;

1- Hipotezler :

$H_0 : \mu = 20$

$H_1 : \mu \neq 20$

2- $Z_{\alpha/2} = 1{,}64$

3- $$Z = \left| \frac{\overline{X}_2 - \mu}{\sigma_{\overline{X}_2}} \right| = \left| \frac{\overline{X}_2 - \mu}{\frac{S_2}{\sqrt{n_2}}} \right| = \left| \frac{21 - 20}{\frac{4}{\sqrt{36}}} \right| = 1{,}5$$

4- $|Z| < |Z_{\alpha/2}|$; 1,5<1,64 olduğundan H_0 kabul edilir. Anakütle ve örnek ortalaması arasında 0,10 anlam seviyesi ile fark yoktur. Bu örnek bu anakütleyi temsil etmektedir ve bu anakütleden alındığı söylenebilir. ($\mu = \overline{X}$)

ÖRNEK 5: 2000 birimlik bir anakütlenin ortalamasının μ=140 olduğu ileri sürülmektedir. Bu anakütleden iadesiz seçimle alınan 121 birimlik örnekte $\overline{X} = 144$ ve S=9 bulunmuştur. 0,05 hata payı ile anakütle ortalaması 140 olarak kabul edilebilir mi?

1- $H_0 : \mu = 140$

$H_1 : \mu \neq 140$

2- $Z_{\alpha/2} = 1{,}96$

3 $$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$\sigma_{\overline{X}}$ için; $\frac{n}{N} = \frac{121}{2000} = 0{,}0605 > 0{,}05$ olduğundan

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{9}{\sqrt{121}} \cdot \sqrt{\frac{2000-121}{2000-1}} = 0{,}79$$

$$Z = \left| \frac{144 - 140}{0{,}79} \right| = 5{,}06$$

4- $|Z| > |Z_{\alpha/2}|$; 5,06>1,96 olduğundan H_0 hipotezi reddedilir. Anakütle ortalaması 140 olarak kabul edilemez.

ÖRNEK 6: Bir bölgede yapılan bir araştırma ile marjinal tüketim eğiliminin 0,68 olduğu belirlenmiştir. Aynı bölgeden n= 500 olan bir örnek alınarak marjinal tüketim eğilimi 0,76 olarak belirlenmiştir. 0,99 güven olasılığı ile marjinal tüketim eğilimi 0,68'den fazla mıdır, inceleyiniz.

Çözüm:

$P = 0{,}68$

$p = 0{,}76$

$n = 500$

$\alpha = 0{,}01$

1- $H_0 : P = 0{,}68$
$H_1 : P > 0{,}68$

2- $Z_\alpha = 2{,}33$

3- $Z = \dfrac{p - P}{\sigma_p}$

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0{,}68)(0{,}32)}{500}} = 0{,}0209$$

$$Z = \frac{0{,}76 - 0{,}68}{0{,}0209} = 3{,}82$$

4- $|Z| > 3{,}82$ olduğundan H_0 hipotezi reddedilir. Marjinal tüketim eğilimi 0,68'den fazladır.

ÖRNEK 7: Bir fabrikada üretilen mamüllerin 0,15'inin hatalı olduğu ileri sürülmektedir. Hata oranını belirlemek amacı ile 250 mamülden oluşan bir örnek alınmış ve 250 mamülden 46 adedinin hatalı olduğu anlaşılmıştır. 0,05 hata payı ile bu fabrika üretiminin 0,15'inin hatalı olduğunu söylemek doğru mudur?

Çözüm:

1- $H_0 : P = 0{,}15$
$H_1 : P \neq 0{,}15$

2- $Z_{\alpha/2} = 1{,}64$

3- $$Z = \left|\frac{p - P}{\sigma_p}\right| = \left|\frac{p - P}{\sqrt{\frac{pq}{n}}}\right|$$

$$p = \frac{46}{250} = 0{,}184$$

$$Z = \left|\frac{0{,}184 - 0{,}15}{\sqrt{\frac{(0{,}15).(0{,}85)}{250}}}\right| = \left|\frac{0{,}034}{0{,}022}\right| = 1{,}54$$

4- $|Z| < |Z_{\alpha/2}|$; 1,54 <1,64 olduğundan H_0 hipotezi kabul edilir. Üretimin 0,05'inin hatalı olduğu söylenebilir.

ÖRNEK 8: İki aynı fabrikada aynı mamül üretilmektedir. Mamüllerin boy uzunluklarını karşılaştırmak için iki fabrika üretiminden birer örnek alınmıştır. Birinci fabrika için $\sigma_1 = 12$ mm. ve ikinci fabrika için $\sigma_2 = 14$ mm. olarak bilinmektedir. Birinci örnek 70 mamülden oluşmuş ve ortalaması 198 mm. ; ikinci örnek 80 mamülden oluşmuş ve ortalaması 201 mm. olarak bulunmuştur. 0,10 hata payı ile fabrikalarda üretilen mamüllerin boy uzunlukları arasında fark olup olmadığını belirleyiniz.

Çözüm:

$\sigma_1 = 12 \quad ; \quad \overline{X}_1 = 198 \quad ; \quad n_1 = 70$

$\sigma_2 = 14 \quad ; \quad \overline{X}_2 = 201 \quad ; \quad n_2 = 80$

1- $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

2- $Z_{\alpha/2} = 1{,}64$

3- $$Z = \left| \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}} \right| = \left| \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

$$= \frac{|198 - 201|}{\sqrt{\frac{12^2}{70} + \frac{14^2}{80}}} = \frac{|3|}{2{,}12} = 1{,}41$$

4- $|Z| < |Z_{\alpha/2}|$; 1,41<1,64 olduğundan H_0 hipotezi kabul edilir. Fabrikalarda üretilen mamüllerin boy uzunlukları arasında fark yoktur.

ÖRNEK 9: A malları 1 kg. lık ambalajlarda piyasaya sürülmektedir. A malının kilogramındaki a maddesi miktarını belirlemek amacı ile bu mallardan biri n_1=10 diğeri n_2=12 birimden oluşan iki örnek alınarak $\overline{X}_1 = 21$ gr. ve S_1 =3 gr.; $\overline{X}_2 = 29$ gr. ve S_2 =4 gr. bulunmuştur. 0,01 hata payı ile ortalamalar arasında fark var mıdır? İnceleyiniz. (a maddesi miktarları dağılımı normal olarak bilinmektedir.)

Çözüm:

1- $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

2- $n_1 + n_2 - 2 = 10+12-2 = 20$; σ_1 ve σ_2 bilinmiyor ve dağılımlar normal olduğundan, $\alpha/2 = 0{,}01/2 = 0{,}005$ ve 20 serbestlik derecesi ile t- dağılımı tablosundan $t_{\alpha/2} = 2{,}845$

3- $$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(10-1).3^2 + (12-1).4^2}{10+12-2} = 12{,}85$$

$$S = \sqrt{12{,}85} = 3{,}58$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 3{,}58\sqrt{\frac{1}{10}+\frac{1}{12}} = 1{,}53$$

$$t = \frac{21-29}{1{,}53} = 5{,}22$$

4- $\left|t\right| > \left|t_{\alpha/2}\right|$; 5,22 >2,845 olduğundan H_1 hipotezi kabul edilir. Örnek ortalamaları arasında fark vardır.

ÖRNEK 10: İki iş kolunda çalışan işçilerin saat ücretlerini karşılaştırmak için iş kollarından birer örnek alınmıştır. Birinci iş kolu için $n_1 = 20$, $\overline{X}_1 = 188$ TL., $S_1^2 = 55$ TL.; ikinci iş kolunda $n_2 = 22$, $\overline{X}_2 = 5$ TL., $S_2^2 = 6$ TL. bulunmuştur. 0,05 hata payı ile birinci iş kolunda çalışan işçilerin ikinci iş kolunda çalışanlardan daha fazla ücret aldığı söylenebilir mi? İnceleyiniz.

Çözüm:

1- $$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

2- $Z_\alpha = 1{,}64$

3- $$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{8-5}{\sqrt{\frac{5}{20}+\frac{6}{22}}} = \frac{3}{0{,}72} = 4{,}16$$

4- $\left|Z\right| > \left|Z_\alpha\right|$; 4,16 > 1,64 olduğundan H_1 hipotezi kabul edilir. Birinci iş kolunda işçilerin ikinci iş kolundan daha az ücret aldıkları söylenebilir.

ÖRNEK 11: A ve B bölgelerinde ilkokul çağında olan çocukların okula gidip gitmedikleri araştırılmaktadır. Bu nedenle A bölgesinden 200, B bölgesinden 250 çocuktan oluşan birer örnek alınmış ve A bölgesinde ilkokul çağındaki çocukların

0,71'inin; B bölgesinde 0,62'sinin okula devam ettiği anlaşılmıştır. 0,10 hata payı ile A bölgesinde ilkokul çağındaki çocukların okula devam etme oranının B bölgesinden fazla olduğu söylenebilir mi?

Çözüm:

1- $H_0 : P_A = P_B$
$H_1 : P_A > P_B$

2- $Z_\alpha = 1{,}28$

3-

$$Z = \frac{|P_A - P_B|}{\sigma_{p_A - p_B}} = \frac{|P_A - P_B|}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}}$$

$$= \frac{|0{,}71 - 0{,}62|}{\sqrt{\frac{(0{,}71).(0{,}29)}{200} + \frac{(0{,}62).(0{,}38)}{250}}} = 2{,}04$$

4- $|Z| > |Z_\alpha|$; 2,04 > 1,28 olduğundan H_1 hipotezi kabul edilir. A bölgesinde okul çağında olan çocukların okula devam etme oranları B bölgesinden büyüktür.

BÖLÜM 12

VARYANS ANALİZİ

Hipotez testleri bölümünde ortalamaların, oranların ve varyansların testlerini açıkladık. Açıklanan testlerde bir veya iki anakütle ile ilgili karşılaştırmalar yapılabilmekteydi. Bu bölümde ikiden fazla anakütle ortalamasının karşılaştırılması için kullanılan varyans analizi (ANOVA=Analysis of Variance) ele alınacaktır.

12.1. İKİDEN FAZLA ORTALAMA ARASINDAKİ FARKLARIN TESTİ

Çeşitli araştırmalarda ikiden fazla anakütle ortalaması arasında fark olup olmadığının diğer bir ifade ile aralarındaki farkların istatistiksel olarak anlamlı olup olmadığının araştırılması gerekebilir. Hipotez testleri bölümünde açıklandığı gibi anakütleler ikişer ikişer eşlendirilerek aralarındaki farkların anlamlılığı test edilebilir.

Ortalamalar ikişer ikişer test edilirken test edilecek ortalama sayısı k ise, C_k^2 veya k(k-1)/2 tane test yapılması gerekecektir. Örneğin, üç ortalama için $C_3^2 = 3$, dört ortalama için $C_4^2 = 6$, beş ortalama için $C_5^2 = 10$ adet ikişerli ortalama testi yapılması gerekmektedir. Bu durumda tüm test işlemleri için 1.tip hata yapma olasılığı artacaktır. Örneğin, üç anakütle ortalaması ikişerli olarak karşılaştırılacak olsun. Hata payı yani 1.tip hata yapma olasılığı 0,05, güven olasılığı 0,95 ise üç testte de doğru sonuç elde etme olasılığı binom dağılımı ile,

$$P(X = 3) = C_3^3 (0{,}95)^3 (0{,}05)^0 = (0{,}95)^3 = 0{,}86$$

olacaktır. Bu durumda en az bir 1.tip hata yapma olasılığı (1-0,86=0,14)'tür. Aynı olasılık 0,05 hata payı ile 5 ortalamanın testi için,

$$P(X = 10) = C_{10}^{10} (0{,}95)^{10} (0{,}05)^0 = (0{,}95)^{10} = 0{,}60$$

ve 1.tip hata yapma olasılığı (1-0,60 =0,40) olacaktır. Görüldüğü gibi ortalama sayısı arttıkça en az bir 1.tip hata yapma olasılığı artmaktadır. Bu açıklamalardan da anlaşılacağı gibi ikiden fazla ortalamanın ikişer ikişer test edilmesi durumunda hata sözkonusudur, ikişerli testler birden fazla ortalamanın eşitliğinin testi için yetersizdir.

Varyans analizi ikiden fazla ortalamanın testi için kullanılan yeterli bir yöntemdir. Yöntemin temeli farklı kaynaklardan hesaplanan varyansların karşılaştırılmasına dayanmaktadır. k ortalama arasındaki fark test edilirken örneklerin varyansı σ^2 olan normal dağılmış anakütlelerden alındığı varsayılmaktadır. Diğer bir ifade ile, örneklerin aynı anakütleden alındığının varsayıldığını söyleyebiliriz. Temel hipotez ortalamaların eşitliğini ifade edecek şekilde oluşturulduğundan, örneklerin alındığı anakütlelerin normal dağıldığı ve varyanslarının eşit (homojen) olduğu kabul edilmektedir.

Varyans analizinin temeli daha önce açıklanan F dağılımına dayanmaktadır. k ortalama arasındaki farkların istatistiksel anlamlılığı varyans analizi ile şöyle yapılmaktadır.

1- Temel hipotez ortalamaların eşitliğini, karşıt hipotez ise temel hipotezin doğru olmadığını belirtecek şekilde hipotezler oluşturulur. Karşıt hipotez tüm ortalamaların eşit olmadığını gösterecektir.

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

$$H_1 : \mu_1 \neq \mu_2 \neq ..... \neq \mu_k$$

Burada H_1 hipotezi ortalamalardan tümünün değil, en az birinin farklı olduğunu ifade etmektedir.

Örneklerin ortalamaları hesaplanarak, k örneğin ortalaması için ortalamaların standart hatası ($S_{\overline{X}}$) hesaplanır.

$$S_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{X}_i - \overline{\overline{X}})^2}{k-1}}$$

Burada $\overline{\overline{X}}$ k örneğin ortalamalarının ($\overline{X}_i$) ortalamasıdır.

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}$$

2- Hesaplanan ortalamaların standart hatası ($S_{\overline{X}}$) ile k örneğin alındığı anakütlenin varyansı tahmin edilir. Tahmin edilen anakütle varyansı S^2 ile ifade edilir ve gruplar arası kareler ortalaması ($KO_{G.Arası}$) olarak adlandırılır. Ortalamaların standart hatası,

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

olduğundan,

$$S = S_{\bar{X}}\sqrt{n}$$

$$S^2 = S_{\bar{X}}^2 .n = \text{ KO}_{\text{G.Arası}}$$

olacaktır. Burada n örnek birim sayısıdır ve tüm örneklerin birim sayılarının eşit olduğu varsayılmaktadır.

3- Grupların ortalamalarına göre ve bağımsız olarak anakütle varyansı tahmin edilir. Tahmin edilen varyansların tartılı ortalamaları alınarak bunlar birleştirilir. Tartı olarak örnek birim sayılarının bir eksiği alınır. Örnek birim sayılarını n_i ile ifade edersek, tartı $(n_i - 1)$ olacaktır. Anakütle varyansının tahmin edilen bu değeri gruplar içi kareler ortalaması ($KO_{G.içi}$) olarak adlandırılır.

$$\text{KO}_{\text{G.İÇİ}} = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{\sum_{i=1}^{k}(n_i - 1)}$$

4- Elde edilen bu iki S^2 arasındaki fark F dağılımı ile tek taraflı olarak test edilir. Test istatistiği,

$$F = \frac{KO_{G.Arası}}{KO_{G.İçi}}$$

dir. F dağılımın serbestlik dereceleri,

$SD_1 = k-1$

$SD_2 = nk-k$

dır. Verilen serbestlik dereceleri ve belirlenen α hata payı ile F dağılımı tablosundan bulunan $F_{\alpha,SD1,SD2}$ değeri ile F test istatistiği karşılaştırılır.

$F < F_{\alpha,SD1,SD2}$ ise H_0 kabul

$F > F_{\alpha,SD1,SD2}$ ise H_1 kabul

edilecektir.

ÖRNEK: Aşağıda öğrencilerin 50 üzerinden üç dersten aldıkları notlar verilmiştir.

Dersler	**Notlar**				
Matematik	29	16	36	15	24
İstatistik	41	28	33	11	37
Bilgisayar	49	29	38	25	19

0,05 hata payı ile ortalamalar arasındaki farkların anlamlılığını (örnek ortalamalarının aynı anakütleden alınıp alınmadığını) inceleyiniz.

Dersler	Notlar					Notların Toplamı	Xi
Matematik	29	16	36	15	14	110	22
İstatistik	41	28	33	11	37	150	30
Bilgisayar	49	29	38	25	19	160	32
							84

1- $H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

$$S_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{X}_i - \overline{\overline{X}})^2}{k-1}}$$

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k} = \frac{84}{3} = 28$$

$$S_{\overline{X}} = \sqrt{\frac{(22-28)^2 + (30-28)^2 + (32-28)^2}{3-1}} = 5{,}29$$

2- $KO_{G.Arası} = S_{\overline{X}}^2 . n = (5{,}29)^2 . (5) = 139{,}92$

3- $KO_{G.İçi} = \dfrac{\Sigma (X_i - \overline{X})^2}{n-1}$

$$S_1^2 = \frac{(29-22)^2 + (16-22)^2 + (36-22)^2 + (15-22)^2 + (14-22)^2}{5-1}$$

$$= 98{,}5$$

$$S_2^2 = \frac{(41-30)^2 + (28-30)^2 + (33-30)^2 + (11-30)^2 + (37-30)^2}{5-1}$$

$$= 136$$

$$S_3^2 = \frac{(49-32)^2 + (29-32)^2 + (38-32)^2 + (25-32)^2 + (19-32)^2}{5-1}$$

$$= 138$$

$$KO_{G.içi} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2}{(n_1-1) + (n_2-1) + (n_3-1)}$$

$$= \frac{(5-1)(98{,}5) + (5-1)(136) + (5-1)(138)}{(5-1) + (5-1) + (5-1)} = 124{,}16$$

4- SD_1= k-1 = 3-1=2

SD_2= [(n.k)-k] = 15-3=12

$F_{\alpha,SD1,SD2} = F_{0,05,2,12}$= 3,89

$$F = \frac{KO_{G.arası}}{KO_{G.İçi}} = \frac{139,92}{124,16} = 1,12$$

1,12<3,89 olduğundan H_0 hipotezi kabul edilir. Ortalamalar arasındaki fark anlamsızdır.

12.2. TEK YÖNLÜ VARYANS ANALİZİ

Tek yönlü varyans analizinde k sayıdaki örnek ortalaması arasındaki fark test edilirken doğrusal model kullanılır. Ortalamaları test edilecek k sayıdaki anakütleyi $A_1,A_2,....,A_k$ ile anakütle ortalamalarını μ_i ve varyanslarını σ^2 ile ifade edelim. Bu anakütleler deneme grubu olarak adlandırılmaktadır. Bu anakütlelerin biraraya gelmesi ile ortalaması μ olan bir anakütle oluşacaktır. Bu büyük anakütlenin ortalaması,

$$\mu = \frac{\sum_{i=1}^{k} \mu_i}{k}$$

olacaktır.

Tek yönlü varyans analizinde temel hipotez daha önce sözedildiği gibi,

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

olarak oluşturulacaktır. Temel hipotezin doğru olması durumunda tüm ortalamalar birbirlerine eşit olduğu gibi, bunlar tüm anakütlelerin ortalaması μ'ye de eşit olacaktır.

$$\mu = \mu_1 = \mu_2 = ... = \mu_k = \mu_i$$

Temel hipotezin doğru olmaması durumunda ise $\mu \neq \mu_i$ olacak ve aralarında a_i ile ifade edebileceğimiz farklar olacaktır.

$$\alpha_i = \mu_i - \mu$$

Burada α_i, A_i anakütlesinin işleyim etkisi olarak tanımlanmaktadır. Bu etki dikkate alınırsa, temel hipotez doğru ise,

$$\alpha_1 = \alpha_2 = ... = \alpha_k = 0$$

$$\sum_{i=1}^{k} \alpha_i = \sum_{i=1}^{k} (\mu_i - \mu) = 0$$

olacaktır. Yani deneme etkileri sıfırdır.

Normal dağılmış herhangi bir i. örneğin j. birimi için,

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

yazılabilir. Burada ε_{ij} hata terimi veya artıktır. ε_{ij}'lerin birbirinden bağımsız oldukları, sıfır ortalama ve σ^2 varyansı ile normal dağıldıkları kabul edilmektedir.

Daha önce $\alpha_i = \mu_i - \mu$ olarak tanımlanmıştı. Buradan,

$$\mu_i = \mu + \alpha_i$$

olarak elde edilir. Bu eşitlik X_{ij} eşitliğinde yerine konursa,

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

bulunur. Elde edilen bu model tek yönlü varyans analizi modelidir.

Tek yönlü varyans analizinde kullanılacak veri Tablo 12.1.'de görüldüğü gibi bir tablo oluşturulacaktır.

Tablo 12.1

Gözlemler	**Deneme Grupları**	**(j)**		
	1	2	.....	k
1	X_{11}	X_{21}	.....	X_{k1}
2	X_{12}	X_{22}	.....	X_{k2}
...	...	....		
N	X_{1n1}	X_{2n2}	.....	X_{knk}
Toplam	$\sum_{j=1}^{n_1} X_{1j}$	$\sum_{j=1}^{n_2} X_{2j}$	.....	$\sum_{j=1}^{n_k} X_{kj}$
n_i	n_1	n_2	.....	n_k

Bu verilerden örnek ortalamaları,

$$\overline{X}_i = \frac{X_{i1} + X_{i2} + ... + X_{in_i}}{n_i} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}$$

tüm örneklerin birleşmesi ile oluşacak büyük örneğin ortalaması,

$$\overline{\overline{X}} = \frac{\sum_{j=1}^{n_1} X_{1j} + \sum_{j=1}^{n_2} X_{2j} + ... + \sum_{j=1}^{n_k} X_{kj}}{N} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{N}$$

$$N = n_1 + n_2 + \ldots\ldots + n_k = \sum_{i=1}^{k} n_i$$

olarak hesaplanacaktır.

Örnek birimleri X_{ij}'lerin genel ortalama $\overline{\overline{X}}$'dan sapmaları, örnek birimlerinin kendi deneme ortalamalarından sapmaları ile deneme ortalamasının genel ortalamadan farklarının toplamına eşit olduğundan,

$$X_{ij} - \overline{\overline{X}} = (X_{ij} - \overline{X}_i) + (\overline{X}_i - \overline{\overline{X}})$$

olarak yazılabilir.

Hata teriminin tahmincisi $(X_{ij} - \overline{X}_i)$ ve deneme etkisinin tahmincisi $(\overline{X}_i - \overline{\overline{X}})$ olacağından örnek birimlerinin büyük ortalamadan farkı bunların toplamına eşittir.

Yukarıda verilen eşitlikten sapmaların kareleri toplamı hesaplanabilir.

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{\overline{X}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[(X_{ij} - \overline{X}_i)^2 + (\overline{X}_i - \overline{\overline{X}})\right] =$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\overline{X}_i - \overline{\overline{X}})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n_1}(X_{ij} - \overline{X}_i)(\overline{X}_i - \overline{\overline{X}})$$

verilen eşitliğin son ifadesinde yer alan $(\overline{X}_i - \overline{\overline{X}})$, i. işlemdeki tüm gözlemler j.gözlemler için sabittir. Aynı ifade de yeralan $(X_{ij} - \overline{X}_i)$ toplamı ise, i. işlemdeki tüm j. gözlemler için sıfırdır. Bu nedenle,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{\overline{X}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\overline{X}_i - \overline{\overline{X}})^2$$

olacaktır. Burada eşitliğin sol tarafı genel sapma karelerin toplamların toplamıdır.

$$KT_G = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{\overline{X}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^{\,2} - N(\overline{\overline{X}})^2$$

Eşitliğin sağ tarafındaki ilk ifade deneme içi sapma kareler toplamıdır.

$$KT_{D.içi} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^{\,2} - \sum_{i=1}^{k} n_i(\overline{X})^2$$

Eşitliğin sağ tarafındaki ikinci ifade ise, denemeler arası sapma kareler toplamıdır.

$$KT_{D.arası} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\overline{X}_i - \overline{\overline{X}})^2 = \sum_{i=1}^{k} n_i \overline{X}_i^2 - N\overline{\overline{X}}^2$$

Bu tanımlamalara göre kısaca,

$KT_G = KT_{D.içi} + KT_{D.arası}$

olacaktır. Bu açıklamalardan da anlaşılacağı gibi birimler arasındaki fark tesadüfi sapmalardan ve farklı anakütlelerden oluşmaktadır.

Tek yönlü varyans analizinde serbestlik dereceleri KT_G için (N-1), $K_{TD.arası}$ için (k-1) ve SSE için (N-k)'dır. Varyans analizinde σ^2'nin tahmincileri kareler ortalaması (KO) olarak adlandırılır. Kareler ortalaması ilgili sapma karelerin ilgili serbestlik derecelerine bölünmesi ile bulunacaktır.

$$KO_{D.arası} = \frac{KT_{D.arası}}{k-1}$$

$$KO_{D.içi} = \frac{KT_{D.İ.İ}}{N-k}$$

$\frac{KO_{D.arası}}{KO_{D.içi}}$ dağılımı (k-1) ve (N-k) serbestlik dereceli F

dağılımıdır. Test istatistiği,

$$F = \frac{KO_{D.arası}}{KO_{D.içi}}$$

olarak hesaplanır.

Bu açıklamalara göre varyans analizi tablosu Tablo 12.2. görüldüğü gibi oluşacaktır.

Tablo 12.2

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F
Deneme Arası	$KT_{D.arası}$	k-1	$KO_{D.arası} = \frac{KT_{D.arası}}{k-1}$	
Deneme İçi	$KT_{D.İçi}=KT_G-KT_{D.arası}$	N-k	$KO_{D.içi} = \frac{KT_{D.İ.İ}}{N-k}$	$F=\frac{KO_{D.arası}}{KO_{D.içi}}$ $F = \frac{MSA}{MSE}$
TOPLAM	SST	N-1		

Tek yönlü varyans analizinde hipotezler daha önce açıklandığı gibi oluşturulur. Belirlenen α hata payı ve (k-1), (N-k) serbestlik dereceleri ile F tablosundan tablo değeri bulunur. Test istatistiği,

$$F=\frac{KO_{D.aras}}{KO_{D.içi}}$$

olarak hesaplanarak, $F < F_{\alpha,k-1,N-k}$ ise H_0 , $F > F_{\alpha,k-1,N-k}$ ise H_1 hipotezi kabul edilir.

ÖRNEK: Aşağıda öğrencilerin 50 üzerinden üç dersten aldıkları notlar verilmiştir.

Dersler	**Notlar**				
Matematik	29	16	36	15	14
İstatistik	41	28	33	11	37
Bilgisayar	49	29	38	25	19

0,05 hata payı ile ortalamalar arasındaki farkların anlamlılığını (örnek ortalamalarının aynı anakütleden alınıp alınmadığını) inceleyiniz.

Dersler	**Notlar**					**Notların Toplamı**	**Toplamlar kareleri**
Matematik	29	16	36	15	14	110	12100
İstatistik	41	28	33	11	37	150	22500
Bilgisayar	49	29	38	25	19	160	25600
						420	60200

1- $H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

$F_{\alpha,k-1,N-k} = 3{,}89$

$\Sigma X_1 = 110 \qquad (\Sigma X_1)^2 = 12100$

$\Sigma X_2 = 150 \qquad (\Sigma X_2)^2 = 22500$

$\Sigma X_3 = 160 \qquad (\Sigma X_3)^2 = 25600$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\Sigma X_1 + \Sigma X_2 + \Sigma X_3) = 420$$

$$\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}\right)^2 = 176400$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 = (29)^2 + (16)^2 + (36)^2 + (15)^2 + (14)^2 + (41)^2 + (28)^2$$

$$+ (33)^2 + (11)^2 + (37)^2 + (49)^2 + (29)^2 + (38)^2 + (25)^2$$

$$+ (19)^2 = 13530$$

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{N} = \frac{420}{15} = 28$$

$$KT_G = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^{\ 2} - N\overline{\overline{X}}^2$$

$$= 13530 - 15(28)^2 = 13530 - 11760 = 1770$$

$$KT_{D.arası} = \sum_{i=1}^{k} n_i \overline{X}_i^2 - N\overline{\overline{X}}^2$$

$$= \left[5\left(\frac{110}{5}\right)^2 + 5\left(\frac{150}{5}\right)^2 + 5\left(\frac{160}{5}\right)^2\right] - 15(28)^2$$

$$= (2420 + 4500 + 5120) - 11760$$

$$= 12040 - 11760 = 280$$

$$KT_{D.İ.İ} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^{\ 2} - \sum_{i=1}^{k} n_i (\overline{X})^2$$

$$= 13530 - 12040 = 1490$$

veya

$$KT_G = KT_{D.İÇİ} + KT_{D.arası}$$

$$1770 = 280 + 1490$$

ilişkisinden yararlanılarak ikisi hesaplanıp üçüncü bu bağıntıdan bulunabilir.

$$KO_{D.arası} = \frac{KT_{D.arası}}{k-1} = \frac{280}{2} = 140$$

$$KO_{D.içi} = \frac{KT_{D.İ.İ}}{N-k} = \frac{1490}{12} = 124{,}16$$

$$F = \frac{KO_{D.arası}}{KO_{D.içi}} = \frac{140}{124{,}16} = 1{,}12$$

ANOVA Tablosu

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F
Denemeler Arası	280	3-1=2	140	
Denemeler İçi	1490	15-3=12	124,66	F=1,12
TOPLAM	1770	14		

$1.12 < 3{,}89$ olduğundan H_0 hipotezi kabul edilir.

12.3. ÇİFT YÖNLÜ VARYANS ANALİZİ

Tek yönlü varyans analizinde incelenenin aksine bazı olaylarda değişkenlerdeki değişmeler iki faktörden kaynaklanır ve bunlar ile ilgili bilgiler bulunabilir. Bu durumda değişmeler iki bölümden oluşmaktadır. Bu nedenle tek yönlü varyans analizi yerine çift yönlü varyans analizi yapılması gerekmektedir. Örneğin daha önce öğrencilerin dersleri ile ilgili örnek çözmüştük. Bu örnekte sadece öğrencilerin derslerden aldıkları notlar yer almıştı. Oysa öğrencilerin orta eğitim yaptıkları okullar, o okulların eğitim şekilleri veya orta öğretimde mezun oldukları bölümler derslerdeki başarıyı etkiliyor olabilir. Bu bilgilerinde eklenmesi ile oluşacak veri tek yönlü analiz etkilemeyeceğinden, çift yönlü varyans analizi ile ortalamalar arasındaki farkların test edilmesi gerekecektir.

Farklı iki değişken sözkonusu olduğunda, değişkenlerin birbirlerine olan etkileri veya bağımsız olup olmadıkları da önemlidir. Bazı olaylarda değişkenler karşılıklı olarak birbirlerini etkilerler, yani bağımsız değillerdir. Bazı olaylarda ise karşılıklı etkileşim sözkonusu değildir. Bu nedenle çift yönlü varyans analizinin etkileşim olmaması ve etkileşim olması durumları için ayrı ayrı incelenmesi gerekmektedir.

Veriler iki değişken için düzenlendiğinde bir matris veya tablo oluşturulur. Çift yönlü varyans analizinde tablonun veya matrisin hücrelerinde birer gözlem yer alması durumu tesadüfi blok düzeni olarak tanımlanmaktadır. Bazı durumlarda ise birden fazla eşlenmiş gözlemler sözkonusudur. Bu gözlemlerin oluşturduğu gruplar blok olarak adlandırılır. Çift yönlü varyans analizinde blokların etkisi sözkonusu olmaktadır.

12.3.1. Etkileşimsiz Çift Yönlü Varyans Analizi

Çift yönlü varyans analizinde analize değişmeleri açıklayacak yeni kaynak eklenmesi ile hata kareler ortalaması (KO_{Hata}) azaltılabilir. Bu durumda bloklarında analize katılması ile varyans analizi modeli genişleyecektir. Etkileşim olmaması durumunda çift yönlü varyans analizi modeli,

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

olur. Burada,

X_{ij} : i. işleyimin j. gözlemini,

μ : genel anakütle ortalamasını

α_i : i. deneme etkisini

β_j : j. blok etkisini

ε_{ij} : hata terimini

ifade etmektedir. X_{ij}'nin dağılımı ortalaması μ ve varyansı σ^2 olan normal dağılım, ε_{ij}'nin dağılımı ise ortalaması sıfır, varyansı σ^2 olan normal dağılımdır. Ayrıca $\sum_{i=1}^{k} \alpha_i = 0$ ve $\sum_{j=1}^{n} \beta_j = 0$ 'dır.

Modelde α_i ve β_j ile ifade edilen etkiler yer aldığından temel hipotez bu etkilerin sıfır olacağını ifade edecek şekilde oluşturulacaktır.

$$H_0 : \alpha_i = 0 \qquad i = 1,2,3,.....,k$$
$$\beta_j = 0 \qquad j = 1,2,3,.....,n$$

Karşıt hipotez en az bir α_i veya en az bir β_j'nin sıfırdan farklı olduğunu ifade etmektedir.

Örnek birimlerinin değerlerinin (X_{ij} gözlemlerinin) büyük örnek ortalamasından farkları,

$$X_{ij} - \overline{\overline{X}} = \left(X_{ij} - \overline{X}_i - \overline{X}_j + \overline{\overline{X}}\right) + \left(\overline{X}_i - \overline{\overline{X}}\right) + \left(\overline{X}_j - \overline{\overline{X}}\right)$$

olacaktır. Sapmaları kareleri toplamı alınırsa,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{ij} - \overline{\overline{X}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_i + \overline{X}_j + \overline{\overline{X}}\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_i - \overline{\overline{X}}\right)^2$$

$$+ \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_j^{\,2} - \overline{\overline{X}}\right)^2$$

olacaktır. Burada $\overline{X}_i$, i. işleyimin gözlemlerinin ortalamasını, $\overline{X}_j$, j. işleyimin gözlemlerinin ortalamasını ve $\overline{\overline{X}}$ ise tüm gözlemlerin ortalamasını ifade etmektedir. Eşitliğin sol tarafı KT_G, sağ tarafı ise sırasıyla KT_H, $KT_{D.arası}$ ve $KT_{B.arası}$'dir. $KT_{B.arası}$ blokların kareleri toplamıdır. KT_G ve $KT_{D.arası}$ daha önce verilen formülerle,

$$KT_G = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^{\ 2} - N\overline{\overline{X}}^2$$

$$KT_{D.arası} = \sum_{i=1}^{k} n_i \overline{X}_i^2 - N\overline{\overline{X}}^2$$

$KT_{Blok\ arası}$ ise,

$$KT_{B.arası} = \sum_{i=1}^{k}\sum_{j=1}^{n} (\overline{X}_j^2 - \overline{\overline{X}})^2 = \sum_{j=1}^{n} \overline{X}_j^2 - nk\overline{\overline{X}}^2$$

olarak bulunacaktır. Burada KT_{Hata} daha önce verilen toplamlar yardımı ile,

$$KT_{Hata} = KT_G - KT_{D.arası} - KT_{B.arası}$$

olarak belirlenebilir. Hesaplanan bu kareler toplamları ile ortalama kareler (MS) şöyle hesaplanacaktır.

$$KO_{D.arası} = \frac{KT_{D.arası}}{k-1}$$

$$KO_{B.arası} = \frac{KT_{B.arası}}{n-1}$$

$$KO_{hata} = \frac{KT_{Hata}}{(k-1)(j-1)}$$

$KO_{D.arası}$ ve $KO_{B.arası}$ için iki F test istatistiği ise,

$$F = \frac{KO_{D.arası}}{KO_{hata}}$$

$$F = \frac{KO_{B.arası}}{KO_{Hata}}$$

olacaktır. Bu test istatistiklerinin birincisi için S_{D1}=(k-1) ve S_{D2}=(k-1)(n-1); ikincisi için S_{D1}=(j-1) ve S_{D2}=(k-1)(n-1) serbestlik dereceleri ve belirlenen α hata payı ile F dağılımı tablosundan bulunacak değerler ile karşılaştırılarak karar verilir. Test hipotezlerde yeralan α_i ve β_j'ler için ayrı ayrı yapılmış olacaktır.

Bu durumda oluşacak varyans analizi tablosu Tablo 12.3.'te görülmektedir.

Tablo 12.3

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kare Ortalaması	F
Gruplar Arası (A)	$KT_{D.arası}$	k-1	$KO_{D.arası}$	$F = \frac{KO_{D.arası}}{KO_{hata}}$
Bloklar Arası (B)	$KT_{B.arası}$	n-1	$KO_{B.arası}$	$F = \frac{KO_{B.arası}}{KO_{Hata}}$
Örnekleme Hatası(E)	KT_H	(k-1)(j-1)	KO_H	
TOPLAM (T)		N-1		

Yukarıdaki tablodan da görüldüğü gibi tek yönlü varyans analizi tablosuna SSB'nin eklenmesi ile çift yönlü varyans analizi tablosu oluşturulmuştur. Bu tablo etkileşim olmaması durumunda çift yönlü varyans analizi tablosudur. Etkileşim olması durumunda tablo değişecektir.

ÖRNEK: Liselerin A,B,C,D ve E bölümlerinden mezun üç öğrencinin 100 üzerinden 3 dersten aldıkları final notları aşağıda verilmiştir.

	Dersler		
Bölümler	**Matematik**	**İstatistik**	**Bilgisayar**
A	77	64	87
B	43	51	62
C	86	79	57
D	27	36	48
E	59	42	76

0,05 hata payı ile öğrencilerin final notları ve liselerden mezun oldukları bölümler arasında fark olup olmadığını test ediniz.

	Dersler(i)				
Bölümler (j)	**Matematik**	**İstatistik**	**Bilgisayar**	**Toplam**	**Ortalama ($\overline{X}_j$)**
A	77	64	87	228	76
B	43	51	62	156	52
C	86	79	57	222	74
D	27	36	48	111	37
E	59	42	76	177	59
Toplam	292	272	330	894	
Ortalama ($\overline{X}_i$)	58,4	54,4	66		$\overline{\overline{X}} = 59,6$

1-

$H_1 : \alpha_i \neq 0$

$\beta_j \neq 0$

$\alpha = 0{,}05$

$F_{\alpha,k-1,(k-1)(n-1)} = F_{0,05,2,8} = 4{,}46$

$F_{\alpha,j-1,(k-1)(n-1)} = F_{0,05,4,8} = 3{,}84$

2- $KT_G = \sum_{i=1}^{3}\sum_{j=1}^{5} X_{ij}^{\ 2} - N\overline{\overline{X}}^2$

$$\sum_{i=1}^{3}\sum_{j=1}^{5} X_{ij}^{\ 2} = (77)^2 + (43)^2 + (86)^2 + (27)^2 + (59)^2 + (64)^2 + (51)^2 +$$

$$(79)^2 + (36)^2 + (42)^2 + (87)^2 + (62)^2 + (57)^2 + (48)^2 +$$

$$(76)^2 = 58124$$

$KT_G = 58124\text{-}15(59{,}6)^2 = 4841{,}6$

$$KT_{D.arası} = \sum_{i=1}^{3} n\overline{X}_i^2 - N(\overline{\overline{X}})^2$$

$$= \left[\left(\frac{292}{5}\right)^2 + \left(\frac{272}{5}\right)^2 + \left(\frac{330}{5}\right)^2 5\right] - 15(59{,}6)^2$$

$$= 53629{,}6 - 53282{,}4 = 347{,}2$$

$$KT_{B.arası} = \sum_{j=1}^{5} \overline{X}_j^2 - nk(\overline{\overline{X}})^2$$

$$= \left[\frac{(228)^2}{3} + \frac{(156)^2}{3} + \frac{(222)^2}{3} + \frac{(111)^2}{3} + \frac{(177)^2}{3}\right] - 15(59{,}6)^2$$

$$= 56418 - 53282{,}4 = 3135{,}6$$

$KT_{Hata} = KT_G - KT_{D.arası} - KT_{B.arası}$

$= 4841{,}6 - 347{,}2 - 3135{,}6 = 1358{,}8$

$$KO_{D.arası} = \frac{KT_{D.arası}}{k-1} = \frac{347{,}2}{3-1} = 173{,}6$$

$$KO_{B.arası} = \frac{KT_{B.arası}}{n-1} = \frac{3135,6}{5-1} = 783,9$$

$$KO_{hata} = \frac{KT_{Hata}}{(k-1)(j-1)} = \frac{1358,8}{(3-1)(5-1)} = 169,85$$

$$F = \frac{KO_{D.arası}}{KO_{hata}} = \frac{173,6}{169,85} = 1,02$$

$$F = \frac{KO_{B.arası}}{KO_{Hata}} = \frac{783,9}{169,85} = 4,61$$

1,02 < 4,46 olduğundan H_0 hipotezi kabul edilir. ($\alpha_i = 0$)

4,61 < 3,84 olduğundan H_1 hipotezi kabul edilir. ($\beta j \neq 0$)

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kare Ortalaması	F
Deneme Arası (A)	347,2	3-1=2	173,6	1,02
Bloklar Arası (B)	3135,6	5-1=4	783,9	4,61
Örnekleme Hatası(E)	1358,8	(3-1)(5-1)=8	169,85	
TOPLAM	4841,6			

12.3.2. Etkileşimli Çift Yönlü Varyans Analizi

Çift yönlü varyans analizinde, iki faktör arasında karşılıklı etkileşim de olabilir. Deneme grupları ile bloklar arasındaki etkiyi de modele dahil edersek, doğrusal çift yönlü varyans analizi modeli,

$$X_{ijh} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijh}$$

i =1,2,3,.....,k

j =1,2,3,.....,n

h =1,2,3,.....,r

olacaktır. Burada,

r : bloklarla deneme grupları birleşiminin sayısını
Xijh : gözlemleri
μ : büyük anakütle ortalamasını
α_i : i. deneme etkisini
β_j : j. blok etkisini
γ_{ij} : i. işleyim ile j. deneme arasındaki etkiyi (etkileşimi)
ε_{ijh} : hata terimini

ifade etmektedir. Gözlemler ve hata teriminin dağılımı normal dağılımdır.

Temel hipotez oluşturulurken etkileşim olmaması durumundan farklı olarak faktörler arası etkileşiminde dikkate alınması gerekmektedir. Burada temel hipotez deneme etkilerinin, blok etkilerinin ve karşılıklı etkileşim sıfır olduğunu belirtecektir.

$$H_0 : \alpha_i = 0 \qquad i = 1,2,3,....., k$$
$$H_0 : \beta_j = 0 \qquad j = 1,2,3,....., n$$
$$H_0 : \gamma_{ij} = 0 \qquad i = 1,2,3,....., k \quad ; \quad j = 1,2,3,....., n$$

Karşıt hipotez ise temel hipotezin doğru olmadığını belirtecek şekilde oluşturulacaktır.

$$H_1 : \alpha_i \neq 0$$
$$H_1 : \beta_j \neq 0$$
$$H_1 : \gamma_{ij} \neq 0$$

Karşıt hipotez α_i ve β_j için en az bir değerin, γ_{ij} için en az bir çifttin değerinin sıfırdan farklı olduğunu ifade etmektedir.

Örnek gözlemlerinin genel örnek ortalamasından farkları,

$$(X_{ijh} - \overline{\overline{X}}) = \left(X_{ijh} - \overline{X}_{ij}\right) + (\overline{X}_i + \overline{\overline{X}}) + \left(\overline{X}_j - \overline{\overline{X}}\right) + \left(\overline{X}_{ij} - \overline{X}_i - \overline{X}_j + \overline{\overline{X}}\right)$$

olacaktır. Sapmaların kareleri toplamı alınırsa,

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{h=1}^{r}\left(X_{ijh} - \overline{\overline{X}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{h=1}^{r}(X_{ijh} - \overline{X}_{ij})^2 + nr\sum_{i=1}^{k}\left(X_i - \overline{\overline{X}}\right)^2$$

$$+ kr\sum_{j=1}^{n}\left(\overline{X}_j - \overline{\overline{X}}\right)^2 + r\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_{ij} - \overline{X}_i - \overline{X}_j + \overline{\overline{X}}\right)^2$$

olacaktır. Burada $\overline{X}_i$, i. işleyim grubunun; $\overline{X}_j$, j. bloğun gözlemlerinin ortalamasıdır. $\overline{X}_{ij}$ ise i. işleyim grubu ile j. bloğun r sayıda gözleminin ortalamasıdır. $\overline{\overline{X}}$ büyük örnek ortalaması, yani (knr) gözlemin ortalamasıdır. Eşitliğin sol tarafı KT_G, sağ tarafı ise sırasıyla KT_H, $KT_{D.arası}$, $KT_{B.arası}$ ve etkileşim için tanımlanan $KT_{etkileşim}$'dır.

Etkileşim olması durumunda çift yönlü varyans analizinde de KT_G, $KT_{D.arası}$, $KT_{B.arası}$ daha önce açıklandığı gibi hesaplanır. $KT_{etkileşim}$ ise,

$$KT_{etkileş} = \frac{1}{r}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\sum_{h=1}^{r}X_{ijh}\right)^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n}X_{ij}\right)^2}{N} - SSA - SSB$$

olacaktır. KT_H yine toplamlardan hareketle,

$KT_H = KT_G - KT_{D.arası} - KT_{B.arası} - KT_{etkileşim}$

olarak bulunabilir.

Hesaplanan bu kareler toplamı ile ortalama kareler (MS) şöyle hesaplanacaktır.

$$KO_D = \frac{KT_D}{k-1}$$

$$KO_B = \frac{KT_B}{n-1}$$

$$KO_E = \frac{KT_E}{(k-1)(n-1)}$$

$$KO_H = \frac{KT_H}{kn(r-1)}$$

KO_D, KO_B ve KO_E için üç F test istatistiği ise,

$$F = \frac{KO_D}{KO_H}$$

$$F = \frac{KO_B}{KO_H}$$

$$F = \frac{KO_E}{KO_H}$$

olacaktır. Bu test istatistiklerinin birincisi için S_{D1}=(k-1) ve S_{D2}=kn(r-1); ikincisi için S_{D1}=(n-1) ve S_{D2}=kn(r-1); üçüncüsü için S_{D1}=(k-1)(n-1) ve S_{D2}=kn(r-1) serbestlik dereceleri ve belirlenen α hata payı ile F dağılımı tablosundan bulunacak değerler ile karşılaştırılarak karar verilir. Test, hipotezlerde yeralan α_i, β_j ve γ_{ij}'ler için ayrı ayrı yapılmış olacaktır. Bu durumda oluşacak varyans analizi tablosu Tablo 11.4.'te görülmektedir.

Tablo 12.4.

Değişim Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F
Gruplar Arası (A)	KT_D	k-1	KO_D	$F = \frac{KO_D}{KO_H}$
Bloklar Arası (B)	KT_B	n-1	KO_B	$F = \frac{KO_B}{KO_H}$
Gruplar Arası ve Bloklararası Etkileşim(I)	KT_E	(k-1)(n-1)	KO_E	$F = \frac{KO_E}{KO_H}$
Örnekleme Hatası(E)	KT_H	Kn(r-1)	KO_H	
TOPLAM	KT_G	N-1		

Görüldüğü gibi etkileşim olmaması durumu ile etkileşim olması durumunda çift yönlü varyans analizi arasındaki fark, faktörler arası etkileşimin üçüncü F istatistiği ile test edilmesidir.

ÖRNEK: Bir önceki örnekte 3 öğrenci için verilen notlar, 6 öğrenci için aşağıda verilmiştir.

	Dersler(i)		
Bölümler (j)	**Matematik**	**İstatistik**	**Bilgisayar**
A	77	64	87
	81	70	89
B	43	51	62
	55	40	70
C	86	79	57
	91	80	87
D	27	36	48
	51	42	63
E	59	42	76
	67	54	62

0,05 hata payı ile tüm farkları test ediniz.

		Dersler(i)			
Bölümler (j)	**Matematik**	**İstatistik**	**Bilgisayar**	**Toplam**	**Ortalama**
A	77	64	87	468	78
	81	70	89		
B	43	51	62	321	53,5
	55	40	70		
C	86	79	57	480	80
	91	80	87		
D	27	36	48	267	44,5
	51	42	63		
E	59	42	76	360	60
	67	54	62		
Toplam	637	558	701	1896	
Ortalama	63,7	55,8	70,1		$\overline{\overline{X}} = 63{,}2$

1-

$$H_1 : \alpha_i \neq 0$$
$$: \beta_j \neq 0$$
$$: \gamma_{ij} \neq 0$$

2- $F_{\alpha,k-1,nk(r-1)} = F_{0,05,2,15} = 3,68$

$F_{\alpha,n-1,nk(r-1)} = F_{0,05,4,15} = 3,06$

$F_{\alpha,(k-1),(n-1),nk(r-1)} = F_{0,05,8,15} = 2,64$

3-

$$KT_G = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{h=1}^{r} X_{ijh}^2 - N(\overline{\overline{X}})^2$$

$$= (77)^2 + (81)^2 + ... + (62)^2 - 30(63,2)^2$$

$$= 128864 - 30(63,2)^2 = 9036,8$$

$$KT_D = \sum_{i=1}^{3} n\overline{X}_i^2 - N\left(\overline{\overline{X}}\right)^2$$

$$= \frac{(637)^2}{10} + \frac{(558)^2}{10} + \frac{(701)^2}{10} - 119827,2$$

$$= 120853,4 - 119827,2 = 1026,2$$

$$KT_B = \sum_{j=1}^{n} \overline{X}_j^2 - nk\left(\overline{\overline{X}}\right)^2$$

$$= \frac{(468)^2}{6} + \frac{(321)^2}{6} + \frac{(480)^2}{6} + \frac{(267)^2}{6} + \frac{(360)^2}{6} - 119827,2$$

$$= 125559 - 119827,2 = 5731,8$$

$$KT_E = \frac{1}{r}\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\sum_{h=1}^{r} X_{ijh}\right) - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}\right)^2}{N} - SSA - SSB$$

$$= \frac{1}{2}[(77+81)^2 + (43+55)^2 + (86+91)^2 + (27+51)^2 + (59+67)^2 +$$

$$(64+70)^2 + (51+40)^2 + (79+80)^2 + (36+42)^2 + (42+54)^2 +$$

$$(87+89)^2 + (62+70)^2 + (57+87)^2 + (48+63)^2 + (76+62)^2] -$$

$$119827,2 - 2154,2 - 5731,8$$

$$= 127588 - 119827,2 - 1026,2 - 5731,8 = 1002,8$$

$$KT_H = KT_G - KT_D - KT_B - KT_E$$
$$= 9036,8 - 1026,2 - 5731,8 - 1002,8$$
$$= 1276$$

$$KO_D = \frac{KT_D}{n-1} = \frac{1026,2}{2} = 513,1$$

$$KO_B = \frac{KT_B}{n-1} = \frac{5731,8}{4} = 1432,95$$

$$KO_E = \frac{KT_E}{(k-1)(n-1)} = \frac{1002,8}{8} = 125,35$$

$$KO_H = \frac{KT_H}{kn(r-1)} = \frac{1276}{15} = 85,06$$

$$F = \frac{KO_D}{KO_H} = \frac{513,1}{85,06} = 6,03$$

$$F = \frac{KO_B}{KO_H} = \frac{1432,95}{85,06} = 16,84$$

$$F = \frac{KO_E}{KO_H} = \frac{125,35}{85,06} = 1,47$$

$6,03 > 3,68$ olduğundan H_1 hipotezi kabul edilir.α_i'ler anlamlı.($\alpha_i \neq 0$)
$16,84 > 3,06$ olduğundan H_1 hipotezi kabul edilir.β_j'ler anlamlı.($\beta_j \neq 0$)
$1,47 < 2,64$ olduğundan H_0 hipotezi kabul edilir.γ_{ij}'ler anlamsız.($\gamma_{ij} = 0$)

Değişim Kaynağı	**Kareler Toplamı**	**Serbestlik Derecesi**	**Kareler Ortalaması**	**F**
Deneme Arası (A)	1026,2	3-1=2	513,1	6,03
Deneme Arası (B)	5731,8	5-1=4	1432,95	16,84
Gruplar Arası Ve Bloklararası Etkileşim(I)	1002,8	(3-1)(5-1)=8	125,35	1,47
Örnekleme Hatası(E)	1276	(3)(5)(2-1)=15	85,06	
TOPLAM	9036,8			

12.4. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir paketleme fabrikası meyva sularının paketlenmesi için üç ayrı teknolojiye sahip dolum makinesi kullanmaktadır. Yapılacak yeni yatırım için teknoloji seçimi yapılacaktır. Kullanılan teknolojinin dolum ağırlıkları üzerinde etkinliği incelenmek istenmektedir. Standart net 1 litre olması gereken paketlerden her makineden 5 adet seçilmiştir ve ağırlık ml olarak ölçülmüştür. %95 güven düzeyinde test ediniz.

Makine1	Makine2	Makine3
1007	995	1002
1001	900	996
1009	1011	1000
996	1000	998
1002	1019	1004
5015	4925	5000

Çözüm:

$$\sum_{j=1}^{5} X_{1j} = 5015 \quad , \quad \sum_{j=1}^{5} X_{2j} = 4925 \quad , \quad \sum_{j=1}^{5} X_{3j} = 5000$$

$$\overline{X}_1 = \frac{5015}{5} = 1003 \quad , \quad \overline{X}_2 = \frac{4925}{5} = 985 \quad , \quad \overline{X}_3 = \frac{5000}{5} = 1000$$

şeklinde örnek ortalamaları hesaplanır. Genel ortalama ise ,

$$\sum_{i=1}^{3}\sum_{j=1}^{5} X_{ij} = \sum_{i=1}^{3} \overline{X}_i n_i = 14940 \quad , \quad \overline{\overline{X}} = \frac{14940}{15} = 996$$

Toplam Varyans : $$S_t^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{5}\left(X_{ij} - \overline{\overline{X}}\right)^2}{n-1} = \frac{10458}{14} = 747$$

Gruplar Arası Varyans: $$S_b^2 = \frac{\sum_{i=1}^{3} n_i\left(\overline{X}_i - \overline{\overline{X}}\right)^2}{k-1} = \frac{930}{2} = 465$$

Gruplar İçi Varyans : $$S_w^2 = \frac{\sum_{i=1}^{3}\sum_{j=1}^{5}\left(X_{ij} - \overline{X}_i\right)^2}{n-k} = \frac{9528}{12} = 794$$

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : Ortalamaların en az biri diğerlerinden farklıdır.

Red bölgesi $F > F_{v1,v2,\alpha}$

Test istatistiği $F = \dfrac{S_b^2}{S_w^2} = 0{,}586 < F_{2,12,0,05} = 3{,}88$

Ortalamaların farklılığını gösterecek yeterli kanıt bulunamamıştır.

ÖRNEK 2: Ekonometri, Maliye ve İktisat bölümlerinde verilen "İstatistiğe Giriş" derslerindeki başarının okunulan bölüme göre farklılık gösterip göstermediği incelenecektir. Bölümlerden öğrenci mevcuduna göre %5 oranında örneklemler oluşturulmuş ve başarı notları listelenmiştir. %1 anlamlılık düzeyinde test ediniz.

Ekonometri	Maliye	İktisat
20	10	10
70	40	20
80	45	25
90	65	45
	90	50
		80
		85

Çözüm:

Grup(i)	$\sum_{j=1}^{n_i} X_{ij}$	n_i	$\overline{X}_i$	$\sum_{j=1}^{n_i} (X_{ij})^2$	$n_i\left(\overline{X}_i - \overline{\overline{X}}\right)^2$
1	260	4	65	19800	722,27
2	250	5	50	16050	12,21
3	315	7	45	19275	301,46
Toplam	825	16	51,5625	55125	1035,94

$$S_w^2 = \frac{\sum_{i=1}^{3}\left[\sum_{j=1}^{n_i}(X_{ij})^2 - n_i(\overline{X}_i)^2\right]}{n-k} = \frac{11550}{16-3} = 888{,}46$$

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{n} = \frac{825}{16} = 51{,}5625$$

$$S_b^2 = \frac{\sum_{i=1}^{k} n_i\left(\overline{X}_i - \overline{\overline{X}}\right)^2}{k-1} = \frac{1035{,}94}{3-1} = 517{,}97$$

$$S_t^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}\right)^2 - \bar{\bar{X}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}X_{ij}}{n-1} = \frac{12585,94}{15} = 839,06$$

Test istatistiği

$$F = \frac{S_b^2}{S_w^2} = 0,583 < F_{2,13,0,01} = 6,70$$

Sıfır hipotezinin reddi için yeterli kanıt yoktur. Bölümlere göre başarı notları ortalamalarında anlamlı bir farklılık yoktur.

ÖRNEK 3: Bir ihracat firması 3 ayrı ülkeye yaptığı ihracatlar arasında miktar açısından farklılık olup olmadığını incelemek istiyor. Bu amaçla üç ayrı ülkeye yapılan ihracat miktarları aşağıdaki gibi toparlanmıştır. Farklılık olup olmadığını 0,01 anlam düzeyinde inceleyiniz.

	A Ülkesi	B Ülkesi	C Ülkesi
	10	50	10
	12	51	15
	17	40	14
	18	30	17
	15		18
			16
$\sum X_{ij}$	72	171	90
$\bar{X}_{.j}$	14,4	42,75	15

Çözüm:

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : Ortalamaların en az biri diğerlerinden farklıdır

$$\bar{X}_{..} = \frac{333}{15} = 22,2$$

$KT_{GENEL} = (10\text{-}21,53)^2 + ...+ (16\text{-}21,53)^2 = 2680,4$

$KT_{G.ARASI} = 5(12,4\text{-}21,53)^2 + 4(42,75\text{-}21,53)^2 + 6(15\text{-}21,53)^2$

$= 2304,45$

$KT_{G.İÇİ} = KT_{GENEL} - KT_{G.ARASI} = 2680,4\text{-}2304,45 = 375,95$

Kaynaklar	sd.	KT	KO
Gruplar Arası	2	2304,45	
Grup İçi	12	375,95	1,305
Genel	14	2680,4	

$F_{0,01,2,12} = 6,93 > F_{hesap} = 1,305$ H_0 kabul

ÖRNEK 4: Üç ayrı eğitim sisteminin etkileri incelenmek için 5'er öğrenci bu sistemler ile çalışarak bir test sınavına alınmışlardır. Bu sınavdan aldıkları notlar aşağıdaki gibidir. Eğitim sistemleri arasında farklılık olup olmadığını 0,01 anlam düzeyinde inceleyiniz.

	A Sistem	B Sistemi	C Sistemi
	70	50	60
	75	65	70
	80	70	80
	90	40	85
	60	55	90
$\sum X_{ij}$	375	280	385
$\overline{X}_{.j}$	75	56	77

Çözüm:

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : Ortalamaların en az biri diğerlerinden farklıdır.

$\overline{\overline{X}} = 69,33$

$$KT_{G.ARASI} = r.\sum_{j=1}^{c}\left(\overline{X}_{.j} - \overline{\overline{X}}\right)^2$$

$$= 5.[(75-69,33)^2 + (56-69,33)^2 + (77-69,33)^2]$$

$$= 1343,3$$

$$KT_{GENEL} = \sum\sum\left(X_{ij} - \overline{X}_{.j}\right)^2$$

$$= (70-75)^2 + (75-75)^2 + ... + (90-77)^2$$

$$= 1650$$

$$KT_{G.İÇİ} = KT_{GENEL} - KT_{G.ARASI}$$

$$= 1650 - 1343,3 = 306,7$$

Kaynaklar	sd.	KT	KO
Gruplar Arası	3-1=2	1343,3	671,65
Grup İçi	14-2=12	306,7	25,55
Genel	15-1=14	1650	

$F_{hesap} = 29{,}66$

$F_{0,01,2,12} = 6{,}93 < F_{hesap} = 29{,}66$ H_0 red

En az bir grup diğerlerinden farklıdır.

BÖLÜM 13

BASİT REGRESYON VE KORELASYON

İstatistiksel araştırmalarda iki veya daha çok değişken arasındaki ilişkinin incelenmesi gerekebilir. Değişkenler arasındaki ilişkinin belirlenmesi için farklı yöntemlerden söz edilebilirse de, en çok kullanılan yöntemler olduklarından ilk olarak akla regresyon ve korelasyon analizi gelmektedir.

Değişkenler arasındaki ilişki matematiksel bir modelle açıklanabileceği gibi, ilişkinin derecesi ve yönü bir katsayı ile de ortaya konabilir. İlişkinin nasıl incelenece-ği konusundaki tercih araştırmanın özelliklerine ve araştırmacıya göre değişecektir.

Bu bölümde önce basit doğrusal regresyon, daha sonra doğrusal korelasyon açıklanacak ve bu konuda gerekli açıklamalar yapıldıktan sonra regresyon ve korelasyon arasındaki ilişki ortaya konacaktır.

13.1. BASİT DOĞRUSAL REGRESYON

Değişkenler arasındaki ilişkinin matematiksel bir model ile açıklanmasına regresyon analizi adı verilmektedir. Basit doğrusal regresyondaki basit kelimesi iki değişken arasındaki ilişkinin incelendiğini, doğrusal kelimesi ise değişkenler arasındaki ilişkiyi açıklamak için kullanılacak matematiksel modelin doğrusal olduğunu ifade etmektedir. İki değişken arasındaki ilişkiyi açıklamak için kullanılacak doğrusal modelin matematiksel ifadesi,

$$Y = a + bX$$

şeklinde olacaktır.

Basit doğrusal regresyonda, değişkenler arasındaki ilişki yukarıda verilen fonksiyona benzer bir fonksiyon ile açıklanırken modelin a ve b parametrelerinin belirlenmesi gerekecek ve bunun için değişkenlerin birlikte değişimini gösteren bir bileşik seriden yararlanılacaktır. Değişkenler arasındaki ilişkiyi açıklayan seri çok sayıda değerden oluşacak ve ilişkiyi belirlemek için serinin tümünün analizde kullanılması gerekecektir. Serinin birim sayısını N ile ifade edersek, N birimin oluşturduğu ana kütlenin bir doğrusal regresyon modeli olacaktır. Anakütle regresyon modeli,

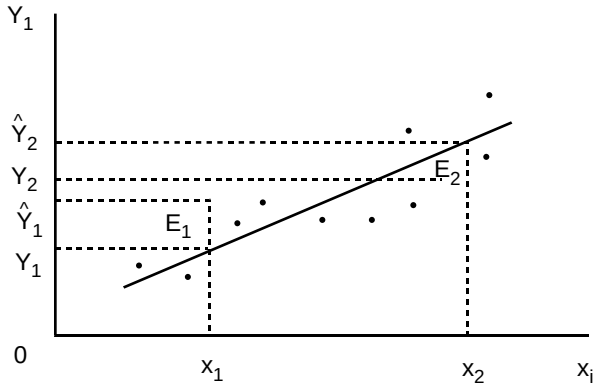
$$Y_i = \alpha + \beta X_i$$

şeklinde yazılırsa bir hata yapılmış olacaktır. Uygulamada karşılaşılabilecek tüm değişkenleri düşünürsek, hangi değişken çifti alınırsa alınsın aralarında tam doğrusal bir ilişki olduğu söylenemez. Değişkenler arasındaki ilişkiyi en iyi açıklayan model doğrusal model olsa bile bazı sapmalar söz konusu olacaktır. Bu nedenle anakütle regresyon modeline hata terimi olarak adlandırılan ve ε_i ile ifade edilen bir değişkenin daha eklenmesi ile modelin,

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

şeklinde ifade edilmesi gerekecektir. Şekil 13.1'de verilen dağılımın noktalarla gösterildiği serpilme diyagramı görülmektedir. Diyagramda yer alan verilerin dağılımına dikkat edilirse bu ilişkinin açıklanması için doğrusal modelin uygun olacağı anlaşılacaktır.

Bu modelin grafiği çizildiğinde elde edilecek doğruya regresyon doğrusu adı verilecektir.



Şekil 13.1

Regresyon doğrusunun bu diyagramda görülen noktalar arasından, değişkenler arasındaki ilişkiyi en iyi açıklayacak şekilde geçirilmesi gerekecektir. Regresyon doğrusu üzerinde yer alacak ve teorik değerler olarak adlandırılan değerler ile gerçek değerler arasındaki fark, hata yani gerçek değerlerden sapmadır. Şekilde gösterilen X_1 değerinin karşılığı olan Y_1 değeri gerçek değer ve regresyon doğrusu üzerinde yer

alan $\hat{Y}_1$ değeri ise teorik değerdir. Bu teorik değer gerçek Y_1 değerinden küçük olduğunda ε_1 hata terimi,

$$\varepsilon_1 = Y_1 - \hat{Y}_1 \qquad \text{ve} \qquad Y_1 > \hat{Y}_1$$

olacağından ε_1 pozitif olacaktır. Diğer taraftan X_2 değerinin karşılığı olarak regresyon doğrusu üzerinde yer alan $\hat{Y}_2$ teorik değeri ile gerçek Y_2 değeri için ε_2 hata terimi,

$$\varepsilon_2 = Y_2 - \hat{Y}_2 \qquad \text{ve} \qquad \hat{Y}_2 > Y_2$$

olduğundan ε_2 negatif olacaktır. Görüldüğü gibi hata terimleri pozitif veya negatif işaretli olabilir. Uygulamada verilerin çok sayıda olması; politikaların veya teknolojinin değişmesinin değişkenler arasındaki ilişkiyi etkilemesi gibi nedenlerle anakütle regresyon modelinin parametreleri anakütleden alınacak örnek ile tahmin edilecektir. Bu nedenle N birimli ana kütleden n birimli bir örnek alınacak. Anakütle doğrusal regresyon modelinin örnekten tahminini,

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$$

olarak gösterilecektir. $\hat{\alpha}$, α 'nın; $\hat{\beta}$, β 'nın tahminini ε_i ise anakütle hata terimi e_i'nin örnekten elde edilecek değerlerini göstermektedir.

Regresyon modelinde Y_i değişkeni etkilenen değişkendir ve bağımlı değişken olarak adlandırılır. Bağımsız değişken olarak adlandırılan X_i değişkeni ise etkileyen, sonuç yaratan değişkendir. Bu nedenle değişkenlerin hangisinin bağımlı hangisinin bağımsız değişken olarak alınacağı önemlidir. Örneğin, gelir, tüketim ilişkisinde gelir tüketimi etkilendiğinden, gelir bağımsız tüketim ise bağımlı değişken olarak ciro, kur ilişkisinde ise ciro karı etkileyeceğinden ciro bağımsız kar bağımlı değişken olarak alınacaktır.

$\hat{\alpha}$, $\hat{\beta}$ daha önce açıklandığı gibi anakütle regresyon modelinin parametrelerinin tahminleridir. $\hat{\alpha}$ 'in tahmin edilen değeri matematiksel olarak $X_i = 0$ olduğundan Y_i'nin alacağı değeri, yani regresyon doğrusunun Y eksenini kesim noktasını gösterecektir. $\hat{\alpha}$ regresyon analizinde bağımsız değişkenin sıfır değerini alması halinde bağımlı değişkenin alacağı değeri göstermektedir. $\hat{\alpha}$ sabit katsayı olarak da adlandırılır.

$\hat{\beta}$ 'in tahmin edilen sayısal değeri matematiksel olarak regresyon doğrusunun eğimini verir $\hat{\beta}$, regresyon analizinde bağımsız değişkenin değerinin 1 birim artması veya azalması karşısında bağımlı değişkenin değerinin ne kadar artacağını veya azalacağını göstermektedir. Bağımsız değişkenin katsayısı olarak da adlandırılan $\hat{\beta}$ değişkenler arasındaki ilişkinin yönünü de gösterir. $\hat{\beta}$ 'in işareti negatif ise değişken-

ler arasında ters yönlü $\hat{\beta}$'nin işareti pozitif ise değişkenler arasında doğru yönlü, ilişki olacaktır. Değişkenler arasında ilişki yoksa $\hat{\beta}$'in sıfır olması gerekecektir.

Regresyon analizi ile ilgili olarak bu bölümde verilecek formüller ile yapılacak tahminlerin geçerli olması, hata terimleri ile ilgili ve temel varsayımlar olarak adlandırılan varsayımların geçerliliğine bağlıdır. Bu varsayımları şöyle sıralayabiliriz.

- Normallik: Hata terimlerinin dağılımı normal dağılımdır.
- Sıfır Ortalama: Hata terimlerinin dağılımlarının ortalaması sıfırdır.
- Sabit Varyans: Hata terimlerinin her birinin dağılımının varyansı sabittir.
- Otokorelasyon olmaması: Hata terimleri arasındaki ilişkiye otokorelasyon denir. Bu varsayıma göre hata terimleri birbirini etkilemeyecektir.
- X_i değişkeninin tesadüfi değişken olmaması

Bu varsayımlardan bir veya birkaçının geçerli olmaması halinde sonuçlarda sapmalar meydana gelecektir. Varsayımlardan sapmalar olarak adlandırabileceğimiz; varsayımların geçerli olmaması durumu oldukça geniş bir konudur ve burada anlatılmayacaktır.

Örnekten anakütle parametrelerinin tahmininde kullanılacak formülleri elde etmek için istatistiksel tahmin teorisinden yararlanılır ve formüller çeşitli yöntemlerle elde edilebilir. Ancak bu yön-temlerden en çok kullanılanı en küçük kareler yöntemi olduğundan burada parametrelerin tahmini için kullanılacak formüllerin bu yöntem ile elde edilmesine yer verilecektir.

13.1.1. Parametrelerin Tahmini

En küçük kareler yönteminde, parametrelerin tahmininde kullanılacak formüller gerçek ve teorik değerler arasındaki farkların kareleri toplamı minimum olacak şekilde belirlenir. Regresyon analizinde gerçek değerler ile teorik değerler arasındaki farklar hata terimi ile ifade edildiğine göre α ve β hata terimlerinin kareleri toplamı minimum olacak şekilde tahmin edilecektir.

Hata terimlerinin kareleri toplamını T ile ifade edersek,

$$T = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

olacaktır. Teorik değerler

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \quad Y_i = \alpha + \beta \; X_i$$

denkleminde X_i değerleri yerine konarak elde edileceğinden,

$$T = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

$$= \sum_{i=1}^{n} \left[Y_i - (\alpha + \beta X_i)\right]^2$$

$$= \sum_{i=1}^{n} \left[Y_i - \alpha - \beta X_i\right]^2$$

yazılabilir. T'yi minimum yapacak α ve β değerlerini belirlemek için, T'nin α ve β 'ya göre kısmi türevleri alınıp sıfıra eşitlenecektir. Kısmi türevler;

$$\frac{\partial T}{\partial \alpha} = 2\sum_{i=1}^{n} \left(Y_i - \alpha - \beta X_i\right)(-1)$$

$$\frac{\partial T}{\partial \beta} = 2\sum_{i=1}^{n} \left(Y_i - \alpha - \beta X_i\right)(-X_i)$$

olacaktır.

Kısmi türevleri sıfıra eşitleyip, tahmin oldukları için α ve β 'yı, $\hat{\alpha}$ ve $\hat{\beta}$ ile göstererek gerekli düzenlemeleri yaparsak,

$$2\sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i\right)(-1) = 0$$

$$2\sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} X_i\right)(-X_i) = 0$$

$$\sum_{i=1}^{n} Y_i - n\hat{\alpha} - \hat{\beta}\sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \hat{\alpha}\sum_{i=1}^{n} X_i - \hat{\beta}\sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

denklemleri elde edilir ve bu denklemler normal denklemler olarak adlandırılmaktadır. α ve β 'nın tahmini için bu iki bilinmeyenli denklem sisteminin çözülmesi

gerekmektedir. T'nin ikinci türevleri alınırsa tahmin edilecek değerlerin minimum olacağı görülecektir.

ÖRNEK: Aşağıda verilen serinin basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.

Y_i	X_i	X_iY_i	X_i^2
2	6	12	36
3	7	21	49
5	9	45	81
8	13	104	169
12	15	180	225
30	50	362	560

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i$$

$$\sum X_iY_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2$$

$$-10/30 = 5\hat{\alpha} + 50\hat{\beta}$$

$$362 = 50\hat{\alpha} + 560\hat{\beta}$$

$$-300 = -50\hat{\alpha} - 500\hat{\beta}$$

$$362 = 50\hat{\alpha} + 560\hat{\beta}$$

$$62 = 0 + 60\hat{\beta}$$

$$\hat{\beta} = \frac{62}{60} = 1{,}03$$

$$30 = 5\hat{\alpha} + 50(1{,}03)$$

$$\hat{\alpha} = \frac{30 - 51{,}5}{5} = -4{,}3$$

İki bilinmeyenli denklemlerin çözümü yerine, birinci denklemden $\hat{\alpha}$ çekilir ve ikinci denklemde yerine konursa,

$$\sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i$$

$$\hat{\alpha} = \frac{\sum Y_i - \hat{\beta}\sum X_i}{n} = \overline{Y} - \hat{\beta}\overline{X}$$

$$\sum X_iY_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_iY_i = \left(\overline{Y} - \hat{\beta}\overline{X}\right)\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_i Y_i = \bar{Y}\sum X_i + \hat{\beta}\bar{X}\sum X_i + \hat{\beta}\sum X_i^2$$

$$\sum X_i Y_i - \bar{Y}\sum X_i = \hat{\beta}(\sum X_i^2 - \bar{X}\sum X_i)$$

$$\hat{\beta} = \frac{\sum X_i Y_i - \bar{Y}\sum X_i}{\sum X_i^2 - \bar{X}\sum X_i}$$

pay ve paydayı $\frac{n}{n}$ ile çarparak,

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\left(\bar{X}\right)^2}$$

elde edilir. Bu durum da $\hat{\alpha}$,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

formülü ile hesaplanabilir. Bu formülleri katsayıların gerçek değerle tahmin formülleri olarak adlandırabiliriz.

ÖRNEK: Normal denklemler ile çözdüğünüz örneği, katsayıların gerçek değerlerden tahmin formülleri ile çözünüz.

$$\sum Y_i = 30 \qquad \sum X_i Y_i = 362$$

$$\sum X_i = 50 \qquad \sum X_i^2 = 560$$

olduğundan ortalamalar,

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{50}{5} = 10$$

olacaktır. Değerler formüllerde yerine konursa,

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\left(\bar{X}\right)^2}$$

$$= \frac{362 - 5(10)(6)}{560 - 5(10)^2} = \frac{62}{60} = 1,03 \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 6 - 1,03(10) = -4,3$$

olarak aynı sonuçlar elde edilmiş olur.

Değişkenler için ortalamadan farklar serileri düzenlenerek katsayılar tahmin edilirse işlemler kolaylaşacaktır. Farklar serileri oluşturulduğunda,

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$$

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) = 0$$

olacağını aritmetik ortalamasının özelliklerinden biliyoruz. Normal denklemlerde X_i yerine (X_i - $\overline{X}$) ve Y_i yerine (Y_i - $\overline{Y}$) koyarsak,

$$\begin{array}{r} \sum Y_i = n\hat{\alpha} + \hat{\beta}\sum X_i \\ \sum X_i Y_i = \hat{\alpha}\sum X_i + \hat{\beta}\sum X_i^2 \\ \hline \sum (Y - \overline{Y}_i) = n\hat{\alpha} + \hat{\beta}\sum (X_i - \overline{X}) \\ \sum (X_i - \overline{X})(Y_i - \overline{Y}) = \hat{\alpha}\sum (X_i - \overline{X}) + \hat{\beta}\sum (X_i - \overline{X})^2 \end{array}$$

Ortalamalardan farkların toplamları sıfır olacağından,

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \hat{\beta}\sum (X_i - \overline{X})^2$$

$$\hat{\beta} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

elde edilir. Bu durumda $\hat{\alpha}$,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

şeklinde tahmin edilebilir. Bu formüller katsayıların ortalamadan farklar ile tahmin formülleri olarak adlandırılabilirler.

ÖRNEK: Daha önce çözülen örneği ortalamadan farklar formülleri ile çözünüz.

Y_i	X_i	$Y_i - \overline{Y}$	$X_i - \overline{X}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(X_i - \overline{X})^2$
2	6	-4	-4	16	16
3	7	-3	-3	9	9
5	9	-1	-1	1	1
8	13	2	3	6	9
12	15	6	5	30	25
30	50	0	0	62	60

$\overline{X} = 10$

$\overline{Y} = 6$

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{62}{60} = 1{,}03$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 6 - 1{,}03\ (10) = -\ 4{,}3$$

olarak aynı sonuçlar bulunur.

Kolaylık olması için $x_i = X_i - \bar{X}$ ve $y_i = Y_i - \bar{Y}$ olarak tanımlarsak,

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

olacaktır.

Gerçek değerlerin toplamları, kareleri ve çarpımlarının toplamları bilindiğinde ortalamadan farklar ile tahmin yapılmak isteniyorsa, ortalamadan farklar serilerinin toplamları kareleri ve çarpımlarının toplamları gerçek değerlerden elde edilebilir.

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i^2 - 2\bar{X}X_i + \bar{X}^2)$$

$$= \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \bar{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2$$

$-2\bar{X}\sum_{i=1}^{n} X_i$ ifadesini $\frac{n}{n}$ ile çarparsak,

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - 2n\bar{X}\frac{\sum X_i}{n} + n\bar{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

olacaktır. Aynı şekilde,

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

olarak elde edilebilir.

$$\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})=\sum_{i=1}^{n}(X_iY_i-Y_i\bar{X}-\bar{Y}X_i+\bar{X}\bar{Y})$$

$$=\sum_{i=1}^{n}X_iY_i-\bar{X}\sum_{i=1}^{n}Y_i-\bar{Y}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}\bar{Y}$$

$-\bar{X}\sum_{i=1}^{n}Y_i$ ve $-\bar{Y}\sum_{i=1}^{n}X_i$ ifadelerini $\frac{n}{n}$ ile çarparsak,

$$\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})=\sum_{i=1}^{n}X_iY_i-n\bar{X}\frac{\sum_{i=1}^{n}Y_i}{n}-n\bar{Y}\frac{\sum_{i=1}^{n}X_i}{n}+n\bar{X}\bar{Y})$$

$$=\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}$$

olarak bulunur.

ÖRNEK: Daha önceki verilen örnekte n =5, $\sum Y_i$ = 30, $\sum X_i$ =50, $\sum X_iY_i$ =362, $\sum X_i^2$ = 560 olarak bilindiğine göre $\sum(X_i-\bar{X})^2$ ve $\sum(X_i-\bar{X})(Y_i-\bar{Y})$ değerlerini hesaplayınız.

$$\bar{X}=\frac{\sum X}{n}=\frac{50}{5}=10$$

$$\bar{Y}=\frac{\sum Y}{n}=\frac{30}{5}=6$$

$$\sum(X_i-\bar{X})^2=\sum X^2-n\bar{X}^2$$

$$= 560\text{-}5(10)^2 = 60$$

$$\sum(X_i-\bar{X})(Y_i-\bar{Y})=\sum X_iY_i-n\bar{X}\bar{Y}$$

$$= 362 - 5(10)(6) = 62$$

bulunur. Bu sonuçlar daha önce farklar serilerinden elde edilen sonuçlar ile aynıdır.

13.1.2. Teorik Değerlerin Hesaplanması

Bağımsız değişkenin herhangi bir değeri için regresyon modelinden hesaplanacak bağımlı değişkenin değerinin teorik değer olarak adlandırıldığından daha önce söz etmiştik. Diğer bir ifade ile X_i değerinin regresyon modelinde yerine konması ile hesaplanan ve $\hat{Y}_i$ gösterilen değerler teorik değerlerdir. X_i değerleri X_1, X_2,X_3 ,....., X_n ise, teorik değerler,

$$\hat{Y}_1 = \hat{\alpha} + \hat{\beta}(X_1)$$

$$\hat{Y}_2 = \hat{\alpha} + \hat{\beta}(X_2)$$

$$\hat{Y}_3 = \hat{\alpha} + \hat{\beta}(X_3)$$

$$\vdots \quad \vdots \quad \vdots$$

$$\hat{Y}_n = \hat{\alpha} + \hat{\beta}(X_n)$$

olacaktır. Bu değerler regresyon doğrusu üzerinde gerçek Y_i değerlerinin karşılığı olan değerlerdir.

$$e_i = Y_i - \hat{Y}_i$$

olduğu ve regresyonun temel varsayımlarından $\bar{e}$ =0 olacağı bilinmektedir.

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n}$$

olduğundan, ancak

$$\sum_{i=1}^{n} e_i = 0$$

ise $\bar{e} = 0$ olacaktır. Bu nedenle,

$$\frac{\sum_{i=1}^{n} e_i}{n} = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)}{n} = 0$$

olarak yazılabilir. Buradan,

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0$$

ve

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{Y}_i = 0$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

bulunur. Bu sonuçlara göre, gerçek Y_i değerleri ile teorik $\hat{Y}_i$ değerlerinin farklarının toplamı sıfır olacaktır. Ayrıca, gerçek Y_i değerleri toplamı ile teorik $\hat{Y}_i$ değerleri toplamları birbirine eşittir.

ÖRNEK: Daha önce verilen örnek için teorik değerleri hesaplayarak,

$$\sum Y_i = \sum \hat{Y}_i \quad \text{ve} \quad \sum (Y_i - \hat{Y}_i) = 0$$

olduğunu gösteriniz.

Regresyon modeli $Y_i = -4{,}3 + 1{,}03X_i$ olarak belirlenmişti.

Y_i	X_i	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$
2	6	1,88	0,12
3	7	2,91	0,09
5	9	4,97	0,03
8	13	9,09	-1,09
12	15	11,15	0,85
30	50	30,00	0

$\hat{Y}_1 = -4{,}3 + 1{,}03\ (6) = 1{,}88$

$\hat{Y}_2 = -4{,}3 + 1{,}03\ (7) = 2{,}91$

$\hat{Y}_3 = -4{,}3 + 1{,}03\ (9) = 4{,}97$

$\hat{Y}_4 = -4{,}3 + 1{,}03\ (13) = 9{,}09$

$\hat{Y}_5 = -4{,}3 + 1{,}03\ (15) = 11{,}15$

13.1.3. Parametre Tahmincilerinin Varyansları

Regresyon modelinin parametrelerinin varyanslarını hesaplayabilmek için ana kütle hata tahminlerinin varyansının hesaplanması gerekecektir. Anakütle hata terimleri ε_i'lerin varyansı,

$$V(\varepsilon_i) = \sigma^2_{\varepsilon_i} = \frac{\sum_{i=1}^{N} (\varepsilon_i - \bar{\varepsilon})^2}{N}$$

olarak hesaplanacaktır. Regresyonun temel varsayımlarından $\bar{e} = 0$ olduğunu bilindiğinden

$$\sigma^2_{\varepsilon_i} = \frac{\sum_{i=1}^{N} (\varepsilon_i)^2}{N}$$

olacaktır. Görüldüğü gibi anakütle hata terimlerinin varyanslarını hesaplayabilmek için anakütle hata terimlerinin bilinmesi gerekmektedir. Ana kütle hata terimleri bilinmeyeceğinden, hata terimlerinin varyansının örnekten tahmin edilmesi gereke-

cektir. Ana kütle hata terimlerinin varyansının örnekten elde edilen tahmini S_e^2 ile gösterilirse, ana kütle hata terimlerinin varyansı örnekten,

$$S_e^2 = \frac{\sum_{i=1}^{n} (\varepsilon_i)^2}{n-2}$$

formülü ile tahmin edilir. Formülde, $\sum e_i^2$ 'nin n yerine (n-2)'ye bölünme sebebi; örnek hata terimleri varyansının, anakütle hata terimleri varyansının sapmalı tahmincisi olmasıdır. Verilen formülle hesaplanan S_e^2 ana kütle hata terimleri varyansının sapmasız tahmincisi olmaktadır.

S_e^2 'nin hesaplanması için hata terimleri kareleri toplamının hesaplanması gerekecektir. Daha önce teorik değerlerin hesaplanmasında açıklandığı gibi, hata terimleri hesaplanarak bunlar yardımı ile hata terimlerinin kareleri toplamı bulunabilir. Hata terimlerinin kareleri toplamının aşağıda verilen formüller ile daha kısa yoldan hesaplanması da mümkündür.

$$Y_i = \hat{\alpha} + \hat{\beta} X_i$$

olduğundan,

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum [Y_i - (\hat{\alpha} + \hat{\beta} X)]^2$$
$$= \sum [Y_i - \hat{\alpha} - \hat{\beta} X_i]^2$$

yazılabilir.

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

olduğundan,

$$\sum e_i^2 = \sum [Y_i - \hat{\alpha} - \hat{\beta} X_i]^2$$
$$= \sum [Y_i - (\overline{Y} - \hat{\beta}\overline{X}) - \hat{\beta} X_i]^2$$

olacaktır. Gerekli düzenlemeleri yaparsak,

$$\sum e_i^2 = \sum [Y_i - \overline{Y} + \hat{\beta}\overline{X} - \hat{\beta} X_i]^2$$
$$= \sum [(Y_i - \overline{Y}) - \hat{\beta}(X_i - \overline{X})]^2$$
$$= \sum [(Y_i - \overline{Y})^2 + \hat{\beta}^2 (X_i - \overline{X})^2 - 2\hat{\beta}(Y_i - \overline{Y})(X_i - \overline{X})]$$
$$= \sum (Y_i - \overline{Y})^2 + \hat{\beta}^2 \sum (X_i - \overline{X})^2 - 2\hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

olarak bulunur.

$$\hat{\beta} = \frac{\sum (Y_i - \overline{Y})(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$$

olduğundan,

$$\hat{\beta} \sum (X_i - \overline{X})^2 = \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

olacaktır. Bu eşitlikten yararlanarak,

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 + \hat{\beta}\hat{\beta} \sum (X_i - \overline{X})^2 - 2\hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum (Y_i - \overline{Y})^2 + \hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X}) - 2\hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum (Y_i - \overline{Y})^2 - \hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

elde edilir. Ortalamalardan farkları,

$x_i = X_i - \overline{X}$

$y_i = Y_i - \overline{Y}$

ile ifade edersek,

$$\sum e_i^2 = \sum y_i^{\,2} - \hat{\beta} \sum x_i y_i$$

olacaktır.

Hata terimlerinin kareleri toplamı ortalamalardan farklar ile hesaplanabildiği gibi, gerçek değerler ile de hesaplanabilir.

$$\sum (Y_i - \overline{Y})^2 = \sum Y^2 - n\overline{Y}^2$$

$$\sum (Y_i - \overline{Y})(X_i - \overline{X}) = \sum X_i Y_i - n\overline{X}\overline{Y}$$

olduğundan bu eşitliklerle,

$$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta} \sum (Y_i - \overline{Y})(X_i - \overline{X})$$

$$= \sum Y_i^2 - n\overline{Y}^2 - \hat{\beta}(\sum X_i Y_i - n\overline{X}\overline{Y})$$

olarak yazılabilir.

$$\overline{Y} = \hat{\alpha} + \hat{\beta}\overline{X}$$

olduğundan,

$$\sum e_i^2 = \sum Y_i^2 - n\overline{Y}\overline{Y} - \hat{\beta} \sum X_i Y_i + n\hat{\beta}\overline{X}\overline{Y}$$

$$= \sum Y_i^2 - n\overline{Y}(\hat{\alpha} + \hat{\beta}\overline{X}) - \hat{\beta} \sum X_i Y_i + n\hat{\beta}\overline{X}\overline{Y}$$

$$= \sum Y_i^2 - n\hat{\alpha}\overline{Y} - n\hat{\beta}\overline{X}\overline{Y} - \hat{\beta}\sum X_i Y_i + n\hat{\beta}\overline{X}\overline{Y}$$

$$= \sum Y_i^2 - n\hat{\alpha}\overline{Y} - \hat{\beta}\sum X_i Y_i$$

olarak bulunur.

$$\overline{Y} = \frac{\sum Y_i}{n}$$

olduğundan ifade,

$$\sum e_i^2 = \sum Y_i^2 - n\hat{\alpha}\frac{\sum Y_i}{n} - \hat{\beta}\sum X_i Y_i$$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

şeklini alacaktır.

ÖRNEK: Daha önce verilen örneğin hata terimlerinin varyansını hesaplayınız.

Hata terimleri daha önce hesaplandığından

e_i	e_i^2
0,12	0,0144
0,09	0,0081
0,03	0,0009
-1,09	1,1881
0,85	0,7225
	1,9340

$\sum e_i^2$ =1,9340 olarak bulunur.

Hata terimlerinin kareleri toplamı gerçek değerler ile hesaplandığında,

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$\sum Y_i^2 = (2)^2 + (3)^2 + (5)^2 + (8)^2 + (12)^2 = 246$$

$$\sum Y_i = 30$$

$$\sum X_i Y_i = 362$$

$$\hat{\alpha} = -4{,}3$$

$$\hat{\beta} = 1{,}03$$

olduğundan,

$$\sum e_i^2 = 246 - (-4{,}3)(30) - 1{,}03(362)$$

$$= 2{,}14$$

olarak bulunur.

Hata terimlerinin kareleri toplamı ortalamalardan farklar ile hesaplandığında,

$$\sum e_i^2 = \sum (Y_i - \bar{Y})^2 - \hat{\beta} \sum (Y_i - \bar{Y})(X_i - \bar{X})$$

Y = 6 olarak bilindiğinden,

$$\sum (Y_i - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$$

$$= 246 - 5(6)^2 = 66$$

$$\sum (Y_i - \bar{Y})(X_i - \bar{X}) = 62$$

$$\sum e_i^2 = 66 - 1{,}03(62) = 2{,}14$$

olarak bulunur.

Görüldüğü gibi hata terimlerinin kareleri toplamı hata terimlerinin teker teker hesaplanarak karelerinin alınması halinde 1,934, diğer şekillerde hesaplandığında ise 2,14 olarak bulunmuştur. Aradaki çok küçük ve önemsiz fark, bağımsız değişkenin katsayısının değerinin iki ondalıklı alınmasından kaynaklanmaktadır.

$\sum e_i^2 = 1{,}934$ olduğundan,

$$S_e^2 = \frac{\sum e_i^2}{n-2} = \frac{1{,}934}{5-2} = 0{,}6446$$

$$S_e = \sqrt{0{,}6446} = 0{,}8029$$

olacaktır.

Regresyon modelinin sabit ve bağımsız değişkenin parametrelerinin varyansları,

$$V(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2 = \sigma_{\varepsilon_i}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$$

$$V(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma_{\varepsilon_i}^2}{\sum (X_i - \bar{X})^2}$$

formülleri ile hesaplanacaktır. $\sigma_{\varepsilon_i}^2$ bilinmeyip örnekten S_e^2 olarak tahmin edileceğinden, parametrelerin varyanslarını tahmin etmek için,

$$S_{\hat{\alpha}}^2 = S_e^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right]$$

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum(X_i - \overline{X})^2}$$

formülleri kullanılacaktır.

ÖRNEK: Daha önce verilen örnek için parametrelerin varyanslarını hesaplayınız.

$$\overline{X} = 10$$
$$S_e^2 = 0{,}6446$$
$$\sum(X_i - \overline{X})^2 = 60$$

olduğundan,

$$S_{\hat{\alpha}}^2 = S_e^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right]$$
$$= 0{,}6446\left[\frac{1}{5} + \frac{(10)^2}{60}\right]$$
$$= 1{,}2032$$
$$S_{\hat{\alpha}} = \sqrt{1{,}2032} = 1{,}0969$$

ve

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum(X_i - \overline{X})^2}$$
$$= \frac{0{,}6446}{60}$$
$$= 0{,}0107$$
$$S_{\hat{\beta}} = \sqrt{0{,}0107} = 0{,}1034$$

olarak bulunur.

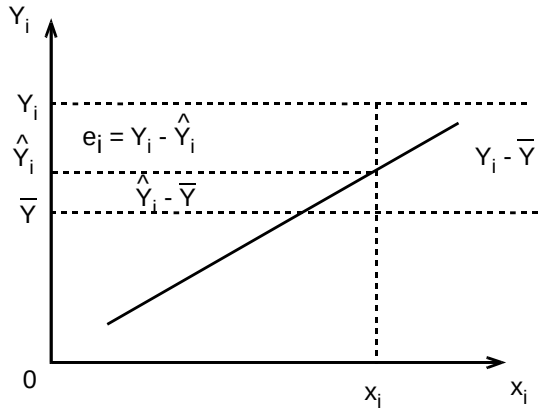
13.1.4. Belirlilik Katsayısı

Bağımlı değişkendeki değişmelerin bağımsız değişken veya bağımsız değişkenler tarafından açıklanma oranını gösteren katsayıya belirlilik katsayısı denmekte-

dir. Basit doğrusal regresyonda tek bağımsız değişken olduğundan, bağımlı değişkendeki değişmeler bu bağımsız değişken tarafından açıklanacaktır. Bağımlı değişkende meydana gelen toplam değişme aritmetik ortalamadan farkların kareleri toplamına eşittir.

$$\text{Toplam Değişme} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} y_i^2$$

Şekil 13.2'de X_i ve Y_i gerçek değerleri ile $\hat{Y}_i$ teorik değeri görülmektedir.



Şekil 13.2

Şekilden de görüldüğü gibi, bağımlı değişkendeki değişme teorik değerin ortalamadan farkı ile hata terimi toplamına eşittir.

$$(Y_i - \overline{Y}) = (\hat{Y}_i - \overline{Y}) + e_i$$

tarafların karelerinin toplamları ise,

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 + \sum e_i^2$$

olarak elde edilir. Burada,

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \text{ Toplam Değişmeyi}$$

$$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = \text{ Açıklanan Değişmeyi}$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \text{Açıklanmayan Değişmeyi}$$

göstermektedir.

Belirlilik katsayısı açıklanan değişmenin toplam değişmeye oranı olarak tanımlandığına göre, R^2 ile ifade edilen belirlilik katsayısı,

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

olacaktır. (1- R^2) ise açıklanmayan değişmenin toplam değişmeye oranını verecektir.

Bağımsız değişken bağımlı değişkendeki değişmeleri hiç açıklamıyorsa açıklanan değişme sıfıra eşit, bağımsız değişken bağımlı değişkendeki değişmelerin tümünü açıklıyorsa açıklanan değişme toplam değişmeye eşit olacağından,

$$0 \le \sum(\hat{Y}_i - \bar{Y})^2 \le \sum(Y_i - \bar{Y})^2$$

olacaktır. Taraflar toplam değişmeye bölünürse,

$$\frac{0}{\sum(Y_i - \bar{Y})^2} \le \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \le \frac{\sum(Y_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$0 \le R^2 \le 1$$

olarak bulunur. R^2, 0 ile 1 arasında değer alır. R^2'nin değeri 1'e yaklaştıkça bağımsız değişkenin bağımlı değişkendeki değişmeleri açıklama oranı artar, 0'a yaklaştıkça bağımsız değişkenin bağımlı değişkendeki değişmeleri açıklama oranı azalır.

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum e_i^2$$

olduğundan,

$$\sum(\hat{Y}_i - \bar{Y})^2 = \sum(Y_i - \bar{Y})^2 - \sum e_i^2$$

olacaktır. Bu eşitliği R^2 formülünde yerine koyarsak,

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

$$= \frac{\sum(Y_i - \bar{Y})^2 - \sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

$$= 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

olarak bulunur. Yani, belirlilik katsayısı açıklanmayan değişmenin toplam değişmeye oranının 1'den çıkarılmasına eşittir. Aynı bağımlı değişken için farklı bağımsız değişkenler ile oluşturulacak regresyon modellerinden belirlilik katsayısı daha büyük olan model ilişkiyi daha iyi açıklayacaktır.

ÖRNEK: Daha önce verilen örneğin belirlilik katsayısını hesaplayarak, bulduğunuz sonucu yorumlayınız.

Daha önce çözdüğümüz örnekte;

$$\sum e_i^2 = 1{,}934$$

ve

$$\sum (Y_i - \overline{Y})^2 = 66$$

olarak bulunmuştu.

Toplam değişme:

$$\sum (Y_i - \overline{Y})^2 = 66$$

Açıklanmayan değişme:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 1{,}934$$

olduğundan,

Açıklanan değişme:

$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = 66 - 1{,}934 = 64{,}066$$

olacaktır.

Açıklanan değişme yukarıda verilen formül ile hesaplanabildiği gibi seriden de hesaplanabilir.

$\overline{Y}$ = 6 olarak bilinmektedir.

$\hat{Y}_i$	$\hat{Y}_i - \overline{Y}$	$(\hat{Y}_i - \overline{Y})^2$
1,88	-4,12	16,9744
2,91	-3,09	9,5481
4,97	-1,03	1,0609
9,09	3,09	9,5481
11,15	5,15	26,5225
	0	63,654

İki hesaplama şekli arasındaki fark katsayıların değerlerinin tam alınmamasından kaynaklanmaktadır.

$$R^2 = \frac{64{,}066}{66} = 0{,}97$$

veya

$$R^2 = 1 - \frac{1{,}934}{66} = 0{,}97$$

olarak bulunur. Bu sonuca göre, bağımsız değişken bağımlı değişkendeki değişmelerin % 97'sini açıklamaktadır. 1- 0,97 = 0,03 olduğundan bağımsız değişken bağımlı değişkendeki değişmelerin % 3'ünü açıklayamamakta ve bu değişme diğer değişkenler tarafından açıklanmaktadır.

13.1.5. Katsayıların Testi

Regresyon modelinin tahmin edilen katsayılarının istatistiksel olarak anlamlı olup olmadığı test edilebilir. Bu amaçla t ve F testi olmak üzere iki farklı test uygulanabilir. F testinin basit regresyonda uygulanması anlamlı olmasa da burada kısaca açıklanacaktır.

13.1.5.1. t -Testi

Katsayıların ayrı ayrı istatistiksel anlamlılığı t testi ile incelenir. Basit regresyonda sabit ve eğim katsayıları olduğuna göre bunlara ayrı ayrı test uygulanır. Katsayıların anlamlılığı sözkonusu olduğundan, sıfırdan farklı olup olmadıklarına belirlenen hata payı ile karar verilecektir.

Katsayılar test edilirken hipotez testleri bölümünde açıklandığı gibi önce hipotezler oluşturulur. Sabit katsayı test edilirken temel hipotez,

$$H_0 : \alpha = 0$$

karşıt hipotez, test çift taraflı için,

$$H_1 : \alpha \neq 0$$

tek taraflı test için ise,

$$H_1 : \alpha > 0$$

veya

$$H_1 : \alpha < 0$$

şeklinde oluşturulacaktır.

Hipotezler oluşturulduktan sonra karar aşamasında kullanılacak kritik değer veya tablo değeri belirlenir. Kritik değerin belirlenmesinde testin tek veya çift taraflı olması yanında örnek birim sayısı da önemlidir. Örnek birim sayısı 30'dan büyük ise örnekleme dağılımının normal dağılım olduğu kabul edilir ve çift taraflı test için $Z_{\alpha/2}$, tek taraflı test için Z_{α} değeri normal dağılım tablosundan belirlenen α hata payı ile bulunur.

Örnek birim sayısı 30'dan küçük ise örnekleme dağılımı t dağılımıdır. Basit regresyonda t dağılımının serbestlik derecesi (n-2)'dir. Belirlenen α hata payı ile t dağılımı tablosundan çift taraflı test için $t_{\alpha/2}$, tek taraflı test için t_{α} değeri bulunacaktır.

Hipotez testlerinde üçüncü adım test istatistiğinin hesaplanmasıdır. Sabit katsayı için test istatistiği,

$$t = \frac{\hat{\alpha} - \alpha}{S_{\hat{\alpha}}}$$

dır. Sabit katsayının sıfıra eşitliği test edildiğinden $\alpha = 0$ alınır ve test istatistiği,

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}}$$

olarak hesaplanır.

Son olarak hipotezlerden birinin kabul kararı hipotez testlerinde açıklandığı gibi verilir. Kısaca tablodan bulunacak değer Z_t ile ifade edilirse, hem tek hem çift taraflı testler ve hem t hem de normal dağılım için $|Z| < |Z_t|$ ise H_0 hipotezi, $|Z| > |Z_t|$ ise H_1 hipotezi kabul edilir.

Sabit katsayının anlamlılığının testinde sıfır hipotezi kabul edilirse ($\alpha = 0$) regresyon modelinin orjinden geçmesi sözkonusudur. Bu sonuç modelin değişkenler arasındaki ilişkiyi iyi açıklamadığı sonucunu doğurur.

Eğim katsayısının testinde hipotezler aynı şekilde oluşturulur. Yani, hipotezlerdeki α 'nın yerini eğim katsayısı β alınır.

Temel Hipotez

$H_0 : \beta = 0$

Karşıt Hipotez

$H_1 : \beta \neq 0$ Çift Taraflı

$H_1 : \beta > 0$
$H_1 : \beta < 0$ Tek Taraflı

Eğim katsayısının testinde tablo değerleri sabit katsayının testi gibi belirlenir. Eğim katsayısı için test istatistiği,

$$t = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}}$$

dir ve $\beta = 0$ olup olmadığı test edildiğinden,

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

olacaktır. Test istatistiği hesaplandıktan sonra hangi hipotezin kabul edileceği kararı aynı şekilde verilir. Eğim katsayısının testinde H_0 hipotezinin kabulü X ve Y değişkenleri arasında belirlenen hata payı ile ilişki olmadığını ifade etmektedir.

ÖRNEK: 14 devrelik veri ile basit doğrusal regresyon modeli,

$$Y_i = 3{,}411 - 0{,}473 X_i$$
$$(1{,}219) \quad (0{,}135)$$

olarak tahmin edilmiştir.

a) 0,05 hata payı ile sabit katsayının anlamlı olup olmadığını test ediniz.

b) 0,01 hata payı ile eğim katsayısının tek taraflı test ediniz.

a) n = 14

$\hat{\alpha}$ = 3,411

$S_{\hat{\alpha}}$ = 1,219

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

$t_{\alpha/2,n-2} = t_{0,05/(14-2=12)} = 2{,}179$

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{3{,}411}{1{,}219} = 2{,}798$$

$|t| > |t_{\alpha/2,n-2}|$

2,798 >2,179 olduğundan H_1 hipotezi kabul, H_0 hipotezi reddedilir.

b) n = 14

$\hat{\beta}$ = -0,473

$S_{\hat{\beta}}$ = 0,135

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$t_{\alpha,n-2} = t_{0,01/12} = 3{,}055$

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{-0{,}473}{0{,}135} = -3{,}503$$

$|t| > |t_{\alpha,n-2}|$

3,503 > 3,055 olduğundan H_1 hipotezi kabul, H_0 hipotezi reddedilir.

13.1.5.2. F Testi

Çoklu regresyon için anlamlı olsa da basit regresyon modelleri için de F testi yapılabilir. F testi bağımsız değişkenlerin tümünün katsayılarının istatistiksel olarak

anlamlı olup olmadığını test eder. Basit regresyonda tek bağımsız değişken olduğundan basit regresyonda F testinin t testinden farklı bir anlamı yoktur.

F testi basit regresyonun bağımsız değişkeninin anlamlı olup olmadığını test edecektir. Bu nedenle temel hipotez,

$$H_0 : \beta = 0$$

olarak oluşturulacaktır. F testinde tek taraflı test yapılamaz. Karşıt hipotez H_0 hipotezinin doğru olmadığını belirteceğinden,

$$H_1 : \beta \neq 0$$

olarak oluşturulur.

F test istatistiği regresyon tarafından açıklanan değişkenin, açıklanmayan değişmeye oranına eşittir. Test istatistiği,

$$F = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 / 1}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i) / (n-2)}$$

dir. Test istatistiğinin dağılımı F dağılımıdır. Serbestlik dereceleri (1) ve (n-2)'dir. F test istatistiği F tablosundan serbestlik dereceleri SD_1=1 ve SD_2 = n-2 ve α hata payı ile bulunan tablo değeri ile karşılaştırılır.

$F < F_{\alpha,1,n-2}$ ise H_0 hipotezi,

$F > F_{\alpha,1,n-2}$ ise H_1 hipotezi

kabul edilir.

F test istatistiği belirlilik katsayısı R^2 ile de hesaplanabilir.

$$F = \frac{R^2}{1-R^2} . \frac{n-2}{1} = \frac{(n-2)R^2}{1-R^2}$$

Basit regresyonda eğim katsayısı için hesaplanan t test istatistiğinin karesi F test istatistiğine eşittir.

$$t_\beta^2 = F$$

Bu eşitlik sadece basit regresyon için geçerlidir.

ÖRNEK: 10 devrelik veri ile regresyon modeli

$$Y_i = 2{,}271 + 0{,}371 X_i$$

$$\sum_{i=1}^{n} (\hat{Y} - \bar{Y})^2 = 20{,}6$$

$$\sum_{i=1}^{n}(Y-\hat{Y}_i)^2 = 8{,}71$$

olarak tahmin edilmiştir. 0,05 hata payı ile modeli F testi ile test ediniz.

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$SD_1 = 1$

$SD_2 = n\text{-}2 = 10\text{-}2 = 8$

$F_{\alpha,1,n-2} = F_{0,05,1,8} = 5{,}32$

$$F = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y-\hat{Y})/(n-2)} = \frac{20{,}6}{8{,}71/8} = 18{,}92$$

$18{,}92 > 5{,}32$ H_1 hipotezi kabul edilir.

Bu örnekte toplam değişme,

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$$

$$= 20{,}6 + 8{,}71 = 29{,}31$$

olacaktır. Bu sonuçla,

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{20{,}6}{29{,}31} = 0{,}7028$$

bulunur ve

$$F = \frac{(n-2)R^2}{1-R^2} = \frac{(8)0{,}7028}{1-0{,}7028} = 18{,}92$$

olarak elde edilir. Bu örnekte β için hesaplanacak t test istatistiği

$$t_\beta^2 = F = 18{,}92$$

$$t_\beta = 4{,}3497$$

olacaktır. $\beta = 0{,}371$ olduğundan,

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}}$$

$$4{,}3497 = \frac{0{,}371}{S_{\hat{\beta}}}$$

$$S_{\hat{\beta}} = 0{,}08529$$

bulunacaktır.

13.1.6. Regresyon Modelinin Tahmin İçin Kullanılması

Regresyon modeli ile bağımsız değişkenin alacağı farklı değerler için bağımlı değişkenin değerleri tahmin edilebilir. Örneğin, bir işletme yöneticisi ürettikleri malın hangi fiyattan kaç adet satılacağını veya geçmiş yılların ciroları ile karları arasındaki ilişkiden yararlanarak gelecek yılın cirosunu belirleyerek karının ne olacağını tahmin etmek isteyebilir.

Regresyon modeli tahmin için kullanılacaksa, bağımsız değişkenin tahmin edilecek değerinin modelde yerine konması gerekecektir. Bağımsız değişkenin X_o değeri için, bağımlı değişkenin tahmini değeri,

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$$

olarak elde edilecektir.

Anakütle Y_o değeri ile örnekten tahmin edilen $\hat{Y}_0$ değeri arasındaki fark tahmin hatası olarak adlandırılır. Tahmin hatasının varyansı,

$$S_T^2 = S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]$$

formülü ile hesaplanır. Tahmin hatasının varyansının karekökü S_T ise tahminin standart hatasıdır.

ÖRNEK: Daha önce verilen örnekte bağımsız değişkenin 20 ve 25 değerlerini alması halinde bağımlı değişkenin değerlerini tahmin ediniz ve her iki durum için tahminin standart hatasını hesaplayınız.

$Y_i = - 4{,}3 + 1{,}03\ X_i$

$S_e^2 = 0{,}6446$

$$\sum (X_i - \overline{X})^2 = 60$$

olarak belirlenmiştir.

$X_o = 20$ için $\hat{Y}_0 = - 4{,}3 + 1{,}03\ (20) = 16{,}3$

ve bu durumda tahminin standart hatası,

$$S_T = \sqrt{S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]}$$

$$= \sqrt{0{,}6446 \left[1 + \frac{1}{5} + \frac{(20 - 10)^2}{60} \right]}$$

$$= 1{,}3593$$

olacaktır.

$X_1 = 25$ için $Y_1 = - 4{,}3 + 1{,}03\ (25) = 21{,}45$

ve bu durumda tahminin standart hatası,

$$S_T = \sqrt{S_e^2\left[1+\frac{1}{n}+\frac{(X_1-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right]}$$

$$= \sqrt{0{,}6446\left[1+\frac{1}{5}+\frac{(25-10)^2}{60}\right]}$$

$$= 1{,}7862$$

olacaktır. Görüldüğü gibi bağımsız değişkenin tahmin edilen değerinin büyümesi ile tahminin standart hatası da büyümektedir.

13.2. KORELASYON KATSAYISI

Regresyon analizinde değişkenler arasındaki ilişki matematiksel bir model ile açıklanmaktaydı. Regresyon modeli değişkenler arasındaki ilişkinin yönünü, bağımsız değişkenin işareti ile belirtmektedir. Ayrıca değişkenlerin karşılıklı değişimlerinin ölçüsü olarak kovaryansın hesaplanabileceğinden daha önce söz edilmişti.

Kovaryans anakütle için,

$$Kov(X,Y) = \sigma_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{N}$$

formülü ile hesaplanacaktır. Anakütle kovaryansı alınacak bir örnek ile tahmin ediliyorsa, formül,

$$S_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$$

olacaktır. Örnek kovaryansı,

$$Kov(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n}$$

anakütle kovaryansının sapmasız tahmincisi değildir. Payda n yerine (n - 1) alındığında formül anakütle varyansının sapmasız tahmincisi olmaktadır.

Kovaryans değişkenler arasındaki ilişkinin yönünü vermekte fakat standart bir ölçü olmadığından ilişkinin derecesini belirlemede yeterli olmamaktadır. Ayrıca, kovaryans standart bir ölçü olmadığından serilerin ilişkilerinin karşılaştırılmasında da yetersiz kalmakta ve bu nedenle korelasyon katsayısı hesaplanmaktadır.

Korelasyon katsayısı, kovaryansın değişkenlerin standart sapmalarına bölünmesi ile elde edilen standart bir ölçüdür. Standart ölçü olduğu için değişkenler arasındaki ilişkinin yönünü belirttiği gibi, ilişkinin derecesini de belirtmektedir.

Verilen tanıma göre anakütle korelasyon katsayısı,

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{N}}{\sigma_X \sigma_Y} = \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{N\sigma_X \sigma_Y}$$

olacaktır.

Anakütle korelasyon katsayısı örnekten tahmin edilirken kovaryans ve standart sapmanın sapmasız tahmincileri kullanılarak,

$$r = \frac{S_{XY}}{S_X S_Y}$$

$$S_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

olduğundan, korelasyon katsayısının tanım formülü,

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)S_X S_Y}$$

olarak elde edilecektir. Burada,

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

ve

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}}$$

formülleri ile hesaplanacaktır.

Korelasyon katsayısının hesaplanmasında regresyonda olduğu gibi bağımlı ve bağımsız değişken ayrımı yoktur. Bu nedenle sadece korelasyon katsayısının hesaplanacağı durumlarda değişkenlerden herhangi biri X_i, diğeri Y_i olarak alınabilir.

Korelasyon katsayısının değeri ∓ 1 arasında değişir.

Korelasyon katsayısı şöyle yorumlanacaktır.

$-1 \le r \le +1$

r =0 ise değişkenler arasında ilişki yoktur.

r < 0 ise değişkenler arasında ters yönlü ilişki vardır.

r > 0 ise değişkenler arasında doğru yönlü ilişki vardır.

r'nin değeri + 1'e yaklaştıkça doğru yönlü ilişki, -1'e yaklaştıkça ters yönlü ilişki kuvvetlenecektir. r'nin değeri sıfıra yaklaşıyorsa ilişki zayıflayacaktır.

r'nin değeri 0,50 civarında ise ilişkinin orta dereceli olduğu söylenebilir.

Örneğin; r = - 0,85 ise değişkenler arasında r < 0 olduğundan ters yönlü, fakat r'nin değeri 1'e yakın olduğundan kuvvetli ilişki olacaktır. Bu sonucu kısaca, değişkenler arasında ters yönlü kuvvetli ilişki vardır şeklinde yorumlayabiliriz. Eğer r = 0,15 ise, değişkenler arasında r > 0 olduğundan doğru yönlü, fakat r'nin değeri 0'a yakın olduğundan zayıf ilişki olacaktır. Bu sonucu kısaca, değişkenler arasında doğru yönlü zayıf ilişki vardır şeklinde yorumlayabiliriz. r'nin değeri sıfıra ne kadar yakın ise ilişkinin zayıf veya bire ne kadar yakın ise ilişkinin kuvvetli olacağı araştırma konusuna ve araştırmacıya göre değişebilir. Bu nedenle kesin sınırların çizilmesi doğru olmayacaktır. Ayrıca bu bölümde verilecek formüller ile hesaplanacak korelasyon katsayısı ile doğrusal ilişkinin yönü ve derecesi belirlendiğinden, yapılacak yorumların tümü doğrusal ilişki ile ilgili olacaktır. Bu şekilde hesaplanan korelasyon katsayısının değerinin küçük olması, değişkenler arasındaki doğrusal ilişkinin zayıf olduğunu gösterecektir.

ÖRNEK: Aşağıda verilen serinin korelasyon katsayısını tanım formülü ile hesaplayınız.

Y_i	X_i	$Y_i - \overline{Y}$	$X_i - \overline{X}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(Y_i - \overline{Y})^2$	$(X_i - \overline{X})^2$
2	5	-3	-2	6	9	4
5	4	0	-3	0	0	9
3	7	-2	0	0	4	0
7	8	2	1	2	4	1
8	11	3	4	12	9	16
25	35			20	26	30

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)S_X S_Y}$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{35}{5} = 7$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{25}{5} = 5$$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{30}{5-1}} = \sqrt{7,5} = 2,7386$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{26}{5-1}} = \sqrt{6,5} = 2,5495$$

$$r = \frac{20}{(5-1)(2,7386)(2,5495)} = \frac{20}{27,9282} = 0,7161$$

tanım formülünde S_x ve S_y formüllerini yerine koyarak, formülde gerekli kısaltmalar yapılırsa,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1} \cdot \frac{\sum(Y_i - \bar{Y})^2}{n-1}}}$$

korelasyon katsayısının ortalamalardan farklar formülü,

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

olarak elde edilecektir. Ortalamadan farklar serileri,

$$x_i = X_i - \bar{X}$$

$$y_i = Y_i - \bar{Y}$$

olarak ifade edilirse formül,

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$$

şeklinde yazılacaktır.

ÖRNEK: Daha önce verilen örnek için korelasyon katsayısını ortalamadan farklar formülü ile hesaplayınız.

Daha önce,

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 20$$

$$\sum(X_i - \bar{X})^2 = 30$$

$$\sum(Y_i - \bar{Y})^2 = 26$$

olarak hesaplanmıştı.

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$$r = \frac{20}{\sqrt{26{,}30}} = \frac{20}{27{,}9284} = 0{,}7161$$

Ortalamadan farkların gerçek değerleri ile hesaplanması için aşağıda verilen formüllerin kullanılabileceğini daha önce göstermiştik.

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y}$$

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$$

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$$

Bu formüllerin, ortalamadan farklar ile hesaplanan korelasyon katsayısı formülünde yerine konması ile korelasyon katsayısının gerçek değerlerden hesaplanma formülü,

$$\mathrm{r} = \frac{\sum \mathrm{X_i Y_i} - \mathrm{n}\bar{\mathrm{X}}\bar{\mathrm{Y}}}{\sqrt{[\sum \mathrm{X_i^2} - \mathrm{n}(\bar{\mathrm{X}})^2][\sum \mathrm{Y_i^2} - \mathrm{n}(\bar{\mathrm{Y}})^2]}}$$

olarak elde edilecektir.

ÖRNEK: Daha önce verilen örnek için korelasyon katsayısını gerçek değerler formülü ile hesaplayınız.

Y_i	X_i	$Y_i X_i$	Y_i^2	X_i^2
2	5	10	4	25
5	4	20	25	16
3	7	21	9	49
7	8	56	49	64
8	11	88	64	121
25	35	195	151	275

$$\bar{X} = 7$$

$$\bar{Y} = 5$$

$$r = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{[\sum X_i^2 - n(\bar{X})^2][\sum Y_i^2 - n(\bar{Y})^2]}}$$

$$= \frac{195 - (5)(7)(5)}{\sqrt{[275 - 5(7)^2][151 - 5(5)^2]}}$$

$$= \frac{20}{\sqrt{30.26}} = \frac{20}{27,9284} = 0,7161$$

Korelasyon katsayısı için standart hata

$$\sigma_r = \sqrt{\frac{1-r^2}{n-2}}$$

formülü ile hesaplanacaktır.

ÖRNEK: Daha önce verilen örnekte n = 5 ve r = 0,71 olarak hesaplanmıştı. Bu örnek için,

$$\sigma_r = \sqrt{\frac{1-(0,71)^2}{5-2}} = 0,40$$

olacaktır.

Korelasyon Katsayısının Testi

Korelasyon katsayısı r, anakütle korelasyon katsayısı ρ 'nun tahmincisidir. Örnekten anakütle korelasyon katsayısı ρ = r olarak tahmin edilir. Korelasyon katsayısı için tahmin hatası,

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

dir.

Tahmin edilen korelasyon katsayısının istatistiksel anlamlılığı için hipotez testi yapılabilir. Temel hipotez,

$$H_0 : \rho = 0$$

şeklinde oluşturulur. Karşıt hipotez ise testin çift veya tek taraflı olmasına göre,

$$H_1 : \rho \neq 0$$

veya

$$H_1 : \rho > 0$$

$$H_1 : \rho < 0$$

olacaktır.

Korelasyon katsayısının anlamlılığının testi için test istatistiği,

$$Z = \frac{r - \rho}{\sigma_r}$$

temel hipotez ρ =0 olduğundan, σ_r 'nin de yerine konması ile,

$$Z = \frac{r}{\sigma_r} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

olacaktır.

Test istatistiğinin dağılımı n'e bağlı olarak normal dağılım veya (n-2) serbestlik dereceli t dağılımıdır. Hipotezlerden hangisinin kabul edileceği kararı t veya normal dağılım tablolarından bulunacak değerler ile test istatistiği karşılaştırılarak verilecektir. Uygun tablo değerini Z_t ile ifade edersek,

$|Z| > |Z_t|$ olması durumunda H_1 hipotezi,

$|Z| < |Z_t|$ olması durumunda H_0 hipotezi kabul edilecektir.

H_0 hipotezinin, yani ρ =0 olmasının kabulü korelasyon katsayısının istatistiksel olarak anlamsız olduğunu ortaya koyacaktır.

ÖRNEK: Daha önce verilen örnekte n=5'ti ve r =0.71 olarak hesaplanmıştı. 0,05 hata payı ile değişkenler arasındaki doğru yönlü ilişki istatistiksel olarak anlamlı mıdır, inceleyiniz.

$H_0 : \rho = 0$

$H_1 : \rho > 0$

$t_{\alpha,n-2} = t_{0,05,3} = 2{,}353$

$$Z = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}71\sqrt{5-2}}{\sqrt{1-(0{,}71)^2}} = 1{,}746$$

1,746 < 2,353 olduğundan H_0 hipotezi kabul edilir.

13.3. BASİT DOĞRUSAL REGRESYON VE KORELASYON ARASINDAKİ İLİŞKİ

Basit doğrusal regresyon modelinin bağımsız değişkeninin katsayısı ortalamadan farklar ile,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

olarak hesaplanıyordu. Pay ve paydayı (n-1)'e bölersek,

$$\hat{\beta} = \frac{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}}{\dfrac{\sum_{i=1}(X_i - \bar{X})^2}{n-1}}$$

$$S_{XY}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

$$S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

olduğundan,

$$\hat{\beta} = \frac{S_{YX}}{S_X^2}$$

olacaktır. Buradan,

$$S_{YX} = \hat{\beta} S_X^2$$

olarak elde edilir. Korelasyon katsayısı,

$$r = \frac{S_{YX}}{S_X S_Y}$$

olduğundan yukarıda yazılan eşitlikle,

$$r = \frac{\hat{\beta} S_X^2}{S_X S_Y} = \hat{\beta} \frac{S_X}{S_Y}$$

ve

$$\hat{\beta} = r \frac{S_Y}{S_X}$$

olarak elde edilir. Görüldüğü gibi S_x ve S_Y ile birlikte $\hat{\beta}$ biliniyorsa r, r biliniyorsa $\hat{\beta}$ hesaplanabilir.

ÖRNEK: Aşağıda verilen seride,

$$\hat{\beta} = r \frac{S_Y}{S_X}$$

olduğunu gösteriniz.

Y_i	X_i	$Y_i-\bar{Y}$	$X_i-\bar{X}$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(Y_i-\bar{Y})^2$	$(X_i-\bar{X})^2$
5	8	-5	-7	35	25	49
8	7	-2	-8	16	4	64
6	12	-4	-3	12	16	9
10	15	0	0	0	0	0
14	23	4	8	32	16	64
17	25	7	10	70	49	100
60	90			165	110	286

$$\bar{X}=\frac{\sum X_i}{n}=\frac{90}{6}=15$$

$$\bar{Y}=\frac{\sum Y_i}{n}=\frac{60}{6}=10$$

$$Y_i=\hat{\alpha}+\hat{\beta}X_i$$

$$\hat{\beta}=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}=\frac{165}{286}=0,57$$

$$r=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$$

$$r=\frac{165}{\sqrt{286.110}}=\frac{165}{177,3696}=0,93$$

$$S_X=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}=\sqrt{\frac{286}{6-1}}=\sqrt{57,2}=7,56$$

$$S_Y=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}}=\sqrt{\frac{110}{6-1}}=\sqrt{22}=4,69$$

$$\hat{\beta}=r\frac{S_Y}{S_X}$$

$$=0,93\frac{4,69}{7,56}=0,57$$

Regresyon modelinde X_i değerlerini yerine koyarak teorik değerler elde edili-yordu.

$$Y_i=\hat{\alpha}+\hat{\beta}X_i$$

eşitliğinde bağımsız değişkenin katsayısı,

$$\hat{\beta} = r\frac{S_Y}{S_X}$$

ve sabit katsayı,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= \overline{Y} - r\frac{S_Y}{S_X}\overline{X}$$

olduğundan, eşitlikleri yerine koyarsak,

$$\hat{Y}_i = \overline{Y} - r\frac{S_X}{S_Y}\overline{X} + r\frac{S_Y}{S_X}X_i$$

$$= \overline{Y} + r\frac{S_X}{S_Y}(X_i - \overline{X})$$

olacak ve bu şekilde tahminde yapılabilecektir.

ÖRNEK: Daha önce verilen örnekte $X_i = 25$ için teorik değeri hesaplayarak,

$$\hat{Y}_i = \overline{Y} + r\frac{S_Y}{S_X}(X_i - \overline{X})$$

olduğunu gösteriniz.

$$\overline{X} = 15$$

$$\overline{Y} = 10$$

$$\hat{\beta} = 0{,}57$$

olduğundan,

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 10 - 0{,}57\ (15) = 1{,}45$$

$$\hat{Y} = 1{,}45 + 0{,}57(25) = 15{,}7$$

bulunur.

$$\hat{Y}_i = \overline{Y} + r\frac{S_Y}{S_X}(X_i - \overline{X})$$

$r = 0{,}93$

$S_Y = 4{,}69$

$S_x = 7{,}56$

olduğundan

$$\hat{Y} = 10 + 0{,}93\frac{4{,}69}{7{,}56}(25 - 15) = 15{,}7$$

olarak elde edilir.

Belirlilik katsayısı,

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

ve

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \hat{\beta}\sum_{i01}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

olduğundan belirlilik katsayısı,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \hat{\beta}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

olarak yazılabilir. Gerekli kısaltmalar yapılırsa,

$$R^2 = 1 - 1 + \frac{\hat{\beta}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

pay ve payda (n-1)'e bölünürse,

$$R^2 = \frac{\hat{\beta}S_{YX}}{S_Y^2}$$

olacaktır. Korelasyon katsayısı,

$$r = \frac{S_{YX}}{S_Y S_X}$$

olduğundan,

$$S_{YX} = rS_Y S_X$$

olarak elde edilir. Bu eşitlik yukarıda verilen R^2 eşitliğinde yerine konursa,

$$R^2 = \frac{\hat{\beta}rS_Y S_X}{S_Y^2} = r\hat{\beta}\frac{S_X}{S_Y}$$

ve

$$r = \hat{\beta}\frac{S_X}{S_Y}$$

olduğundan,

$$R^2 = r.r = r^2$$

olarak bulunur. Görüldüğü gibi basit doğrusal regresyonda korelasyon katsayısının karesi belirlilik katsayısına eşittir.

ÖRNEK: Daha önce verilen örnekte

$$r^2 = R^2$$

olduğunu gösteriniz.

Örnekte r = 0,93 , regresyon modeli,

$Y_i = 1{,}45 + 0{,}57\ X_i$

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = 165$$

$$\sum (X_i - \overline{X})^2 = 110$$

olarak belirlenmişti.

$$\sum (Y_i - \hat{Y})^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta}\sum (X_i - \overline{X})(Y_i - \overline{Y})$$

$$= 110 - 0{,}57\ (165) = 15{,}95$$

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \overline{Y})^2} = 1 - \frac{15{,}95}{110}$$

$$= 1 - 0{,}145 = 0{,}855$$

$$r^2 = (0{,}93)^2 = 0{,}8649$$

olarak bulunur. Hesaplama şekilleri arasındaki fark değerlerin yaklaşık alınmasından kaynaklanmaktadır.

Regresyon modelinde X_i bağımsız, Y_i ise bağımlı değişken olarak alınmaktaydı. X_i bağımlı, Y_i bağımsız değişken alınarak da regresyon modeli oluşturulabilir. X_i'nin bağımsız, Y_i'nin bağımlı değişken olduğu regresyon modeli,

$$Y_i = \hat{\alpha}_{XY} - \hat{\beta}_{YX} X_i$$

ve X_i'nin bağımlı, Y_i'nin bağımsız değişken olduğu regresyon modeli,

$$X_i = \hat{\alpha}_{XY} + \hat{\beta}_{YX} Y_i$$

şeklinde ifade edilirse, ikinci model ile ilgili hesaplamaların yapılabilmesi için daha önce verilen formüllerde X_i ile Y_i'nin yer değiştirmesi yeterli olacaktır.

Yukarıda verilen modellerin bağımsız değişkenlerinin katsayıları,

$$\hat{\beta}_{YX} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

formülleri ile hesaplanacaktır. Verilen bu katsayılar çarpılarak kare kökleri alınırsa,

$$\sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = r$$

olarak elde edilir. Bu durumda β_{YX} ve β_{XY} 'nin işaretleri birbiri ile aynı olacak ve korelasyon katsayısı, regresyon modelinin katsayıları ile aynı işareti taşıyacaktır.

ÖRNEK: Aşağıda verilen seri için,

$$r = \sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}}$$

olduğunu gösteriniz.

Y_i	X_i	$Y_i - \bar{Y}$	$X_i - \bar{X}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})^2$
2	17	-3	5	-15	9	25
4	14	-1	2	-2	1	4
3	11	-2	-1	2	4	1
6	10	1	-2	-2	1	4
10	8	5	-4	-20	25	16
25	60			-37	40	50

$$\bar{X} = \frac{\sum X_i}{n} = \frac{60}{5} = 12$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{25}{5} = 5$$

$$\hat{\beta}_{YX} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{-37}{50} = -0{,}74$$

$$\hat{\beta}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$= \frac{-37}{40} = -0{,}925$$

$$r = \sqrt{\hat{\beta}_{YX}\hat{\beta}_{XY}}$$

$$= \sqrt{(-0{,}74)(-0{,}925)} = -0{,}827$$

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$$= \frac{-37}{\sqrt{50.40}} = \frac{-37}{44{,}721} = -0{,}827$$

13.4. EĞRİSEL KORELASYON VE REGRESYON

Buraya kadar yapılan açıklamalarda iki değişken arasındaki ilişkinin doğrusal olduğu kabul ediliyordu. Değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarda, bu ilişkiyi açıklayacak eğrisel modelin belirlenmesi gerekecektir.

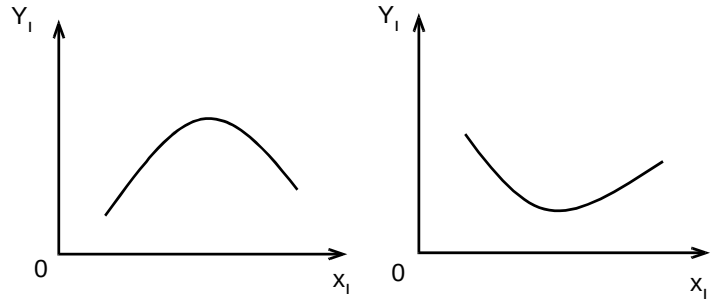
Değişkenler arasındaki eğrisel ilişkiyi açıklamak için kullanılacak modelin matematiksel şeklini belirlemek için yapılabilecek en basit işlem serinin grafiğini çizmektir. Çizilen grafik yardımı ile model belirlenebilirse de, grafikler yanıltıcı sonuçlar verebilirler. Bu nedenle, ilişkiyi açıklamak için uygun olabilecek modellerin denenerek, bunlardan ilişkiyi en iyi açıklayanının, yani belirlilik katsayısı en yüksek olanının seçilmesi daha doğru olacaktır.

Değişkenler arasındaki ilişkiyi açıklamak için çeşitli eğrisel modeller kullanılabilir. Burada en çok kullanılan birkaç eğrisel modele yer verilecektir.

Serinin tek maksimumu veya minimumu varsa değişkenler arasındaki ilişki,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

modeli ile açıklanabilir.



Şekil 13.3

Yukarıda verilen anakütle regresyon modeli alınacak örnekten,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \varepsilon_i$$

olarak tahmin edilecektir. Modelin parametrelerinin en küçük kareler yöntemi ile tahmini, doğrusal modelin parametrelerinin tahminine benzer şekilde yapılarak model için normal denklemler,

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3$$

$$\sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4$$

olarak elde edilir. Bu durumda hata terimlerinin kareleri toplamı,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_i^2 Y_i$$

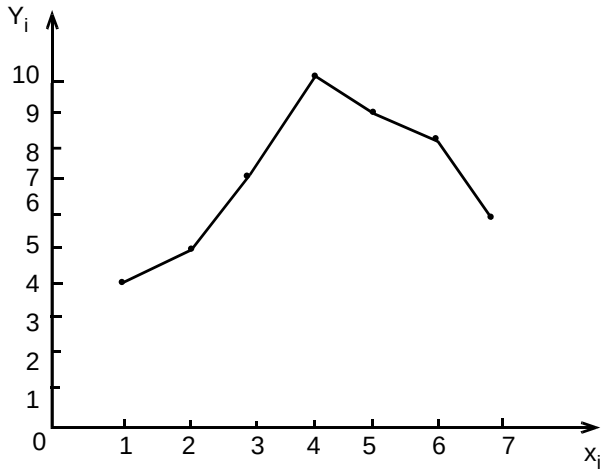
formülü ile hesaplanacaktır. Korelasyon katsayısı ise daha sonra çoklu regresyonda yapılacak açıklamalara göre belirlenecektir.

ÖRNEK: Aşağıdaki serinin grafiğini çizerek,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \varepsilon_i$$

regresyon modelinin parametrelerini tahmin ediniz.

Y_i	X_i	X_i^2	X_i^3	X_i^4	$Y_i\ X_i$	$X_i^2\ Y_i$
4	1	1	1	1	4	4
5	2	4	8	16	10	20
7	3	9	27	81	21	63
10	4	16	64	256	40	160
9	5	25	125	625	45	225
8	6	36	216	1296	48	288
6	7	49	343	2401	42	294
49	28	140	784	4676	210	1054



$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^3$$

$$\sum_{i=1}^{n} Y_i X_i^2 = \hat{\beta}_0 \sum_{i=1}^{n} X_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} X_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^4$$

$$49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$-4/49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$

$$-196 = -28\hat{\beta}_0 - 112\hat{\beta}_1 - 560\hat{\beta}_2$$
$$210 = 28\hat{\beta}_0 + 140\hat{\beta}_1 + 784\hat{\beta}_2$$

$$14 = 0 + 28\hat{\beta}_1 + 224\hat{\beta}_2$$

$$-20/49 = 7\hat{\beta}_0 + 28\hat{\beta}_1 + 140\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$-980 = -140\hat{\beta}_0 - 560\hat{\beta}_1 - 2800\hat{\beta}_2$$
$$1054 = 140\hat{\beta}_0 + 784\hat{\beta}_1 + 4676\hat{\beta}_2$$

$$74 = 0 + 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-8/14 = 28\hat{\beta}_1 + 224\hat{\beta}_2$$
$$74 = 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-112 = -224\hat{\beta}_1 - 1792\hat{\beta}_2$$
$$74 = 224\hat{\beta}_1 + 1876\hat{\beta}_2$$

$$-38 = 0 + 84\hat{\beta}_2$$

$$\beta_2 = \frac{-38}{84} = -0,45$$

$$14 = 28\hat{\beta}_1 + 224(-0,45)$$

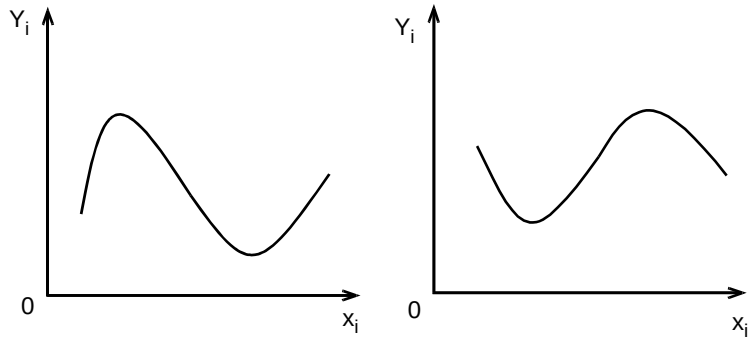
$$\beta_1 = \frac{14 + 100,8}{28} = 4,1$$

$$49 = 7\hat{\beta}_0 + 28(4,1) + 140(-0,45)$$

$$\beta_0 = \frac{49 - 114,8 + 63}{7} = -0,4$$

$$Y_i = -0,4 + 4,1X_i - 0,45X_i^2$$

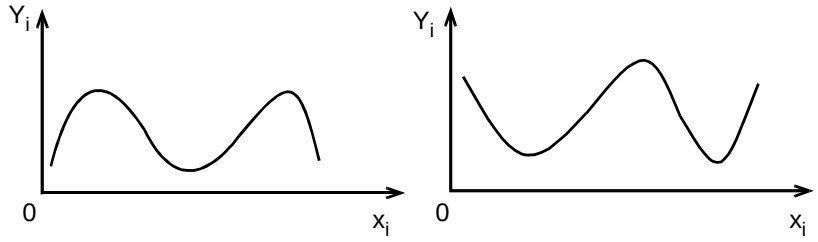
Serinin bir maksimumu ve bir minimumu varsa (Şekil 13.4) örnek regresyon modeli,



Şekil 13.4

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3 + \varepsilon_i$$

serinin iki maksimumu ve bir minimumu veya bir maksimumu ve iki minimumu varsa(Şekil 13.5) örnek regresyon modeli,

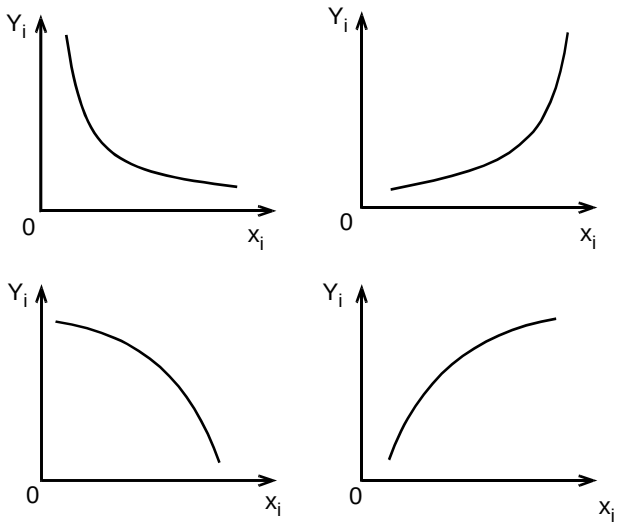


Şekil 13.5

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3 + \hat{\beta}_4 X_i^4 + \varepsilon_i$$

olacaktır. Görüldüğü gibi denklemde bağımsız değişkenin kuvveti serinin maksimum ve minimum sayıları toplamından bir fazladır. Bu tür modeller için normal denklemler doğrusal modelin normal denklemlerinin elde edilmesine benzer şekilde elde edilecek ve çözüm için tahmin edilecek parametre sayısı kadar denklem olacaktır.

Serinin maksimum ve minimumu yoksa, fakat ilişki doğrusal değilse (Şekil 13.6), değişkenler arasındaki ilişki yarı logaritmik veya tam logaritmik modellerle açıklanabilir.



Şekil 13.6

Değişkenlerden sadece birinin logaritmasının alındığı modeller yarı logaritmik model olarak adlandırılmaktadır.

$$Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

ve

$$\ln Y_i = \hat{\alpha} + \hat{\beta} X_i + e_i$$

modelleri yarı logaritmik modellerdir. Logaritmalar 10 veya e tabanına göre alınabilir. Her iki değişkenin logaritmasının alındığı modele ise tam logaritmik model denilmektedir.

$$Y_i = \hat{\alpha}.X_i^{\hat{\beta}}.e^{e_i}$$

modelin logaritması alınırsa,

$$\ln Y_i = \ln \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

olacaktır. Logaritmik modellerde $\ln e_i$ olması gereken hata terimleri kolaylık olması için e_i ile ifade edilmiştir.

Yarı logaritmik ve tam logaritmik modellerin normal denklemleri ve korelasyon katsayısı tanım formülleri doğrusal model ile aynı şekilde elde edilirler. Elde edilecek formüller, Xi yerine $\ln X_i$ ve Y_i yerine $\ln Y_i$'nin yer aldığı doğrusal regresyon formülleri ile aynı olduklarından burada formüllerin çıkartılması gösterilmeyecektir. Ayrıca, logaritma alındığında ortalamadan farklar formülleri ile yapılacak işlemler güçleşeceğinden sadece gerçek değerler ile ilgili formüller verilecektir.

$$Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$$

Yarı logaritmik regresyon modeli için normal denklemler,

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} \ln X_i$$

$$\sum_{i=1}^{n} Y_i (\ln X_i) = \hat{\alpha} \sum_{i=1}^{n} \ln X_i + \hat{\beta} \sum_{i=1}^{n} (\ln X_i)^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i (\ln X_i) - n\bar{Y}(\overline{\ln X})}{\sum_{i=1}^{n} (\ln X_i)^2 - n(\overline{\ln X})^2}$$

ve

$$\hat{\alpha} = \bar{Y} - \hat{\beta}(\overline{\ln X})$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n} Y_i (\ln X_i) - n\bar{Y}(\overline{\ln X})}{\sqrt{[\sum_{i=1}^{n} (\ln X_i)^2 - n(\overline{\ln X})^2][\sum_{i=1}^{n} Y_i^2 - n(\bar{Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\alpha} \sum_{i=1}^{n} Y_i - \hat{\beta} \sum_{i=1}^{n} Y_i (\ln X_i)$$

olacaktır.

ÖRNEK: Aşağıda verilen seri için $Y_i = \hat{\alpha} + \hat{\beta} \ln X_i + e_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $X_6 = 10$ için $\hat{Y}_6$ değerini tahmin ediniz.

Y_i	X_i	$\ln X_i$	$(\ln X_i)^2$	$Y_i (\ln X_i)$	Y_i^2
1	2	0,69314	0,48044	0,69314	1
4	5	1,60943	2,59026	6,43772	16
7	6	1,79175	3,21036	12,54225	49
13	8	2,07944	4,32407	27,03272	169
15	9	2,19722	4,82777	32,9583	225
40	30	8,37098	15,4329	79,66413	460

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{40}{5} = 8$$

$$(\overline{\ln X}) = \frac{\sum \ln X_i}{n} = \frac{8,37098}{5} = 1,67419$$

$$\hat{\beta} = \frac{\sum Y_i (\ln X_i) - n\bar{Y}(\overline{\ln X})}{\sum (\ln X_i)^2 - n(\overline{\ln X})^2}$$

$$\hat{\beta} = \frac{79,66413 - (5)(8)(1,67419)}{15,43290 - 5(1,67419)^2}$$

$$= \frac{12,69653}{1,41834} = 8,95$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}(\overline{\ln X})$$

$$= 8 - 8,95(1,67419)$$

$$= 8\text{-}14,98 = \text{-}6,98$$

$$Y_i = -6,98 + 8,95 \ln X_i$$

$$r = \frac{\sum Y_i (\ln X_i) - n\bar{Y}(\overline{\ln X})}{\sqrt{[\sum (\ln X_i)^2 - n(\overline{\ln X})^2][\sum Y_i^2 - n(\bar{Y})^2]}}$$

$$r = \frac{79,66413 - 5(8)(1,67413)}{\sqrt{[15,43290 - 5(1,67419)^2][460 - 5(8)^2]}}$$

$$= \frac{12,69653}{\sqrt{(1,41834)(140)}}$$

$$= \frac{12,69653}{14,09140} = 0,90$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\alpha}\sum_{i=1}^{n} Y_i - \hat{\beta}\sum_{i=1}^{n} Y_i(\ln X_i)$$

= 460 - (-6,98) (40) - 8,95 (79,66413)

= 460 + 279,2 - 712,99

= 26,21

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{26,21}{5-2} = 8,73$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{26,21}{140} = 1 - 0,187 = 0,813$$

$\hat{Y}_6$ = -6,98 + 8,95 ($\ln X_6$)

= -6,98 + 8,95 (ln10)

= -6,98 + 8,95 (2,30258)

= 13,62

Yarı logaritmik regresyon modeli,

$$\ln Y_i = \hat{\alpha} + \hat{\beta} X_i$$

ise normal denklemler,

$$\sum_{i=1}^{n} (\ln Y_i) = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i(\ln Y_i) = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i(\ln Y_i) - n(\overline{X})(\overline{\ln Y})}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

ve

$$\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}\overline{X}$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n} X_i(\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sqrt{[\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2][\sum_{i=1}^{n} (\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı ise,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \ln Y_i^2 - \hat{\alpha}\sum_{i=1}^{n} (\ln Y_i) - \hat{\beta}\sum_{i=1}^{n} X_i(\ln Y_i)$$

olacaktır.

ÖRNEK: Daha önce verilen seri için $\ln Y_i = \hat{\alpha} + \hat{\beta}X_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $X_6 = 10$ için $\hat{Y}_6$ değerini tahmin ediniz.

Y_i	X_i	$\ln Y_i$	X_i^2	$X_i(\ln Y_i)$	$(\ln Y_i)^2$
1	2	0	4	0	0
4	5	1,38629	25	6,93145	1,92179
7	6	1,94591	36	11,67546	3,78656
13	8	2,56494	64	20,51952	6,57891
15	9	2,70805	81	24,37245	7,33353
40	30	8,60519	210	63,49888	19,62079

$$\overline{X} = \frac{\sum X_i}{n} = \frac{30}{5} = 6$$

$$(\overline{\ln Y}) = \frac{\sum \ln Y_i}{n} = \frac{8,60519}{5} = 1,72103$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i(\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sum_{i=1}^{n} X_i^{\,2} - n(\overline{X})^2}$$

$$\hat{\beta} = \frac{63{,}49888 - (5)(6)(1{,}72103)}{210 - 5(6)^2}$$

$$= \frac{11{,}86798}{30} = 0{,}39$$

$$\hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{X})$$

$$= 1{,}72103 - 0{,}39(6)$$

$$= 1{,}72103 - 2{,}34 = -0{,}62$$

$$\ln Y_i = -0{,}62 + 0{,}39 X_i$$

$$r = \frac{\sum X_i(\ln Y_i) - n\overline{X}(\overline{\ln Y})}{\sqrt{[\sum X_i^{\,2} - n(\overline{X})^2][\sum (\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

$$r = \frac{63{,}49888 - 5(6)(1{,}72103)}{\sqrt{[210 - 5(6)^2][19{,}62079 - 5(1{,}72103)^2]}}$$

$$= \frac{11{,}86798}{\sqrt{(30)(4{,}81)}} = \frac{11{,}86798}{12{,}0124} = 0{,}98$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\ln Y_i)^2 - \hat{\alpha}\sum_{i=1}^{n} (\ln Y_i) - \hat{\beta}\sum_{i=1}^{n} X_i(\ln Y_i)$$

$$= 19{,}62079 - (-0{,}62)\,(8{,}60519) - 0{,}39\,(63{,}49888)$$

$$= 0{,}19144$$

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{0{,}19144}{5-2} = 0{,}06381$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\ln Y_i - (\overline{\ln Y})]^2}$$

$$\sum[\ln Y_i - (\overline{\ln Y})]^2 = \sum(\ln Y_i)^2 - n(\overline{\ln Y})^2$$

$$= 19{,}62079 - 5(1{,}72103)^2 = 4{,}81106$$

$$R^2 = 1 - \frac{0{,}19144}{4{,}81106}$$

$$= 1 - 0{,}039 = 0{,}961$$

$$\ln \hat{Y}_6 = -0{,}62 + 0{,}39 X_6$$

$$= -0{,}62 + 0{,}39(10)$$

$$= -0{,}62 + 3{,}9 = 3{,}28$$

$$\hat{Y}_6 = 26{,}57$$

Tam logaritmik regresyon modeli,

$$\ln Y_i = \ln \hat{\alpha} + \hat{\beta} \ln X_i$$

ise normal denklemler,

$$\sum_{i=1}^{n}(\ln Y_i) = n \ln \hat{\alpha} + \hat{\beta} \sum_{i=1}^{n}(\ln X_i)$$

$$\sum_{i=1}^{n}(\ln Y_i)(\ln X_i) = \ln \hat{\alpha} \sum_{i=1}^{n}(\ln X_i) + \hat{\beta} \sum_{i=1}^{n}(\ln X_i)^2$$

olacaktır. Bu denklemlerden,

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sum_{i=1}^{n}(\ln X)^2 - n(\overline{\ln X})^2}$$

ve

$$\ln \hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{\ln X})$$

olarak elde edilir. Bu durumda korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n}(\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sqrt{[\sum_{i=1}^{n}(\ln X_i)^2 - n(\overline{\ln X})^2][\sum_{i=1}^{n}(\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

formülü ile hesaplanacaktır. Hata terimlerinin kareleri toplamı ise,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \ln Y_i^2 - (\ln \hat{\alpha}) \sum_{i=1}^{n} \ln Y_i - \hat{\beta} \sum_{i=1}^{n}(\ln X_i)(\ln Y_i)$$

olacaktır.

ÖRNEK: Daha önce verdiğimiz $\ln Y_i = \ln \hat{\alpha} + \hat{\beta} \ln X_i$ regresyon modelinin parametrelerini, S_e^2 değerini, korelasyon katsayısını, belirlilik katsayısını hesaplayınız ve $\hat{X}_6 = 10$ için Y_6 değerini tahmin ediniz.

Y_i	X_i	$\ln Y_i$	$\ln X_i$	$(\ln X_i)(\ln Y_i)$
1	2	0	0,69314	0
4	5	1,38629	1,60943	2,23113
7	6	1,94591	1,79175	3,48658
13	8	2,56494	2,07944	5,33363
15	9	2,70805	2,19722	5,95018
40	30	8,60519	8,37098	17,00152

Daha önce,

$(\overline{\ln Y}) = 1,72103 \quad \sum (\ln Y_i)^2 = 19,62079$

$(\overline{\ln X}) = 1,67419 \quad \sum (\ln X_i)^2 = 15,43290$

olarak hesaplanmıştı.

$$\hat{\beta} = \frac{\sum (\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sum (\ln X_i)^2 - n(\overline{\ln X})^2}$$

$$= \frac{17,00152 - 5(1,67419)(1,72103)}{15,43290 - 5(1,67419)^2}$$

$$= \frac{2,59486}{1,41834} = 1,82$$

$$\ln \hat{\alpha} = (\overline{\ln Y}) - \hat{\beta}(\overline{\ln X})$$

$$= 1,72103 - 1,82(1,67419)$$

$$= 1,72103 - 3,04702$$

$$= -1,32$$

$$\ln Y_i = -1,32 + 1,82(\ln X_i)$$

$$r = \frac{\sum (\ln X_i)(\ln Y_i) - n(\overline{\ln X})(\overline{\ln Y})}{\sqrt{[\sum (\ln X_i)^2 - n(\overline{\ln X})^2][\sum (\ln Y_i)^2 - n(\overline{\ln Y})^2]}}$$

$$= \frac{17,00152 - 5(1,67419)(1,72103)}{\sqrt{[15,43290 - 5(1,67419)^2][19,62079 - 5(1,72103)^2}}$$

$$= \frac{2,59486}{\sqrt{1,41834 .4,81106}}$$

$$= \frac{2,59486}{2,61222} = 0,99$$

$$\sum e_i^2 = \sum (\ln Y_i)^2 - (\ln \hat{\alpha}) \sum \ln Y_i - \hat{\beta} \sum (\ln X_i)(\ln Y_i)$$

$$= 19,62079 - (-1,32)(8,60519) - 1,82(17,00152)$$

$$= 0,03687$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\ln Y_i - (\overline{\ln Y})]^2}$$

$$= 1 - \frac{0,03687}{4,81106}$$

$$= 1 - 0,007 = 0,993$$

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

$$= \frac{0,03687}{5-2} = 0,01229$$

$$\ln \hat{Y}_6 = -1,32 + 1,82(\ln X_6)$$

$$= -1,32 + 1,82(2,30258)$$

$$= 2,87069$$

$$\hat{Y}_6 = 17,64$$

Aynı değişkenler için farklı matematiksel yapıdaki modeller denendiğinde bunlardan hangisinin değişkenler arasındaki ilişkiyi daha iyi açıkladığının belirlenmesi gerekecektir. Modellerden hata terimlerinin kareleri toplamı veya varyansı küçük olan, değişkenler arasındaki ilişkiyi daha iyi açıklamaktadır. Modellerin hata terimlerinin varyansı veya karelerinin toplamı ile karşılaştırılabilmesi için modellerin bağımlı değişkenlerinin aynı yapıda olması gerekmektedir. Bu durumda hata terimlerinin varyansı veya kareleri toplamı daha küçük olan model değişkenler arasındaki ilişkiyi daha iyi açıklayacaktır.

Uygun modelin belirlenmesinde belirlilik katsayısı da kullanılabilir. Bağımlı değişkenler aynı yapıda olmadığında da modellerin belirlilik katsayısı ile karşılaştırılması mümkündür. Bağımlı değişkenlerin bazıları gerçek değerlerle, bazıları logaritmik olarak ifade ediliyorsa bu modeller hata terimlerinin tahmincilerinin varyansları veya karelerinin toplamı ile karşılaştırılamazlar.

Bu durumda belirlilik katsayısı büyük olan model, değişkenler arasındaki ilişkiyi daha iyi açıklayacağından bu modelin seçilmesi uygun olacaktır.

ÖRNEK: Daha önce verilen örnek için tam logaritmik ve yarı logaritmik modeller belirlenmişti. Aynı seri için,

$$Y_i = \hat{\alpha} + \hat{\beta} X_i$$

ve

$$Y_i = \hat{\alpha} + \hat{\beta} X_i^2$$

Modellerinin katsayılarını tahmin ediniz.

Y_i	X_i	X_i^2	$X_i Y_i$	Y_i^2	$X_i^2 Y_i$	X_i^3	X_i^4
1	2	4	2	1	4	8	16
4	5	25	20	16	100	125	625
7	6	36	42	49	252	216	1296
13	8	64	104	169	832	512	4096
15	9	81	135	225	1215	729	5661
40	30	210	303	460	2403	1590	12594

$$\overline{X} = 6$$

$$\overline{Y} = 8$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$= \frac{303 - 5(6)(8)}{210 - 5(6)^2}$$

$$= \frac{63}{30} = 2{,}1$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 8 - 2{,}1(6) = -4{,}6$$

$$Y_i = -4{,}6 + 2{,}1 X_i$$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$= 460 - (-4{,}6)(40) - 2{,}1(303) \quad = 7{,}7$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y - \overline{Y})^2}$$

$$= 1 - \frac{\sum e_i^2}{\sum Y^2 - n\overline{Y}^2}$$

$$= 1 - \frac{7,7}{140}$$

$$= 1 - 0,055 = 0,945$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 + \hat{\beta}_2 \sum X_i^3$$

$$\sum X_i^2 Y_i = \hat{\beta}_0 \sum X_i^2 + \hat{\beta}_1 \sum X_i^3 + \hat{\beta}_2 \sum X_i^4$$

$$-6/40 = 5\hat{\beta}_0 + 30\hat{\beta}_1 + 210\hat{\beta}_2$$

$$303 = 30\hat{\beta}_0 + 210\hat{\beta}_1 + 1590\hat{\beta}_2$$

$$-240 = -30\hat{\beta}_0 - 180\hat{\beta}_1 - 1260\hat{\beta}_2$$

$$303 = 30\hat{\beta}_0 + 210\hat{\beta}_1 + 1590\hat{\beta}_2$$

$$63 = 0 + 30\hat{\beta}_1 + 330\hat{\beta}_2$$

$$-42/40 = 5\hat{\beta}_0 + 30\hat{\beta}_1 + 210\hat{\beta}_2$$

$$2403 = 210\hat{\beta}_0 + 1590\hat{\beta}_1 + 12594\hat{\beta}_2$$

$$-1680 = -210\hat{\beta}_0 - 1260\hat{\beta}_1 - 8820\hat{\beta}_2$$

$$2403 = 210\hat{\beta}_0 + 1590\hat{\beta}_1 + 12594\hat{\beta}_2$$

$$723 = 0 + 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$-11/63 = 30\hat{\beta}_1 + 330\hat{\beta}_2$$

$$723 = 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$-693 = -330\hat{\beta}_1 - 3630\hat{\beta}_2$$

$$723 = 330\hat{\beta}_1 + 3774\hat{\beta}_2$$

$$30 = 144\hat{\beta}_2$$

$$\hat{\beta}_2 = \frac{30}{144} = 0{,}20$$

$$63 = 30\hat{\beta}_1 + 330(0{,}20)$$

$$\hat{\beta}_1 = \frac{63 - 66}{30}$$

$$= \frac{-3}{30} = -0{,}10$$

$$40 = 5\,\hat{\beta}_0 + 30(-0{,}10) + 210(0{,}20)$$

$$\hat{\beta}_0 = \frac{40 + 3 - 42}{5}$$

$$= \frac{1}{5} = 0{,}2$$

$$Y_i = 0{,}2 - 0{,}1X_i + 0{,}2X_i^2$$

$$\sum e_i^2 = \sum Y^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum X_i Y_i - \hat{\beta}_2 \sum X_i^2 Y_i$$

$$= 460 - 0{,}2(40) - (-0{,}1)(303) - 0{,}2(2403) = 1{,}7$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum Y_i^2 - n\bar{Y}^2}$$

$$= 1 - \frac{1{,}7}{460 - 5(8)^2}$$

$$= 1 - \frac{1{,}7}{140}$$

$$= 1 - 0{,}012 \qquad = 0{,}988$$

13.5. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir bölgede yaşayan ailelerin yıllık ortalama gelirleri ile ulaşım harcamaları aşağıda verilmiştir. (1=1000 TL)

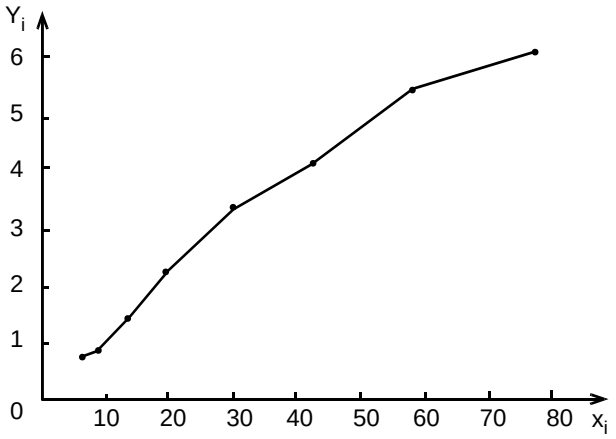
Yıllar	Gelir	Ulaşım H.
2002	5,2	0,7
2003	8,5	0,9
2004	13,4	1,5
2005	20,2	2,3
2006	30,5	3,4
2007	42,0	4,1
2008	59,3	5,2
2009	80,9	5,9
	260,0	24,0

a) Serinin grafiğini çiziniz.
b) Basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.
c) Parametre tahmincilerinin varyanslarını hesaplayınız.
d) Korelasyon katsayısını hesaplayınız.
e) Belirlilik katsayısını hesaplayınız.
f) 2010 yılında ailelerin ortalama gelirlerinin 2009 yılına göre % 60 artacağı varsayımı ile 2010 yılı ortalama ulaşım harcamalarını tahmin ederek, tahminin standart hatasını hesaplayınız.
g) Sabit ve eğim katsayılarını 0,05 hata payı ile çift taraflı olarak test edin.
h) F testini yaparak t_{β}^{2} =F bağıntısını gösterin.

Çözüm:

Y_i	X_i	X_iY_i	X_i^2	Y_i^2
0,7	5,2	3,64	27,04	0,49
0,9	8,5	7,65	72,25	0,81
1,5	13,4	20,10	179,56	2,25
2,3	20,2	46,46	408,04	5,29
3,4	30,5	103,70	930,25	11,56
4,1	42,0	172,20	1764,00	16,81
5,2	59,3	308,26	3516,49	27,04
5,9	80,9	477,31	6544,81	34,81
24,0	260,0	1139,42	13442,44	99,06

a)



b) $\overline{X} = \dfrac{\sum X_i}{n} = \dfrac{260}{8} = 32{,}5$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{24}{8} = 3$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$= \frac{1139{,}42 - 8(32{,}5)(3)}{13442{,}44 - 8(32{,}5)^2}$$

$$= \frac{359{,}42}{4992{,}44} = 0{,}071$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 3 - 0{,}071(32{,}5) = 0{,}692$$

$$Y_i = 0{,}692 + 0{,}071 X_i$$

Gelirin bin Türk lirası artması halinde ulaşım harcamaları (0,071 x 1.000 =) 71 TL artmaktadır. Gelir olmasa bile ($X_i = 0$) (0,692 x 1.000 =) 692 TL ulaşım harcaması yapılacaktır.

a) $S_e^2 = \dfrac{\sum e_i^2}{n-2}$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

$$= 99{,}06 - 0{,}692(24) - 0{,}071(1139{,}42)$$

$$= 1{,}553$$

$$S_e^2 = \frac{1{,}553}{8-2} = 0{,}258$$

$$S_{\hat{\alpha}}^2 = S_e^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right]$$

$$\sum (X_i - \overline{X})^2 = \sum X_i^2 - n\overline{X}^2$$

$$= 13442{,}44 - 8(32{,}5)^2 \quad = 4992{,}44$$

$$S_{\hat{\alpha}}^2 = 0{,}258\left[\frac{1}{8} + \frac{(32{,}5)^2}{4992{,}44}\right]$$

$$= 0{,}08683$$

$$S_{\hat{\beta}}^2 = \frac{S_e^2}{\sum (X_i - \overline{X})^2}$$

$$= \frac{0{,}258}{4992{,}44} = 0{,}000051$$

d)

$$r = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sqrt{[\sum X_i{}^2 - n\overline{X}^2][\sum Y_i^2 - n\overline{Y}^2]}}$$

$$= \frac{1139{,}42 - 8(32{,}5)(3)}{\sqrt{[13442{,}44 - 8(32{,}5)^2][99{,}06 - 8(3)^2]}}$$

$$= \frac{359{,}42}{\sqrt{(4992{,}44).(27{,}06)}}$$

$$= \frac{359{,}42}{367{,}55} = 0{,}977$$

e) $R^2 = r^2 = (0{,}977)^2 = 0{,}954$

ve

$$\sum e_i^2 = 1{,}553$$

$$\sum (Y - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

$$= 99{,}06 - 8(3)^2 = 27{,}06$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{1,553}{27,06}$$

$$= 1 - 0,057 = 0,943$$

Gelir, ulaşım harcamalarındaki değişmelerin yaklaşık % 95'ini açıklamaktadır.

f) Artış oranı 0,60 olursa 2010 yılı geliri,

(80,9)(1,6) = 129,44

olacaktır.

$$\hat{Y}_{10} = 0,692 + 0,071\ X_{10}$$

$$= 0,692 + 0,071\ (129,44) = 9,882$$

$$S_T = \sqrt{S_e^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]}$$

$$= \sqrt{0,258 \left[1 + \frac{1}{8} + \frac{(129,44 - 32,5)^2}{4992,44} \right]}$$

$$= \sqrt{0,77588} = 0,880$$

g) $Y_i = 0,692 + 0,071 X_i$

$S_{\hat{\alpha}}^2 = 0,08683$ $\qquad$ $S_{\hat{\alpha}} = 0,2946$

$S_{\hat{\beta}}^2 = 0,000051$ $\qquad$ $S_{\hat{\beta}} = 0,0071$

olarak bulunmuştur.

α 'nın testi:

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

$t_{\alpha/2,n-2} = t_{0,05/2,8-2} = 2,447$

$$t = \frac{\hat{\alpha}}{S_{\hat{\alpha}}} = \frac{0,692}{0,2946} = 2,348$$

$|2,348| < |2,447|$ olduğundan H_0 hipotezi kabul edilir.

β 'nın testi:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$t_{\alpha/2,n-2} = t_{0,05/2,8-2} = 2,447$

$$t = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{0,071}{0,0071} = 10$$

$|10| > |2,447|$ olduğundan H_1 hipotezi kabul edilir.

h) $H_0 : \beta = 0$

$H_1 : \beta \neq 0$

$F_{\alpha,1,n-2} = F_{0,05,1,6} = 5,99$

$$F = \frac{R^2}{1-R^2} . \frac{n-2}{1}$$

$$= \frac{0,943}{1-0,943} . \frac{6}{1} = 99,26$$

$|99,26| > |5,99|$ olduğundan H_1 hipotezi kabul edilir.

$t_\beta = 10$

$t_\beta{}^2 = (10)^2 = 100 \cong 99,26$

sapma işlemlerden kaynaklanmaktadır.

ÖRNEK 2: Bir ailenin yedi yıllık geliri ve gıda harcamaları aşağıda verilmiştir. (1 = 100 T.L.)

Yıllar	Gelir	Gıda Harcamaları
2003	13	3
2004	18	4
2005	24	6
2006	31	9
2007	25	8
2008	45	15
2009	54	11
	210	56

a) $Y_i = \beta_0 + \beta_1 X_i + e_i$ regresyon modelinin parametrelerini tahmin ediniz.

b) Hata terimlerinin varyansını tahmin ediniz.

c) Korelasyon katsayısını hesaplayınız.

d) Belirlilik katsayısını hesaplayınız.

e) Ailenin 2010 yılı gelirinin en az 650 T.L., ve en çok 800 T.L. olacağını tahmin edildiğine göre kişinin 2010 yılı gıda harcamalarını tahmin ediniz.

Çözüm:

X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X})(Y_i - \overline{Y})$	$(X_i - \overline{X})^2$	$(Y_i - \overline{Y})^2$
13	3	-17	-5	85	289	25
18	4	-12	-4	48	144	16
24	6	-6	-2	12	36	4
31	9	1	1	1	1	1
25	8	-5	0	0	25	0
45	15	15	7	105	225	49
54	11	24	3	72	576	9
210	56	0	0	323	1296	104

a) $\overline{X} = \dfrac{\sum X_i}{n} = \dfrac{210}{7} = 30$

$\overline{Y} = \dfrac{\sum Y_i}{n} = \dfrac{56}{7} = 8$

$\hat{\beta} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$

$= \dfrac{323}{1296} = 0{,}249$

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$

$= 8 - 0{,}249\ (30) = 0{,}53$

$Y_i = 0{,}53 + 0{,}249\ X_i$

Gelirin 100 TL artması gıda harlamalarını 100 x 0,249 = 24,9 lira arttıracaktır. Ailenin gelirinin olmaması durumunda gıda harcamaları 100 x 0,53 = 53 TL olacaktır.

b) $S_e^2 = \dfrac{\sum e_i^2}{n-2}$

$\sum e_i^2 = \sum (Y_i - \overline{Y})^2 - \hat{\beta} \sum (X_i - \overline{X})(Y_i - \overline{Y})$

$= 104 - 0{,}249(323) = 23{,}573$

$S_e^2 = \dfrac{23{,}573}{7-2} = 4{,}7146$

c) $$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

$$= \frac{323}{\sqrt{1296.104}}$$

$$= \frac{323}{367,12} = 0,879$$

Gelir ile gıda harcamaları arasında doğru yönlü kuvvetli ilişki vardır.

d) $R^2 = r^2 = (0,879)^2 = 0,772$

veya

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y - \overline{Y})^2}$$

$$= 1 - \frac{23,573}{104}$$

$$= 1 - 0,226 = 0,774$$

Gelir, gıda harcamalarındaki değişimin %77,4 ünü açıklamaktadır.

e) Yi = 0,53 + 0,24 9 X_i

$\hat{X}_{10,1} = 650$

$\hat{X}_{10,2} = 800$

$\hat{Y}_{10,1}$ = 0,53 + 0,249(650) = 167,15

$\hat{Y}_{10,2}$ = 0,53 + 0,249(800) = 204,5

Ailenin 2010 yılı geliri 650 TL olursa gıda harcamaları 100 x 167,15 = 16715 lira, 800 TL olursa gıda harcamaları 100 x 204,5 = 20450 TL olacaktır.

ÖRNEK 3: Aşağıda bir işletmenin 12 yıllık satışları ve karları ile ilgili olarak aşağıdaki değerler hesaplanmıştır. (1=100 T.L.)

$$\sum_{i=1}^{12} Y_i = 96 \qquad \sum_{i=1}^{12} Y_i^2 = 1064 \qquad \sum_{i=1}^{12} X_i Y_i = 6564$$

$$\sum_{i=1}^{12} X_i = 666 \qquad \sum_{i=1}^{12} X_i^2 = 42922$$

Bu değerlerden yararlanarak,

a) Basit doğrusal regresyon modelinin parametrelerini tahmin ediniz.

b) Korelasyon katsayısını hesaplayınız.

c) Belirlilik katsayısını hesaplayınız.

Çözüm:

a) $\bar{Y} = \dfrac{\sum Y_i}{n} = \dfrac{96}{12} = 8$

$\bar{X} = \dfrac{\sum X_i}{n} = \dfrac{666}{12} = 55{,}5$

$\hat{\beta} = \dfrac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2}$

$= \dfrac{6564 - 12(55{,}5)(8)}{42922 - 12(55{,}5)^2}$

$= \dfrac{1236}{5959} = 0{,}207$

$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

$= 8 - 0{,}207(55{,}5) = -3{,}488$

$Y_i = -3{,}488 + 0{,}207\ X_i$

Satışların 100 T.L. artışı, karı (100)(0,207) = 20,7 T.L. arttıracaktır. Satış yapılmadığında işletmenin zararı (100) (-3,488) = - 348,8 - T.L. olacaktır.

b) $r = \dfrac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{[\sum X_i^2 - n(\bar{X})^2][\sum Y_i^2 - n(\bar{Y})^2]}}$

$= \dfrac{6564 - 12(55,5)(8)}{\sqrt{[42922 - 12(55,5)^2][1064 - 12(8)^2]}}$

$= \dfrac{1236}{\sqrt{5959.296}} = 0{,}930$

satışlar ve kar arasında doğru yönlü kuvvetli ilişki vardır.

c) $R^2 = r^2 = (0{,}93)^2 = 0{,}8649$

veya

$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$

$= 1064 - (-3{,}488)\ (96) - (0{,}207)\ (6564) = 40{,}1$

$R^2 = 1 - \dfrac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n(\bar{Y})^2$$

$$= 1064 - 12(8)^2 = 296$$

$$R^2 = 1 - \frac{40,1}{296}$$

$$= 1 - 0,135 = 0,865$$

Satışlar kardaki değişmelerin % 86, 5'ini açıklamaktadır.

ÖRNEK 4: Aşağıda bir malın üretim miktarları ile maliyetleri verilmiştir.

Maliyet (1=1000 TL)	Üretim Miktarı (1=1000 Adet)
22	1
18	2
16	3
14	4
12	7
11	10

a) $Y_i = \alpha + \beta X_i$

b) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$

c) $Y_i = \alpha + \beta \log X_i$

d) $\log Y_i = \alpha + \beta X_i$

e) $\log Y_i = \alpha + \beta \log X_i$

modellerini tahmin ederek, belirlilik katsayılarını hesaplayınız.

Çözüm:

X_i	Y_i	X_iY_i	X_i^2	Y_i^2	X_i^3	X_i^4	$X_i^2Y_i$
1	22	22	1	484	1	1	22
2	18	36	4	324	8	16	72
3	16	48	9	256	27	81	144
4	14	56	16	196	64	256	224
7	12	84	49	144	343	2401	588
10	11	110	100	121	1000	10000	1100
27	93	356	179	1525	1443	12755	2150

a) $Y_i = \hat{\alpha} + \hat{\beta} X_i$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{27}{6} = 4,5$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{93}{6} = 15{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n(\bar{X})^2}$$

$$= \frac{356 - 6(4{,}5)(15{,}5)}{179 - 6(4{,}5)^2}$$

$$= \frac{-62{,}5}{57{,}5} = -1{,}086$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

= 15,5 – (-1,086)(4,5) = 20,387

$Y_i = 20{,}387 - 1{,}086\ X_i$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha}\sum Y_i - \hat{\beta}\sum X_i Y_i$$

= 1525 – 20,387(93) – (1,086)(356) =15,625

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n(\bar{Y})^2$$

= 1525 – 6(15,5)^2 =83,5

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{15{,}625}{83{,}5}$$

= 1 – 0,187 = 0,813

b) $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

$$\sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 + \hat{\beta}_2 \sum X_i^3$$

$$\sum X_i^2 Y_i = \hat{\beta}_0 \sum X_i^2 + \hat{\beta}_1 \sum X_i^3 + \hat{\beta}_2 \sum X_i^4$$

$$-4{,}5 / 93 = 6\hat{\beta}_0 + 27\hat{\beta}_1 + 179\hat{\beta}_2$$

$$356 = 27\hat{\beta}_0 + 179\hat{\beta}_1 + 1443\hat{\beta}_2$$

$$-418{,}5 = -27\hat{\beta}_0 - 121{,}5\hat{\beta}_1 - 805{,}5\hat{\beta}_2$$

$$356 = 27\hat{\beta}_0 + 179\hat{\beta}_1 + 1443\hat{\beta}_2$$

$$-62{,}5 = 0 + 57{,}5\hat{\beta}_1 + 637{,}5\hat{\beta}_2$$

$$-179/93 = 6\hat{\beta}_0 + 27\hat{\beta}_1 + 179\hat{\beta}_2$$

$$6/2150 = 179\hat{\beta}_0 + 1443\hat{\beta}_1 + 12755\hat{\beta}_2$$

$$-16647 = -1074\hat{\beta}_0 - 4833\hat{\beta}_1 - 32041\hat{\beta}_2$$

$$12900 = 1074\hat{\beta}_0 + 8658\hat{\beta}_1 + 76530\hat{\beta}_2$$

$$-3747 = 0 + 3825\hat{\beta}_1 + 44489\hat{\beta}_2$$

$$-3825/-62{,}5 = 57{,}5\hat{\beta}_1 + 637{,}5\hat{\beta}_2$$

$$57{,}5/-3747 = 3825\hat{\beta}_1 + 44489\hat{\beta}_2$$

$$239062{,}5 = -219937{,}5\hat{\beta}_1 - 2438437{,}5\hat{\beta}_2$$

$$-215452{,}5 = 219937{,}5\hat{\beta}_1 + 2558117{,}5\hat{\beta}_2$$

$$23610 = 119680\hat{\beta}_2$$

$$\hat{\beta}_2 = \frac{23610}{119680} = 0{,}197$$

$$-62{,}5 = 57{,}5\hat{\beta}_1 + 637{,}5(0{,}197)$$

$$\hat{\beta}_1 = \frac{-62{,}5 - 125{,}5875}{57{,}5} = -3{,}271$$

$$93 = 6\,\hat{\beta}_0 + 27(-3{,}271) + 179(0{,}197)$$

$$\hat{\beta}_0 = \frac{93 + 88{,}317 - 35{,}263}{6} = 24{,}342$$

$$Y_i = 24{,}342 - 3{,}217\,X_i + 0{,}197\,X_i^2$$

$$\sum e_i^2 = \sum Y^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum X_i Y_i - \hat{\beta}_2 \sum X_i^2 Y_i$$

$$= 1525 - 24{,}342(93) - (-3{,}271)(356) - 0{,}197(2150)$$

$$= 2{,}12$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{2{,}12}{83{,}5} = 1 - 0{,}025 = 0{,}975$$

c) $Y_i = \hat{\alpha} + \hat{\beta} \log X_i$

$$\hat{\beta} = \frac{\sum Y_i (\log X_i) - n\bar{Y}(\overline{\log X})}{\sum (\log X)^2 - n(\overline{\log X})^2}$$

$\log X_i$	$(\log X_i)^2$	$Y_i(\log X_i)$
0,00000	0,00000	0,00000
0,30103	0,09061	5,41854
0,47712	0,22764	7,63392
0,60205	0,36246	8,42870
0,84509	0,71418	10,14108
1,00000	1,00000	11,00000
3,22529	2,39489	42,62224

$$(\overline{\log X}) = \frac{\sum \log X_i}{n} = \frac{3{,}22529}{6} = 0{,}53754$$

$$\beta = \frac{42{,}62224 - 6(15{,}5)(0{,}53754)}{2{,}39489 - 6(0{,}53754)^2}$$

$$= \frac{-7{,}36898}{0{,}66119} = -11{,}14503$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}(\overline{\log X})$$

$$= 15{,}5 - (-11{,}14503)(0{,}53754) = 21{,}49089$$

$Y_i = 21{,}49089 - 11{,}14503 \log X_i$

$$\sum e_i^2 = \sum Y_i^2 - \hat{\alpha} \sum Y_i - \hat{\beta} \sum Y_i (\log X_i)$$

$$= 1525 - 21{,}49089(93) - (-11{,}14503)(42{,}62224)$$

$$= 1{,}37337$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{1{,}37337}{83{,}5}$$

$= 1 - 0{,}016 = 0{,}984$

d) $\log X_i = \hat{\alpha} + \hat{\beta} X_i$

$$\beta = \frac{\sum X_i(\log Y_i) - n\bar{X}(\overline{\log Y})}{\sum X^2 - n(\bar{X})^2}$$

$\log Y_i$	$(\log Y_i)^2$	$X_i(\log Y_i)$
1,34242	1,80209	1,34242
1,25527	1,57570	2,51054
1,20412	1,44990	3,61236
1,14612	1,31360	4,58448
1,07918	1,16463	7,55426
1,04139	1,08449	10,41390
7,06850	8,39041	30,01796

$$(\overline{\log Y}) = \frac{\sum \log Y_i}{n} = \frac{7{,}06850}{6} = 1{,}17808$$

$$\hat{\beta} = \frac{30{,}01796 - 6(4{,}5)(1{,}17808)}{179 - 6(4{,}5)^2} = \frac{-1{,}7902}{57{,}5} = -0{,}03113$$

$$\hat{\alpha} = (\overline{\log Y}) - \hat{\beta}\bar{X}$$

$= 1{,}17808 - (-0{,}03113)(4{,}5) = 1{,}31816$

$\log Y_i = 1{,}31816 - 0{,}03113 X_i$

$$\sum e_i^2 = \sum (\log Y_i)^2 - \hat{\alpha}\sum \log Y_i - \hat{\beta}\sum X_i(\log Y_i)$$

$= 8{,}39041 - 1{,}31816(7{,}06850) - (-0{,}03113)(30{,}01796)$

$= 0{,}00745$

$$\sum [\log Y_i - (\overline{\log Y})]^2 = \sum (\log Y_i)^2 - n(\overline{\log Y})^2$$

$= 8{,}39041\text{-}6(1{,}17808)^2 = 0{,}06317$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\log Y - (\overline{\log Y})]^2} = 1 - \frac{0{,}00745}{0{,}06317}$$

$= 1 - 0{,}117 = 0{,}883$

e) $\log X_i = \log \hat{\alpha} + \hat{\beta} \log X_i$

$$\beta = \frac{\sum (\log X_i)(\log Y_i) - n(\overline{\log X})(\overline{\log Y})}{\sum (\log X)^2 - n(\overline{\log X})^2}$$

$(\log X_i)(\log Y_i)$
0,00000
0,37787
0,57450
0,60002
0,91200
1,04139
3,59578

$$\hat{\beta} = \frac{3{,}59578 - 6(0{,}53754)(1{,}17808)}{2{,}39489 - 6(0{,}53754)^2}$$

$$= \frac{-0{,}20381}{0{,}66119} = -0{,}30824$$

$$\log \hat{\alpha} = (\overline{\log Y}) - \hat{\beta}(\overline{\log X})$$

$$= 1{,}17808 - (-0{,}30824)(0{,}53754) = 1{,}34377$$

$$\log Y_i = 1{,}34377 - 0{,}30824(\log X_i)$$

$$\sum e_i^2 = \sum (\log Y_i)^2 - (\log \hat{\alpha}) \sum \log Y_i - \hat{\beta} \sum (\log X_i)(\log Y_i)$$

$$= 8{,}39041 - 1{,}34377(7{,}06850) - (-0{,}30824)(3{,}59578)$$

$$= 0{,}000335$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\log Y_i - (\overline{\log Y})]^2}$$

$$= 1 - \frac{0{,}000335}{0{,}06317}$$

$$= 1 - 0{,}0053 = 0{,}9947$$

BÖLÜM 14

ÇOKLU REGRESYON VE KORELASYON

Basit regresyonda iki değişken arasındaki ilişki matematiksel bir model ile belirleniyor ve bağımlı değişkendeki değişmeler bir tek bağımsız değişken tarafından açıklanıyordu. Uygulamada bir değişkendeki değişmelerin birden fazla değişken tarafından açıklanması, veya ikiden fazla değişken arasındaki ilişkinin derecesinin belirlenmesi gerekebilir. Regresyon modelinde birden fazla bağımsız değişken bulunması durumu çoklu regresyon ve ikiden fazla değişken arasındaki ilişkinin incelenmesi durumu çoklu korelasyon olarak adlandırılır. Örneğin gelir ile tüketim arasındaki ilişki söz konusu olduğunda basit regresyonda tüketiminin gelire bağlı olduğu kabul edilerek model kuruluyordu. Ancak tüketim gelirden başka, fiyat nüfus, gibi değişkenler tarafından etkilenebilir. Gelir dışındaki değişkenlerin de modelde bulunması isteniyorsa, çoklu regresyon modelinin oluşturulması gerekecektir. Aynı şekilde bu değişkenler arasındaki ilişkinin derecesi belirlenmek isteniyorsa çoklu korelasyon katsayısı hesaplanacaktır.

14.1. ÇOKLU REGRESYON

Regresyon modelinde bir bağımlı ve birden fazla bağımsız değişkenin bulunduğu anakütle çoklu regresyon modeli genel olarak

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

şeklinde gösterilebilir. Bu modelde Y_i, bağımlı yani etkilenen değişken, X_{i2}, X_{i3} .. X_{ik} ise bağımsız, yani etkileyen değişkenler e_i ise hata terimidir. Modelde k-1 sayıda bağımsız değişken, k sayıda değişken ve katsayı bulunmaktadır.

Anakütle regresyon modelinin parametreleri alınacak bir örnek ile,

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \cdots + \hat{\beta}_k X_{ik} + \varepsilon_i$$

olarak tahmin edilecektir.

Basit doğrusal regresyon ile ilgili açıklamalarda verilen temel varsayımlar çoklu regresyon için de aynen geçerlidir. Ayrıca çoklu regresyonun basit regresyondan farklı iki temel varsayımı daha vardır. Çoklu regresyonun basit regresyondan farklı varsayımları şunlardır.

-Çoklu doğrusal bağlılık olmaması: bağımsız değişkenler arasında tam doğrusal ilişki olmasına çoklu doğrusal bağlılık adı verilmektedir.

-Gözlem sayısının tahmin edilecek parametre sayısından fazla olması (n > k): Bu varsayım denklem sisteminin çözülmesi için gereklidir.

Çoklu regresyon ile ilgili olarak yapacağımız açıklamaların gerekliliği, basit regresyonun temel varsayımları ile birlikte yukarıda verilen varsayımların da geçerliliğine bağlıdır.

14.1.1. Parametrelerin Tahmini

Çoklu regresyon modelinin parametrelerini tahmin etmek için kullanılacak normal denklemler, basit regresyonda olduğu gibi en küçük kareler yöntemi ile belirlenebilir. Bunun için gerçek ve teorik Y_i değerleri arasındaki farkların kareleri toplamının,

$$T = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

minimize edilmesi gerekecektir. Çoklu regresyonda anakütle teorik değerleri

$$\hat{Y}_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

olduğundan,

$$T = \sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_k X_{ik})^2$$

olarak yazılabilir. T'nin $\beta_1, \beta_2, \cdots \beta_k$ 'ya göre kısmi türevleri alınarak, sıfıra eşitlenirse,

$$\frac{\partial T}{\partial \beta_1} = 2\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_k X_{ik})(-1) = 0$$

$$\frac{\partial T}{\partial \beta_2} = 2\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_k X_{ik})(-X_{i2}) = 0$$

$$\frac{\partial T}{\partial \beta_3} = 2\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_k X_{ik})(-X_{i3}) = 0$$

$$\vdots$$

$$\vdots$$

$$\frac{\partial T}{\partial \beta_k} = 2\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3} - \cdots - \beta_k X_{ik})(-X_{ik}) = 0$$

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{i2} - \hat{\beta}_3 X_{i3} - \cdots - \hat{\beta}_k X_{ik}) = 0$$

$$\sum_{i=1}^{n}(Y_i X_{i2} - \hat{\beta}_1 X_{i2} - \hat{\beta}_2 X_{i2}^{\,2} - \hat{\beta}_3 X_{i2} X_{i3} - \cdots - \hat{\beta}_k X_{i2} X_{ik}) = 0$$

$$\sum_{i=1}^{n}(Y_i X_{i3} - \hat{\beta}_1 X_{i3} - \hat{\beta}_2 X_{i2} X_{i3} - \hat{\beta}_3 X_{i3}^{\,2} - \cdots - \hat{\beta}_k X_{i3} X_{ik}) = 0$$

$$\vdots$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i X_{ik} - \hat{\beta}_1 X_{ik} - \hat{\beta}_2 X_{i2} X_{ik} - \hat{\beta}_3 X_{i3} X_{ik} - \cdots - \hat{\beta}_k X_{ik}^{\,2}) = 0$$

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} X_{i3} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ik}$$

$$\sum_{i=1}^{n} X_{i2} Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{i2} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}^{\,2} + \hat{\beta}_3 \sum_{i=1}^{n} X_{i2} X_{i3} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{12} X_{ik}$$

$$\sum_{i=1}^{n} X_{i3} Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{i3} + \hat{\beta}_2 \sum_{i=1}^{n} X_{i3} X_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} X_{i3}^{\,2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{13} X_{ik}$$

$$\vdots$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ik} Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{ik} + \hat{\beta}_2 \sum_{i=1}^{n} X_{ik} X_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} X_{ik} X_{i3} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ik}^{\,2}$$

olarak normal denklemler elde edilir.

ÖRNEK: Aşağıda verilen $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$ modelinin parametrelerini tahmin ediniz.

Y_i	X_{i2}	X_{i3}	$X_{i2}Y_i$	$X_{i3}Y_i$	$X_{i2}X_{i3}$	X_{i2}^2	X_{i3}^2	Y_i^2
10	120	4	1200	40	480	14400	16	100
13	135	8	1755	104	1080	18225	64	169
19	132	5	2508	95	660	17424	25	361
18	140	8	2520	144	1120	19600	64	324
15	153	10	2295	150	1530	23409	100	225
75	680	35	10278	533	4870	93058	269	1179

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{75}{5} = 15$$

$$\bar{X}_2 = \frac{\sum X_{i2}}{n} = \frac{680}{5} = 136$$

$$\bar{X}_3 = \frac{\sum X_{i3}}{n} = \frac{35}{5} = 7$$

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{i2} + \hat{\beta}_3 \sum X_{i3}$$

$$\sum X_{i2}Y_i = \hat{\beta}_1 \sum X_{i2} + \hat{\beta}_2 \sum X_{i2}^2 + \hat{\beta}_3 \sum X_{i2}X_{i3}$$

$$\sum X_{i3}Y_i = \hat{\beta}_1 \sum X_{i3} + \hat{\beta}_2 \sum X_{i3}X_{i2} + \hat{\beta}_3 \sum X_{i3}^2$$

$$-136/75 = 5\hat{\beta}_1 + 680\hat{\beta}_2 + 35\hat{\beta}_3$$

$$10278 = 680\hat{\beta}_1 + 93058\hat{\beta}_2 + 4870\hat{\beta}_3$$

$$-10200 = -680\hat{\beta}_1 - 92480\hat{\beta}_2 - 4760\hat{\beta}_3$$

$$10278 = 680\hat{\beta}_1 + 93058\hat{\beta}_2 + 4870\hat{\beta}_3$$

$$78 = 578\hat{\beta}_2 + 110\hat{\beta}_3$$

$$-7/75 = 5\hat{\beta}_1 + 680\hat{\beta}_2 + 35\hat{\beta}_3$$

$$533 = 35\hat{\beta}_1 + 4870\hat{\beta}_2 + 269\hat{\beta}_3$$

$$8 = 110\hat{\beta}_2 + 24\hat{\beta}_3$$

$$78 = 578\hat{\beta}_2 + 110\hat{\beta}_3$$
$$8 = 110\hat{\beta}_2 + 24\hat{\beta}_3$$

$$\hat{\beta}_1 = -45{,}55$$
$$\hat{\beta}_2 = 0{,}56$$
$$\hat{\beta}_3 = -2{,}23$$

$Y_i = -45{,}55 + 0{,}56X_{i2} - 2{,}23X_{i3}$

Normal denklemlerde gerçek değerler yerine ortalamalardan farklar,

$y_i = Y_i - \overline{Y}$

$x_{i2} = X_{i2} - \overline{X}_2$

$x_{i3} = X_{i3} - \overline{X}_3$

$\vdots$

$x_{ik} = X_{ik} - \overline{X}_k$

olarak yerine konulursa normal denklemler,

$$\sum_{i=1}^{n} x_{i2}y_i = \hat{\beta}_2 \sum_{i=1}^{n} {x_{i2}}^2 + \hat{\beta}_3 \sum_{i=1}^{n} x_{i2}x_{i3} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{12}x_{ik}$$

$$\sum_{i=1}^{n} x_{i3}y_i = \hat{\beta}_2 \sum_{i=1}^{n} x_{i3}x_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} {x_{i3}}^2 + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{13}x_{ik}$$

$\vdots$

$\vdots$

$$\sum_{i=1}^{n} x_{ik}y_i = \hat{\beta}_1 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_2 \sum_{i=1}^{n} x_{ik}x_{i2} + \hat{\beta}_3 \sum_{i=1}^{n} x_{ik}x_{i3} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} {x_{ik}}^2$$

olacaktır. Görüldüğü gibi ortalamadan farkların toplamı sıfır olduğundan denklem sayısı ve tahmin edilecek parametre sayısı azalmıştır. Bu durumda,

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_2 - \hat{\beta}_3\overline{X}_3 - \cdots - \hat{\beta}_k\overline{X}_k$$

olarak belirlenir.

ÖRNEK: Daha önce verilen seri için $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$ modelinin parametrelerini ortalamadan farklar ile tahmin ediniz.

Y_i	X_{i2}	X_{i3}	y_i	x_{i2}	x_{i3}	$x_{i2}y_i$	$X_{i3}y_i$	$x_{i2}x_{i3}$	x_{i2}^2	x_{i3}^2	y_i^2
10	120	4	-5	-16	-3	80	15	48	256	9	25
13	135	8	-2	-1	1	2	-2	-1	1	1	4
19	132	5	4	-4	-2	-16	-8	8	16	4	16
18	140	8	3	4	1	12	3	4	16	1	9
15	153	10	0	17	3	0	0	51	289	9	0
75	680	35	0	0	0	78	8	110	578	24	54

$\overline{Y} = 15$

$\overline{X}_2 = 136$

$\overline{X}_3 = 7$

$$\sum x_{i2}y_i = \hat{\beta}_2 \sum x_{i2}^2 + \hat{\beta}_3 \sum x_{i2}x_{i3}$$

$$\sum x_{i3}y_i = \hat{\beta}_2 \sum x_{i3}x_{i2} + \hat{\beta}_3 \sum x_{i3}^2$$

$$78 = 578\hat{\beta}_2 + 110\hat{\beta}_3$$

$$8 = 110\hat{\beta}_2 + 24\hat{\beta}_3$$

$\hat{\beta}_2 = 0{,}56$

$\hat{\beta}_3 = -2{,}23$

$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_2 - \hat{\beta}_3\overline{X}_3$

$= 15 - 0{,}56(136) - (-2{,}23)(7) = -45{,}5$

$Y_i = -45{,}55 + 0{,}56X_{i2} - 2{,}23X_{i3}$

Çoklu regresyon modellerinin parametreleri matrisler ile de tahmin edilebilir. Daha önce verilen, parametrelerin gerçek değerler ile tahmin edildiği normal denklemler aşağıda görüldüğü gibi matrisler ile ifade edilebilir.

$$\begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i2}Y_i \\ \sum_{i=1}^{n} X_{i3}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ik}Y_i \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \begin{bmatrix} n & \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i3} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \sum_{i=1}^{n} X_{i2}X_{i3} & \cdots & \sum_{i=1}^{n} X_{i2}X_{ik} \\ \sum_{i=1}^{n} X_{i3} & \sum_{i=1}^{n} X_{i2}X_{i3} & \sum_{i=1}^{n} X_{i3}^2 & \cdots & \sum_{i=1}^{n} X_{i3}X_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \sum_{i=1}^{n} X_{i3}X_{ik} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix}$$

Sistemde yeralan bilinmeyen β katsayıları, katsayılar matrisinde yeraldığından bu matris yalnız bırakılırsa,

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i3} & \cdots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \sum_{i=1}^{n} X_{i2}X_{i3} & \cdots & \sum_{i=1}^{n} X_{i2}X_{ik} \\ \sum_{i=1}^{n} X_{i3} & \sum_{i=1}^{n} X_{i2}X_{i3} & \sum_{i=1}^{n} X_{i3}^2 & \cdots & \sum_{i=1}^{n} X_{i3}X_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \sum_{i=1}^{n} X_{i3}X_{ik} & \cdots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i2}Y_i \\ \sum_{i=1}^{n} X_{i3}Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ik}Y_i \end{bmatrix}$$

olacaktır. Sistemin çözümü ters matris alınmasını gerektirecektir.

ÖRNEK: Daha önce verilen seri için $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$ modelinin parametrelerini matrisler ile tahmin ediniz.

$\sum Y_i = 75$ $\quad\quad$ $\sum X_{i2}Y_i = 10278$ $\quad\quad$ $\sum X_{i2}^2 = 93058$

$\sum X_{i2} = 680$ $\quad\quad$ $\sum X_{i3}Y_i = 533$ $\quad\quad$ $\sum X_{i3}^2 = 269$

$\sum X_{i3} = 35$ $\quad\quad$ $\sum X_{i3}X_{i2} = 4870$ $\quad\quad$ n = 5

olarak hesaplanmıştı.

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum X_{i2}Y_i \\ \sum X_{i3}Y_i \end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}5 & 680 & 35\\ 680 & 93058 & 4870\\ 35 & 4870 & 269\end{bmatrix}^{-1}\begin{bmatrix}75\\ 10278\\ 533\end{bmatrix}$$

Sistemin çözülmesi için önce ters matrisin alınması gerekecektir. Tersi alınacak matrisi (X′X) ile gösterirsek,

$$(X'X)^{-1}=\frac{1}{D}(X'X)^{*}$$

Burada D matrisinin determinantını $(X'X)^{*}$ ise matrisin devriğini ifade etmektedir.

D = [(5)(93058)(269) + (680)(4870)(35) + (35)(680)(4870)] – [(35)(93058)(35) + (4870)(4870)(5) + (269)(680)(680)]
= 8860

$$(X'X)^{*}=\begin{bmatrix}\begin{vmatrix}93058 & 4870\\ 4870 & 269\end{vmatrix} & \begin{vmatrix}680 & 4870\\ 35 & 269\end{vmatrix} & \begin{vmatrix}680 & 93058\\ 35 & 4870\end{vmatrix}\\ \begin{vmatrix}680 & 35\\ 4870 & 269\end{vmatrix} & \begin{vmatrix}5 & 35\\ 35 & 269\end{vmatrix} & \begin{vmatrix}5 & 680\\ 35 & 4870\end{vmatrix}\\ \begin{vmatrix}680 & 35\\ 93058 & 4870\end{vmatrix} & \begin{vmatrix}5 & 35\\ 680 & 4870\end{vmatrix} & \begin{vmatrix}5 & 680\\ 680 & 93058\end{vmatrix}\end{bmatrix}$$

$$=\begin{bmatrix}\overset{+}{1315702} & \overset{-}{12470} & \overset{+}{54570}\\ \overset{-}{12470} & \overset{+}{120} & \overset{-}{550}\\ \overset{+}{54570} & \overset{-}{550} & \overset{+}{2890}\end{bmatrix}$$

$$(X'X)^{-1}=\frac{1}{8860}\begin{bmatrix}1315702 & -12470 & 54570\\ -12470 & 120 & -550\\ 54570 & -550 & 2890\end{bmatrix}$$

$$=\begin{bmatrix}148{,}499097 & -1{,}407449209 & 6{,}159142212\\ -1{,}407449209 & 0{,}013544018 & -0{,}062076749\\ 6{,}159142212 & -0{,}062076749 & 0{,}326185101\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}148{,}499097 & -1{,}407449209 & 6{,}159142212\\ -1{,}407449209 & 0{,}013544018 & -0{,}062076749\\ 6{,}159142212 & -0{,}062076749 & 0{,}326185101\end{bmatrix}\begin{bmatrix}75\\ 10278\\ 533\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}11137,43227 & -14465,76297 & +3282,822798\\ -105,5586906 & +139,205417 & -33,08690721\\ 461,9356659 & -638,0248262 & +173,8566588\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}-45,507902\\ 0,55981919\\ -2,2325015\end{bmatrix}$$

Çoklu regresyon modellerinin parametrelerinin tahmininde kullanılan ortalamadan farklar ile elde edilen denklemler de matrisler ile gösterilebilir.

$$\begin{bmatrix}\sum_{i=1}^{n} x_{i2}y_i\\ \sum_{i=1}^{n} x_{i3}y_i\\ \vdots\\ \sum_{i=1}^{n} x_{ik}y_i\end{bmatrix}=\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\\ \vdots\\ \vdots\\ \vdots\\ \hat{\beta}_k\end{bmatrix}\begin{bmatrix}\sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2}x_{i3} & \cdots & \sum_{i=1}^{n} x_{i2}x_{ik}\\ \sum_{i=1}^{n} x_{i2}x_{i3} & \sum_{i=1}^{n} x_{i3}^2 & \cdots & \sum_{i=1}^{n} x_{i3}x_{ik}\\ \vdots & \vdots & & \vdots\\ \sum_{i=1}^{n} x_{i2}x_{ik} & \sum_{i=1}^{n} x_{i3}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2\end{bmatrix}$$

Katsayıların yer aldığı matris yalnız bırakılırsa,

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\\ \vdots\\ \vdots\\ \vdots\\ \hat{\beta}_k\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2}x_{i3} & \cdots & \sum_{i=1}^{n} x_{i2}x_{ik}\\ \sum_{i=1}^{n} x_{i2}x_{i3} & \sum_{i=1}^{n} x_{i3}^2 & \cdots & \sum_{i=1}^{n} x_{i3}x_{ik}\\ \vdots & \vdots & & \vdots\\ \sum_{i=1}^{n} x_{i2}x_{ik} & \sum_{i=1}^{n} x_{i3}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^{n} x_{i2}y_i\\ \sum_{i=1}^{n} x_{i3}y_i\\ \vdots\\ \sum_{i=1}^{n} x_{ik}y_i\end{bmatrix}$$

olacaktır. Bu durumda sabit katsayı

$$\hat{\beta}_1=\overline{Y}-\hat{\beta}_2\overline{X}_2-\hat{\beta}_3\overline{X}_3-\cdots-\hat{\beta}_k\overline{X}_k$$

olarak belirlenir.

ÖRNEK: Daha önce örnek olarak verilen seri için $Y_i=\hat{\beta}_1+\hat{\beta}_2X_{i2}+\hat{\beta}_3X_{i3}$ çoklu regresyon modelinin parametrelerini ortalamadan farklar ile ifade edilen matrislerle tahmin ediniz.

Daha önce,

$\bar{Y} = 15$ $\quad \sum x_{i2}y_i = 78$ $\quad \sum x_{i2}^2 = 578$

$\bar{X}_2 = 136$ $\quad \sum x_{i3}y_i = 8$ $\quad \sum x_{i3}^2 = 24$

$\bar{X}_3 = 7$ $\quad \sum x_{i2}x_{i3} = 110$

olarak hesaplanmıştı.

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}\sum x_{i2}^2 & \sum x_{i2}x_{i3}\\ \sum x_{i2}x_{i3} & \sum x_{i3}^2\end{bmatrix}^{-1}\begin{bmatrix}\sum x_{i2}y_i\\ \sum x_{i3}y_i\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}578 & 110\\ 110 & 24\end{bmatrix}^{-1}\begin{bmatrix}78\\ 8\end{bmatrix}$$

Tersi alınacak matrisi $(\underline{X}'\underline{X})$ ile gösterirsek

$$(\underline{X}'\underline{X})^{-1} = \frac{1}{D}(\underline{X}'\underline{X})^*$$

olacaktır.

D = (578)(24) – (110)(10) = 1772

$$(\underline{X}'\underline{X})^* = \begin{bmatrix}\overset{+}{24} & \overset{-}{110}\\ \overset{-}{110} & \overset{+}{578}\end{bmatrix}$$

$$(\underline{X}'\underline{X})^{-1} = \frac{1}{1772}\begin{bmatrix}24 & -110\\ -110 & 578\end{bmatrix}$$

$$(\underline{X}'\underline{X})^{-1} = \begin{bmatrix}0{,}013544018 & -0{,}062076749\\ -0{,}062076749 & 0{,}326185101\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}0{,}013544018 & -0{,}062076749\\ -0{,}062076749 & 0{,}326185101\end{bmatrix}\begin{bmatrix}78\\ 8\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}1{,}056433404 & -0{,}496613992\\ -4{,}841986422 & 2{,}609480808\end{bmatrix} = \begin{bmatrix}0{,}559819412\\ -2{,}232505614\end{bmatrix}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3$$

= 15 – (0,559819412)(136) – (-2,232505614)(7)
= -45,50790074

14.1.2. Parametre Tahmincilerinin Varyansları

Daha önce basit regresyonda açıklandığı gibi, çoklu regresyonda da parametrelerin varyanslarını hesaplayabilmek için, hata terimlerinin varyansının örnekten tahmin edilmesi gerekecektir. Hata terimlerinin varyansının sapmasız tahmincisi,

$$S_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-k} = \frac{\sum e_i^2}{n-k}$$

olduğundan, hata terimlerinin varyansını tahmin edebilmek için önce hata terimlerinin karelerinin toplamı hesaplanacaktır.

Hata terimlerinin kareleri toplamı teorik değerler ile seriden hesaplanabilir. Teorik değerlerin bulunabilmesi için basit regresyonda olduğu gibi, bağımsız değişkenlerin değerlerinin regresyon modelinde yerine konması gerekecektir.

Çoklu regresyonda hata terimlerinin kareleri toplamı gerçek değerlerle,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\beta}_1 \sum_{i=1}^{n} Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_{i2} Y_i - \hat{\beta}_3 \sum_{i=1}^{n} X_{i3} Y_i - \cdots - \hat{\beta}_k \sum_{i=1}^{n} X_{ik} Y_i$$

veya ortalamadan farklarla,

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} y_i^2 - \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} y_i - \hat{\beta}_3 \sum_{i=1}^{n} x_{i3} y_i - \cdots - \hat{\beta}_k \sum_{i=1}^{n} x_{ik} y_i$$

formülleri ile de hesaplanabilir.

Çoklu regresyonda parametrelerin varyansları varyans-kovaryans matrisinden tahmin edilebilir. Daha önce verilen $(X'X)^{-1}$ matrisinin S_e^2 ile çarpımı varyansları ve kovaryansları verecektir.

$$\begin{bmatrix} Var(\hat{\beta}_1) & Kov(\hat{\beta}_1\hat{\beta}_2) & \cdots & Kov(\hat{\beta}_1\hat{\beta}_k) \\ Kov(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_2) & \cdots & Kov(\hat{\beta}_2\hat{\beta}_k) \\ \vdots & \vdots & & \vdots \\ Kov(\hat{\beta}_1\hat{\beta}_k) & Kov(\hat{\beta}_2\hat{\beta}_k) & \cdots & Var(\hat{\beta}_3) \end{bmatrix} = S_e^2 (X'X)^{-1}$$

$$= S_e^2 \begin{bmatrix} n & \sum X_{i2} & \cdots & \sum X_{ik} \\ \sum X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{ik} \\ \vdots & \vdots & & \vdots \\ \sum X_{ik} & \sum X_{i2}X_{ik} & \cdots & \sum X_{ik}^2 \end{bmatrix}^{-1}$$

ÖRNEK: Daha önce verilen örneğin parametrelerinin varyanslarını hesaplayınız.

$$S_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-k} = \frac{\sum e_i^2}{n-k}$$

$Y_i = -45{,}51 + 0{,}56\ X_{i2} - 2{,}23\ X_{i3}$

$\hat{Y}_1 = -45{,}51 + 0{,}56(120) - 2{,}23(4) = 12{,}77$

$\hat{Y}_2 = -45{,}51 + 0{,}56(135) - 2{,}23(8) = 12{,}25$

$\hat{Y}_3 = -45{,}51 + 0{,}56(132) - 2{,}23(5) = 17{,}26$

$\hat{Y}_4 = -45{,}51 + 0{,}56(140) - 2{,}23(8) = 15{,}05$

$\hat{Y}_5 = -45{,}51 + 0{,}56(153) - 2{,}23(10) = 17{,}87$

Y_i	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
10	-2,777	7,6729
13	0,75	0,5625
19	1,74	3,0276
18	2,95	8,7025
15	-2,87	8,2369
		28,2024

veya

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} Y_i^2 - \hat{\beta}_1 \sum_{i=1}^{n} Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_{i2}Y_i - \hat{\beta}_3 \sum_{i=1}^{n} X_{i3}Y_i$$

$$= 1179 - (-45{,}55)(75) - (0{,}56)(10278) - (-2{,}23)(533)$$

$$= 28{,}16$$

$$S_e^2 = \frac{28{,}2024}{5-3} = 14{,}1012$$

$$\begin{bmatrix} Var(\hat{\beta}_1) & Kov(\hat{\beta}_1\hat{\beta}_2) & Kov(\hat{\beta}_1\hat{\beta}_3) \\ Kov(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_2) & Kov(\hat{\beta}_2\hat{\beta}_3) \\ Kov(\hat{\beta}_1\hat{\beta}_3) & Kov(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_3) \end{bmatrix} = S_e^2 \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i3}X_{i2} & \sum X_{i3}^2 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} Var(\hat{\beta}_1) & & \\ & Var(\hat{\beta}_2) & \\ & & Var(\hat{\beta}_3) \end{bmatrix} = 14{,}1012 \begin{bmatrix} 148{,}499097 & & \\ & 0{,}13544018 & \\ & & 0{,}326185101 \end{bmatrix}$$

$$S_{\hat{\beta}_1}^2 = Var\hat{\beta}_1 = 2094{,}015466$$

$$S_{\hat{\beta}_2}^2 = Var\hat{\beta}_2 = 0{,}190986906$$

$$S_{\hat{\beta}_3}^2 = Var\hat{\beta}_3 = 4{,}599600063$$

$(X'X)^{-1}$ matrisinin köşegeni varyansları verdiğinden, yukarıda daha önce hesaplanan matrisin ilgili değerleri verilmiştir.

Ortalamadan farklar matrisleri söz konusu olduğunda varyans - kovaryans matrisi

$$\begin{bmatrix} Var(\hat{\beta}_2) & Kov(\hat{\beta}_2\hat{\beta}_3) & \cdots & Kov(\hat{\beta}_1\hat{\beta}_k) \\ Kov(\hat{\beta}_2\hat{\beta}_3) & Var(\hat{\beta}_3) & \cdots & Kov(\hat{\beta}_2\hat{\beta}_k) \\ \vdots & \vdots & & \vdots \\ Kov(\hat{\beta}_2\hat{\beta}_k) & Kov(\hat{\beta}_3\hat{\beta}_k) & \cdots & Var(\hat{\beta}_k) \end{bmatrix} = S_e^2 (\underline{X}' \underline{X})^{-1}$$

$$= S_e^2 \begin{bmatrix} \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \cdots & \sum x_{i2}x_{ik} \\ \sum x_{i2}x_{i3} & \sum x_{i3}^2 & \cdots & \sum x_{i3}X_{ik} \\ \vdots & \vdots & & \vdots \\ \sum x_{i2}x_{ik} & \sum x_{i3}x_{ik} & \cdots & \sum x_{ik}^2 \end{bmatrix}^{-1}$$

olacaktır. Görüldüğü gibi bu durumda sabit katsayının varyansı belirlenememektedir.

ÖRNEK: Daha önce verilen örneğin parametrelerinin varyanslarını ortalamadan farklar serileri ile hesaplayınız.

Daha önce,

S_e^2 = 14, 1012 ve $(\underline{X}' \underline{X})^{-1}$ matrisi,

$$(\underline{X}'\,\underline{X})^{-1} = \begin{bmatrix} 0{,}013544018 & -0{,}062076749 \\ -0{,}062076749 & 0{,}326185101 \end{bmatrix}$$

olarak belirlenmiştir. Bu durumda bağımsız değişkenlerin katsayılarının varyansları,

$$\begin{bmatrix} Var(\hat{\beta}_2) & Kov(\hat{\beta}_2,\hat{\beta}_3) \\ Kov(\hat{\beta}_2,\hat{\beta}_3) & Var(\hat{\beta}_3) \end{bmatrix} = S_e^2(\underline{X}'\,\underline{X})^{-1}$$

$$\begin{bmatrix} Var(\hat{\beta}_2) & Kov(\hat{\beta}_2,\hat{\beta}_3) \\ Kov(\hat{\beta}_2,\hat{\beta}_3) & Var(\hat{\beta}_3) \end{bmatrix} = S_e^2\begin{bmatrix} \sum x_{i2}^2 & \sum x_{i2}x_{i3} \\ \sum x_{i2}x_{i3} & \sum x_{i3}^2 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} Var(\hat{\beta}_2) & \\ & Var(\hat{\beta}_3) \end{bmatrix} = 14{,}1012\begin{bmatrix} 0{,}13544018 & \\ & 0{,}326185101 \end{bmatrix}$$

$$Var\hat{\beta}_2 = 0{,}190986906$$

$$Var\hat{\beta}_3 = 4{,}599600063$$

olarak elde edilir.

14.1.3. Belirlilik Katsayısı

Çoklu regresyonda belirlilik katsayısı bağımsız değişkenlerin bağımlı değişkenlerdeki değişmeleri açıklama oranını verecek ve basit regresyonda olduğu gibi açıklanan değişmelerin, toplam değişmeye oranı ile,

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

veya açıklanmayan değişmenin toplam değişmeye oranının 1'den çıkartılması ile,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

hesaplanacaktır. $\hat{Y}_i$ değerleri veya $\sum e_i^2$ değerleri daha önce parametrelerin varyansları bölümünde verilen formüllerle hesaplanarak R^2 bulunabilir. Ayrıca açıklanan değişmenin hesaplanabilmesi için,

$$\sum(\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_2 \sum_{i=1}^{n}(Y_i - \overline{Y})(X_{i2} - \overline{X}_2) + \hat{\beta}_3 \sum_{i=1}^{n}(Y_i - \overline{Y})(X_{i3} - \overline{X}_3)$$

$$+ \cdots + \hat{\beta}_k \sum_{i=1}^{n}(Y_i - \overline{Y})(X_{ik} - \overline{X}_k)$$

formülünden de yararlanılabilir.

Çoklu regresyonda modele eklenecek her yeni bağımsız değişkenin, açıklanamayan değişmelerin bir bölümünü daha açıklaması beklenecektir. Aynı bağımlı değişken için bağımsız değişken sayıları ile farklı modeller oluşturulabilir.

Bu farklı modelleri karşılaştırmak ve ilişkiyi daha iyi açıklayan modeli belirlemek için düzeltilmiş belirlilik katsayıları hesaplanacaktır.

Düzeltilmiş belirlilik katsayısı $\overline{R}^2$ ile ifade edilebilir ve bu katsayının hesaplanabilmesi için farklı formüller kullanılabilir. Bu formüllerden ikisi,

$$\overline{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2)$$

$$\overline{R}^2 = R^2 - (1-R^2)\left(\frac{n-1}{n-k}\right)$$

olarak gösterilebilir. Formüllerden görüldüğü gibi $\overline{R}^2$ hesaplanırken n ve k değerleri dikkate alınmaktadır. Düzeltilmiş belirlilik katsayısı hesaplandığında $\overline{R}^2 \leq R^2$ olacaktır.

ÖRNEK: Daha önce verilen örnek için belirlilik katsayısını ve düzeltilmiş belirlilik katsayısını hesaplayınız.

Belirlilik katsayısı daha önce hesaplanan teorik değerlerden yararlanılarak aşağıda görüldüğü gibi hesaplanabilir.

Y_i	$\hat{Y}_i$	$\hat{Y}_i - \overline{Y}$	$(\hat{Y}_i - \overline{Y})^2$
10	12,77	-2,23	4,9729
13	12,25	-2,75	7,5625
19	17,26	2,26	5,1076
18	15,05	0,05	0,0025
15	17,87	2,87	8,2369
			25,8824

$$\bar{Y} = 15$$

$$\sum (Y_i - \bar{Y})^2 = 54$$

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$= \frac{25{,}8824}{54} = 0{,}4793$$

Belirlilik katsayısı daha önce hesaplanan hata terimlerinin karelerinin toplamından yararlanılarak da hesaplanabilir.

$$\sum e_i^2 = \sum Y_i^2 - \hat{\beta}_1 \sum Y_i - \hat{\beta}_2 \sum X_{i2} Y_i - \hat{\beta}_3 \sum X_{i3} Y_i$$
$$= 28{,}16$$

olarak hesaplanmıştı.

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$= 1 - \frac{28{,}16}{54}$$
$$= 1 - 0{,}5214 = 0{,}4786$$

$R^2 = 0{,}47$ olarak kabul edildiğinde düzeltilmiş belirlilik katsayısı,

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$$

$$= 0{,}47 - \left(\frac{3-1}{5-3}\right)(1 - 0{,}47) = -0{,}06$$

veya

$$\bar{R}^2 = R^2 - (1 - R^2)\left(\frac{n-1}{n-k}\right)$$
$$= 1 - (1 - 0{,}47)\left(\frac{5-1}{5-3}\right) = -0{,}06$$

olarak bulunur.

14.1.4. Katsayıların Testi

Basit regresyonda olduğu gibi çoklu regresyonda da katsayıların istatistiksel olarak anlamlı olup olmadığı test edilebilir. Çoklu regresyonda basit regresyonun aksine F testi de anlam kazanmaktadır.

14.1.4.1. t -Testi

Sabit katsayı ve tüm bağımsız değişkenlerin katsayıları ayrı ayrı t testi ile test edilerek istatistiksel olarak anlamlı olup olmadıkları belirlenebilir. Çoklu regresyonda da t testi basit regresyonda açıklandığı gibi yapılacaktır.

Test edilecek katsayı β_i ise temel hipotez,

$H_0 : \beta_i = 0$

karşıt hipotez ise çift taraflı test için,

$H_1 : \beta_i \neq 0$

Tek taraflı test için

$H_1 : \beta_i > 0$

veya

$H_1 : \beta_i < 0$

olacaktır. tablolardan kritik değerler α hata payı ile belirlenir. t dağılımının serbestlik derecesi (n-k)'dır. Test istatistiği,

$$t = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$

temel hipotez $\beta_i = 0$ olduğundan,

$$t = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$$

olacaktır. Hipotezlerden birinin kabul kararı aynı şekilde verilecektir.

14.1.4.2. F Testi

Basit regresyonda söz edildiği gibi F testi ile bağımsız değişkenlerin katsayılarının tümü birlikte test edilir. Temel hipotez bağımsız değişkenlerin katsayılarının anlamsız olduğunu ifade eder. Modelin sabit katsayısı β_1 ise temel hipotez

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k$$

olacaktır. Bu hipotezin kabulü tüm bağımsız değişkenlerin istatistiksel olarak anlamsız olduğu anlamına gelmemektedir. Bu hipotez bağımsız değişken katsayılarından en az birinin istatistikel olarak anlamsız olduğunu ifade etmektedir.

F testinde tek taraflı test yapılamaz. Karşıt hipotez,

$$H_1 : \beta_2 \neq \beta_3 \neq \cdots \neq \beta_k$$

olarak H_0 hipotezinin doğru olmadığını ifade edecek şekilde yazılabilir. Buna göre bütün bağımsız değişken katsayıları anlamlıdır.

Çoklu regresyon modelinde tahmin edilen katsayı sayısı k ise F test istatistiği

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / (k-1)}{\sum (Y_i - \hat{Y}_i) / (n-k)}$$

olacaktır. Test istatistiği belirlilik katsayısı ile,

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$$

olarak da hesaplanabilir.

F tablo değeri belirlenecek α hata payı ile tablodan bulunacaktır. Serbestlik dereceleri S_{D1}=k-1 ve S_{D2} =n-k'dır.

Hesaplanan F değeri tablo değerinde büyük ise H_1 hipotezi, küçük ise H_0 hipotezi kabul edilecektir.

F ve t testleri farklı sonuç verebilirler. Bu temel varsayımlardan sapma olmasının göstergesidir. Çoklu doğrusal bağıntı olması durumunda t ve F testleri farklı sonuç verebilmektedirler.

14.1.5. Çoklu Regresyon Modelinin Tahminde Kullanılması

Basit regresyonda olduğu gibi çoklu regresyonda da gelecek devrelerin değerleri tahmin edilebilir. Bu durumda modelde yer alan bağımsız değişkenlerin ilgili devreye ait değerlerinin tahmin edilmesi gerekecektir. Tahmin edilen bağımsız değişkenlere ait değerlerin modelde yerine konması ile, ilgili devrenin bağımlı değişken değeri tahmin edilmiş olacaktır. Tahmin edilecek bağımlı değişken değerini $\hat{Y}_0$ ile ve aynı devreye ait bağımsız değişkenlerin değerlerini,

$X_{0,2}$; $X_{0,3}$; .... ; $X_{0,k}$

ile ifade edersek,

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_{0,2} + \hat{\beta}_3 X_{0,3} + \cdots + \hat{\beta}_k X_{0,k}$$

olacaktır.

Tahminin standart hatası matrisler ile,

$$S_T^2 = S_e^2 \left[1 + \frac{1}{n} + (X_i - \overline{X})' (\underline{X}' \underline{X})^{-1} (X_i - \overline{X}) \right]$$

olarak gösterebilir. Burada $(X_i - \overline{X})$ matrisi,

$$(X_i - \overline{X}) = \begin{bmatrix} X_{0,2} - \overline{X}_2 \\ X_{0,3} - \overline{X}_3 \\ \vdots \\ X_{0,k} - \overline{X}_k \end{bmatrix}$$

dır ve bağımsız değişkenlerin tahmin edilen değerleri ile ortalamaları arasındaki farklardan oluşmaktadır.

ÖRNEK: Daha önce verilen örnek için $X_{6,2}$ = 160 ve $X_{6,3}$ = 12 olarak tahmin edildiğine göre, 6. devrenin bağımlı değişken değerini $(\hat{Y}_6)$ tahmin ediniz ve tahminin standart hatasını hesaplayınız.

Y_i = -45,51 + 0,56 X_{i2} – 2,23 X_{i3}

$\hat{Y}_6$ = -45,51 + 0,56(160) – 2,23(12) = 17,33

$\overline{X}_2$ = 136 ve $\overline{X}_3$ = 7 olarak bilindiğinden,

$$(X_i - \overline{X}) = \begin{bmatrix} X_{0,2} - \overline{X}_2 \\ X_{0,3} - \overline{X}_3 \end{bmatrix} = \begin{bmatrix} 160 - 136 \\ 12 - 7 \end{bmatrix} = \begin{bmatrix} 24 \\ 5 \end{bmatrix}$$

olacaktır. $(\underline{X}' \underline{X})^{-1}$ matrisi de daha önce belirlendiğinden,

$$S_T^2 = 14{,}1012 \left\{ 1 + \frac{1}{5} + \begin{bmatrix} 24 & 5 \end{bmatrix} \begin{bmatrix} 0{,}013544018 & -0{,}062076749 \\ -0{,}062076749 & 0{,}326185101 \end{bmatrix} \begin{bmatrix} 24 \\ 5 \end{bmatrix} \right\}$$

$$S_T^2 = 131{,}51359$$

$$S_T = \sqrt{131{,}51359} = 11{,}46$$

olarak bulunur.

14.1.6. Eğrisel Çoklu Regresyon

Basit regresyonda olduğu gibi çoklu regresyonda da değişkenler arasında eğrisel ilişki olduğunda matematiksel şekilleri farklı modeller kurulacaktır. Bu durumda değişkenlerin tümü için aynı dönüşüm yapılabileceği gibi, değişkenlerin bazıları doğrusal bazıları ise eğrisel olarak alınabilir. Bu durumda daha önce verilen formüllerde, basit regresyonda yapıldığı gibi, değişiklikler yapılacaktır. Ancak buradaki açıklamalarda sadece tam logaritmik model örnek olarak verilecek diğer modellere değinilmeyecektir. Anakütle regresyon modeli,

$$Y_i = \beta_1 X_{i2}^{\beta_2} X_{i3}^{\beta_3} \cdots X_{ik}^{\beta_k} e_{\varepsilon_i}$$

olarak gösterildiğinde, doğal veya on tabanına göre logaritmasının alınması ile örnek regresyon modeli,

$$\ln Y_i = \ln \hat{\beta}_1 + \hat{\beta}_2 \ln X_{i2} + \hat{\beta}_3 \ln X_{i3} + \cdots + \hat{\beta}_k \ln X_{ik}$$

olarak tahmin edilecektir. Bunun için daha önce verilen formüllerde değişkenlerin logaritmalarının konulması yeterli olacaktır.

ÖRNEK: Daha önce verilen örnek için

$$\ln Y_i = \ln \hat{\beta}_1 + \hat{\beta}_2 \ln X_{i2} + \hat{\beta}_3 \ln X_{i3}$$

çoklu regresyon modelinin parametrelerini tahmin ederek, belirlilik katsayısını hesaplayınız.

Y_i	X_{i2}	X_{i3}	$\ln Y_i$	$\ln X_{i2}$	$\ln X_{i3}$
10	120	4	2,30258	4,78749	1,38629
13	135	8	2,56494	4,90527	2,07944
19	132	5	2,94443	4,88280	1,60943
18	140	8	2,89037	4,94164	2,07944
15	153	10	2,70803	5,03043	2,30258
75	680	35	13,41035	24,54763	9,45718

$(\ln Y_i)(\ln X_{i2})$	$(\ln Y_i)(\ln X_{i3})$	$(\ln X_{i3})(\ln X_{i2})$	$(\ln X_{i2})^2$	$(\ln X_{i3})^2$	$(\ln Y_i)^2$
11,02357	3,19204	6,63684	22,92006	1,92179	5,30187
12,58172	5,33363	10,20021	24,06167	4,32407	6,57891
14,37706	4,73885	7,85852	23,84173	2,59026	8,66966
14,28316	6,01035	10,27584	24,41980	4,32407	8,35423
13,62255	6,23545	11,58296	25,30522	5,30187	7,33342
65,88806	25,51032	46,55437	120,5485	18,46206	36,23809

$$(\overline{\ln Y}) = \frac{\sum \ln Y_i}{n} = \frac{13{,}41035}{5} = 2{,}68207$$

$$(\overline{\ln X_2}) = \frac{\sum \ln X_{i2}}{n} = \frac{24{,}54763}{5} = 4{,}90952$$

$$(\overline{\ln X_3}) = \frac{\sum \ln X_{i3}}{n} = \frac{9{,}45718}{5} = 1{,}89143$$

$$\sum [\ln X_{i2} - (\overline{\ln X_2})]^2 = \sum (\ln X_{i2})^2 - n(\overline{\ln X_2})^2$$

$$= 120{,}54848 - 5(4{,}90952)^2 = 0{,}03154$$

$$\sum [\ln X_{i3} - (\overline{\ln X_3})]^2 = \sum (\ln X_{i3})^2 - n(\overline{\ln X_3})^2$$

$$= 18{,}46206 - 5(1{,}89143)^2 = 0{,}57452$$

$$\sum [\ln X_{i2} - (\overline{\ln X_2})][\ln X_{i3} - (\overline{\ln X_3})] = \sum (\ln X_{i2})(\ln X_{i3}) - n(\overline{\ln X_2})(\overline{\ln X_3})$$

$$= 46{,}55437 - 5(4{,}90952)(1{,}89143) = 0{,}12430$$

$$\sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i2} - (\overline{\ln X_2})] = \sum (\ln Y_2)(\ln X_{i2}) - n(\overline{\ln Y})(\overline{\ln X_2})$$

$$= 65{,}88806 - 5(2{,}68207)(4{,}90952) = 0{,}04967$$

$$\sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i3} - (\overline{\ln X_3})] = \sum (\ln Y_2)(\ln X_{i3}) - n(\overline{\ln Y})(\overline{\ln X_3})$$

$$= 25{,}51032 - 5(2{,}68207)(1{,}89143) = 0{,}14558$$

$$\sum [\ln Y_i - (\overline{\ln Y})]^2 = \sum (\ln Y_i)^2 - n(\overline{\ln Y})^2$$

$$= 36{,}23809 - 5(2{,}68207)^2 = 0{,}27059$$

$$\ln Y_i = \ln \hat{\beta}_1 + \hat{\beta}_2 \ln X_{i2} + \hat{\beta}_3 \ln X_{i3}$$

$$\sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i2} - (\overline{\ln X_2})] = \hat{\beta}_2 \sum [\ln X_{i2} - (\overline{\ln X_2})]^2 + \hat{\beta}_3 \sum [\ln X_{i2} - (\overline{\ln X_2})][\ln X_{i3} - (\overline{\ln X_3})]$$

$$\sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i3} - (\overline{\ln X_3})] = \hat{\beta}_2 \sum [\ln X_{i2} - (\overline{\ln X_2})][\ln X_{i3} - (\overline{\ln X_3})] + \hat{\beta}_3 \sum [\ln X_{i3} - (\overline{\ln X_3})]^2$$

$$-0{,}12430 / 0{,}04967 = 0{,}03154\,\hat{\beta}_2 + 0{,}12430\,\hat{\beta}_3$$

$$0{,}03154 / 0{,}14558 = 0{,}12430\,\hat{\beta}_2 + 0{,}57452\,\hat{\beta}_3$$

$$-0{,}00617398 = -0{,}00392042\,\hat{\beta}_2 - 0{,}01545049\,\hat{\beta}_3$$

$$0{,}00459159 = 0{,}00392042\,\hat{\beta}_2 + 0{,}01812036\,\hat{\beta}_3$$

$$-0{,}00158239 = 0 + 0{,}00266987\,\hat{\beta}_3$$

$$\hat{\beta}_3 = \frac{-0{,}00158239}{0{,}00266987} = -0{,}59268$$

$$0{,}04967 = 0{,}03154\,\hat{\beta}_2 + 0{,}12430\,(-0{,}59268)$$

$$\hat{\beta}_2 = \frac{0{,}04967 + 0{,}07367}{0{,}03154} = 3{,}91058$$

$$\ln \hat{\beta}_1 = \overline{\ln Y} - \hat{\beta}_2(\overline{\ln X_2}) - \hat{\beta}_3(\overline{\ln X_3})$$

$$\ln \hat{\beta}_1 = 2{,}68207 - (3{,}91058)(4{,}90952) - (-0{,}59268)(1{,}89143)$$

$$= -15{,}39598$$

$$\ln Y_i = -15{,}39598 + 3{,}91058\ \ln X_{i2} - 0{,}59268\ \ln X_{i3}$$

$$\sum e_i^2 = \sum [\ln Y_i - (\overline{\ln Y})]^2 - \hat{\beta}_2 \sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i2} - (\overline{\ln X_2})]$$

$$- \hat{\beta}_3 \sum [\ln Y_i - (\overline{\ln Y})][\ln X_{i3} - (\overline{\ln X_3})]$$

$$= 0{,}27059 - (3{,}91058)(0{,}04967) - (-0{,}59268)(0{,}14558)$$

$$= 0{,}16263$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum [\ln Y_i - (\overline{\ln Y})]^2}$$

$$= 1 - \frac{0{,}16263}{0{,}27059}$$

$$= 1 - 0{,}601 = 0{,}399$$

14.2. ÇOKLU VE KISMİ KORELASYON

Basit regresyonda iki değişken arasındaki ilişkinin derecesi korelasyon katsayısının hesaplanması ile belirlenmekteydi. Ancak ikiden fazla değişken söz konusu olduğunda da çoklu korelasyon katsayısı hesaplanacak ve bu durumda iki değişken arasındaki ilişkinin derecesini belirlemek için kısmi korelasyon katsayısının hesaplanması gerekecektir.

14.2.1. Çoklu Korelasyon Katsayısı

Belirlilik katsayısı çoklu regresyon modelinde yer alan bağımsız değişkenlerin, bağımlı değişkendeki değişmeleri açıklama oranını göstermekte ve

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

olarak hesaplanmaktaydı. Çoklu regresyon modeli için hesaplanan belirlilik katsayısı ise çoklu belirlilik katsayısı olarak adlandırılmakta, $R^2_{Y.X_2X_3\cdots X_k}$ şeklinde gösterilmektedir. İndiste yer alan ($Y.X_2X_3...X_k$) ifadesi Y bağımlı değişkenindeki değişmelerin, $X_2X_3...X_k$ bağımsız değişkenleri tarafından açıklandığını göstermektedir.

Çoklu korelasyon katsayısı ise çoklu belirlilik katsayısının karekökü olarak tanımlanabilir ve aşağıda görülen formül ile,

$$R_{Y.X_2X_3\cdots X_k} = \sqrt{R^2_{Y.X_2X_2\cdots X_k}}$$

hesaplanabilir. Bu formül ile hesaplanan çoklu korelasyon katsayısının işareti ise daima pozitif olarak alınacaktır. Belirlilik katsayısının hesaplanması için gerekli formüller daha önce verilmişti. Bu formüllere göre hesaplanan belirlilik katsayısının karekökü alınarak çoklu korelasyon katsayısı hesaplanmaktadır.

ÖRNEK: Daha önce verilen örnek için çoklu korelasyon katsayısını hesaplayınız.Daha önce verilen örnekte Y bağımlı, X_2 ve X_3 bağımsız değişkenlerdi ve R^2 = 0,4786 olarak hesaplanmıştı. Bu verilerden yararlanılarak

$$R_{Y.X_2X_3} = \sqrt{R^2_{Y.X_2X_2}} = \sqrt{0,4786} = 0,6918$$

olarak bulunur. Bu değer değişkenler arasında kuvvetli ilişki olduğunu göstermektedir.

14.2.2. Kısmi Korelasyon Katsayıları

Çoklu belirlilik katsayısı bağımlı değişkendeki değişmelerin bağımsız değişkenler tarafından açıklanma oranını, çoklu korelasyon katsayısı ise değişkenler arasındaki ilişkinin derecesini göstermekte, fakat değişkenlerin birbirleri ile olan ilişkilerini belirtememektedirler. İki değişken arasındaki ilişkiyi diğer değişkenler sabit tutularak belirlemek için kısmi korelasyon katsayıları hesaplanır. kısmi korelasyon katsayılarının hesaplanması için önce kısmi belirlilik katsayılarının hesaplanması gerekecektir.

Y, X_2 ve X_3 değişkenlerinin yer aldığı bir analizde YX_2, YX_3 ve X_2 X_3 olmak Üzere Üç değişken çifti olacaktır. Y ve X_2 değişkenleri için kısmi belirlilik katsayısı $r^2_{YX_2.X_3}$ şeklinde ifade edilir ve

$$r^2_{YX_3.X_2} = 1 - \frac{1 - R^2_{Y.X_2X_3}}{1 - r^2_{YX_2}}$$

olarak tanımlanır. İndiste noktadan sonra yer alan değişkenlere ilişkide yer verilmekte, fakat sabit tutulduğu kabul edilmektedir. $r^2_{YX_3.X_2}$,X_2 sabit tutulduğuna göre Y ve X_3 değişkenleri için kısmi belirlilik katsayısını ifade etmektedir. Formülde yer alan $R^2_{Y.X_2X_3}$ çoklu belirlilik katsayısı, $r^2_{YX_3}$ ise Y ve X_2 değişkenleri için hesaplanacak korelasyon katsayısının karesidir ve

$$r^2_{YX_2} = 1 - \left[\frac{\sum_{i=1}^{n} x_{i2} y_i}{\sqrt{\sum_{i=1}^{n} x^2_{i2} \sum_{i=1}^{n} y^2_i}} \right] = \frac{\left(\sum_{i=1}^{n} x_{i2} y_i \right)^2}{\sum_{i=1}^{n} x^2_{i2} \sum_{i=1}^{n} y^2_i}$$

$$x_{i2} = X_{i2} - \overline{X}_2$$

$$y_i = Y_i - \overline{Y}$$

olarak hesaplanacaktır. Kısmi belirlilik katsayıları çoklu regresyonda modele yeni değişkenlerin eklenmesi ile bağımlı değişkendeki açıklanamayan değişmelerin nasıl azaldığının açıklanmasında yardımcı olmaktadırlar.

Yukarıda Y_1, X_2 değişken çifti için hesaplanan kısmi belirlilik katsayısı diğer değişken çiftleri için de aynı şekilde hesaplanabilir. Değişken sayısının artması halinde de verilen formüllere benzer formüller elde edilecektir. Modele X_4 değişkeninin eklenmesi halinde kısmi belirlilik katsayısı,

$$r^2_{YX4.X_2X_3} = 1 - \frac{1 - R^2_{Y.X_2X_3X_4}}{1 - R^2_{YX_2X_3}}$$

olacaktır. $r^2_{YX_4.X_2X_3}$ X_2 ve X_3'ün sabit tutulmaları halinde Y ve X_4 arasındaki ilişkiyi açıklamaktadır. $r^2_{YX_4.X_2X_3}$ diğer bir şekilde açıklanmayan değişmelerin oranları ile de ifade edilebilir. Bu durumda kısmi belirlilik katsayısı

$$r^2_{YX4.X_2X_3} = \frac{(1-R^2_{Y.X_2X_3})(1-R^2_{Y.X_2X_3X_4})}{1-R^2_{YX_2X_3}}$$

olarak ifade edilebilir. Formülün payında Y'deki değişmelerin X_2 ve X_3 tarafından açıklanmayan kısmının oranı ile; X_2, X_3 ve X_4 tarafından açıklanmayan kısmının oranı arasındaki fark yer almaktadır. $r^2_{YX_4.X_2X_3}$ ilişkiye X_4 değişkeninin eklenmesi ile Y'deki değişmelerin X_2 ve X_3 değişkenleri tarafından açıklanmayan kısmındaki azalış oranını göstermektedir.

Kısmi korelasyon katsayıları, kısmi belirlilik katsayılarının kareköklerinin alınması ile bulunurlar ve değişkenlerin bir kısmının ilişkide yer almaları fakat sabit tutulmaları halinde, sabit tutulmayan değişkenler arasındaki ilişkinin derecesini gösterirler. Bağımlı değişken ile bağımsız değişkenler için kısmi korelasyon katsayıları hesaplandığında, kısmi korelasyon katsayısı büyük olan bağımsız değişkenin, bağımlı değişkendeki değişmeleri diğer bağımsız değişkenlerden daha çok açıkladığı sonucuna varılacaktır.

ÖRNEK: Daha önce verilen örnek için $r^2_{YX_2.X_3}$ ve $r^2_{YX_3.X_2}$ kısmi korelasyon katsayılarını hesaplayınız.

$$r^2_{YX_2.X_3} = 1 - \frac{1-R^2_{Y.X_2X_3}}{1-r^2_{YX_3}}$$

$$r^2_{YX_3.X_2} = 1 - \frac{1-R^2_{Y.X_2X_3}}{1-r^2_{YX_2}}$$

Daha önce

$R^2_{Y.X_2X_3} = 0{,}4786$ $\quad \sum y_i^2 = 54$ $\quad \sum x_{i2}^2 = 578$

$\sum x_{i3}^2 = 24$ $\quad \sum x_{i2}y_i = 78$ $\quad \sum x_{i3}y_i = 8$

olarak hesaplanmıştı.

$$r^2_{YX_2} = \frac{(\sum x_{i2}y_i)^2}{\sum y_i^2 \sum x_{i2}^2}$$

$$= \frac{(78)^2}{54.578} = 0{,}1949$$

$$r^2_{YX_3} = \frac{(\sum x_{i3}y_i)^2}{\sum y_i^2 \sum x_{i3}^2}$$

$$= \frac{(8)^2}{54.24} = 0{,}0493$$

$$r^2_{YX_2.X_3} = 1 - \frac{1 - R^2_{Y.X_2X_3}}{1 - r^2_{YX_3}}$$

$$= 1 - \frac{1 - 0{,}4786}{1 - 0{,}0493}$$

$$= 1 - 0{,}5484 = 0{,}4516$$

$$r^2_{YX_2.X_3} = \sqrt{0{,}4516} = 0{,}6720$$

$$r^2_{YX_3.X_2} = 1 - \frac{1 - R^2_{Y.X_2X_3}}{1 - r^2_{YX_2}}$$

$$= 1 - \frac{1 - 0{,}4786}{1 - 0{,}1949}$$

$$= 1 - 0{,}6476 = 0{,}3524$$

$$r^2_{YX_3.X_2} = \sqrt{0{,}3524} = 0{,}5936$$

olacaktır.

14.3. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Bir bölgenin yıllık A malı tüketim miktarları malın yıllık ortalama reel fiyatı ve bu bölgede yaşayan ailelerin yıllık reel gelirleri aşağıda verilmiştir.

Yıllar	A Malı Tüketim Miktarı (1=10000 Adet)	Malın Reel Fiyatı (1=1 TL)	Ailelerin Yıllık Reel Geliri (1= 10000 TL)
2002	24	53	2,0
2003	28	51	2,4
2004	33	47	2,6
2005	26	52	2,5
2006	21	60	2,3
2007	30	54	2,6
2008	37	51	2,7
2009	41	40	2,9
	240	408	20,0

a) A malının tüketim miktarı için çoklu regresyon modelinin parametrelerini tahmin ediniz.
b) Parametrelerin varyanslarını hesaplayınız.
c) Belirlilik katsayısını hesaplayınız.
d) 2010 yılında malın reel fiyatı 46 TL. ve ailelerin reel gelirleri 30000 TL olarak tahmin edildiğine göre, 2010 yılı A malı tüketim miktarını tahmin ediniz.
e) Çoklu korelasyon katsayısını hesaplayınız.
f) $r_{YX_3.X_2}$ kısmi korelasyon katsayısını hesaplayınız.
g) Tahmin ettiğiniz sabit katsayıyı çift, diğerlerini tek taraflı olarak 0,05 hata payı ile test ediniz.
h) 0,05 hata payı ile F testi ile katsayıların anlamlılığını inceleyiniz.

Çözüm:

Y_i	X_{i2}	X_{i3}	Y_iX_{i2}	Y_iX_{i3}	$X_{i2}X_{i3}$	Y_i^2	X_{i2}^2	X_{i3}^2
24	53	2,0	1272	48,0	106,0	576	2809	4,00
28	51	2,4	1428	67,2	122,4	784	2601	5,76
33	47	2,6	1551	85,8	122,2	1089	2209	6,76
26	52	2,5	1352	65,0	130,0	676	2704	6,25
21	60	2,3	1260	48,3	138,0	441	3600	5,29
30	54	2,6	1620	78,0	140,4	900	2916	6,76
37	51	2,7	1887	99,9	137,7	1369	2601	7,29
41	40	2,9	1640	118,9	116,0	1681	1600	8,41
240	408	20,0	12010	611,1	1012,7	7516	21040	50,52

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{240}{8} = 30$$

$$\overline{X}_2 = \frac{\sum X_{i2}}{n} = \frac{408}{8} = 51$$

$$\overline{X}_3 = \frac{\sum X_{i3}}{n} = \frac{20,0}{8} = 2,5$$

a) $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} n & \sum X_{i2} & \sum X_{i3} \\ \sum X_{i2} & \sum X_{i2}^2 & \sum X_{i2}X_{i3} \\ \sum X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum X_{i2}Y_i \\ \sum X_{i3}Y_i \end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}8 & 408 & 20{,}0\\ 408 & 21040 & 1012{,}7\\ 20{,}0 & 1012{,}7 & 50{,}52\end{bmatrix}^{-1}\begin{bmatrix}240\\ 12010\\ 611{,}1\end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix}8 & 408 & 20{,}0\\ 408 & 21040 & 1012{,}7\\ 20{,}0 & 1012{,}7 & 50{,}52\end{bmatrix}^{-1}$$

D = [(8.21040.50,52) + (408.1012,7.20,0) + (20,0.408.1012,7)] – [(20,0.21040.20) + (1012,7.1012,7.8) + (50,52.408.408)]
= 538,8

$$(X'X)^{*} = \begin{bmatrix}\begin{vmatrix}21040 & 1012{,}7\\ 1012{,}7 & 50{,}52\end{vmatrix} & \begin{vmatrix}408 & 1012{,}7\\ 20{,}0 & 50{,}52\end{vmatrix} & \begin{vmatrix}408 & 21040\\ 20{,}0 & 1012{,}7\end{vmatrix}\\ \begin{vmatrix}408 & 20{,}0\\ 1012{,}7 & 50{,}52\end{vmatrix} & \begin{vmatrix}8 & 20{,}0\\ 20{,}0 & 50{,}52\end{vmatrix} & \begin{vmatrix}8 & 408\\ 20{,}0 & 1012{,}7\end{vmatrix}\\ \begin{vmatrix}408 & 20{,}0\\ 21040 & 1012{,}7\end{vmatrix} & \begin{vmatrix}8 & 20{,}0\\ 408 & 1012{,}7\end{vmatrix} & \begin{vmatrix}8 & 408\\ 408 & 21040\end{vmatrix}\end{bmatrix}$$

$$= \begin{bmatrix}\overset{+}{37379{,}51} & \overset{-}{358{,}16} & \overset{+}{-7618{,}4}\\ \overset{-}{358{,}16} & \overset{+}{4{,}16} & \overset{-}{-58{,}4}\\ \overset{+}{-7618{,}4} & \overset{-}{-58{,}4} & \overset{+}{1856}\end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{538{,}8}\begin{bmatrix}37379{,}51 & -358{,}16 & -7618{,}4\\ -358{,}16 & 4{,}16 & 58{,}4\\ -7618{,}4 & 58{,}4 & 1856\end{bmatrix}$$

$$= \begin{bmatrix}69{,}37548255 & -0{,}664736451 & -14{,}13956941\\ -0{,}664736451 & 0{,}007720861 & 0{,}108389012\\ -14{,}13956941 & 0{,}108389012 & 3{,}444691907\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix} = \begin{bmatrix}69{,}37548255 & -0{,}664736451 & -14{,}13956941\\ -0{,}664736451 & 0{,}007720861 & 0{,}108389012\\ -14{,}13956941 & 0{,}108389012 & 3{,}444691907\end{bmatrix}\begin{bmatrix}240\\ 12010\\ 611{,}1\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}16650,11581 & -7983,484776 & -8640,690866\\ -159,5367482 & +92,72754061 & +66,23652523\\ -3393,496658 & +1301,752034 & +2105,051224\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}25,940168\\ -0,57268236\\ 13,3066\end{bmatrix}$$

$Y_i = 25,9401 - 0,5726\ X_{i2} + 13,3066\ X_{i3}$

b) $S_e^2=\dfrac{\sum e_i^2}{n-k}$

$$\sum_{i=1}^{n} e_i^2=\sum_{i=1}^{n} Y_i^2-\hat{\beta}_1\sum_{i=1}^{n} Y_i-\hat{\beta}_2\sum_{i=1}^{n} X_{i2}Y_i-\hat{\beta}_3\sum_{i=1}^{n} X_{i3}Y_i$$

$$= 7516 - (25,9401)(240) - (-0,5726)(12010) - (13,3066)(611,1)$$

$$= 35,63874$$

$$S_e^2=\frac{35,63874}{8-3}=7,1277$$

$$\begin{bmatrix}Var(\hat{\beta}_1) & Kov(\hat{\beta}_1\hat{\beta}_2) & Kov(\hat{\beta}_1\hat{\beta}_3)\\ Kov(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_2) & Kov(\hat{\beta}_2\hat{\beta}_3)\\ Kov(\hat{\beta}_1\hat{\beta}_3) & Kov(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_3)\end{bmatrix}=S_e^2(X'X)^{-1}$$

$$=7,1277\begin{bmatrix}69,37548255 & -0,664736451 & -14,13956941\\ -0,664736451 & 0,007720861 & 0,108389012\\ -14,13956941 & 0,108389012 & 3,444691907\end{bmatrix}$$

$S^2_{\hat{\beta}_1} = (7,1277)(69,37548255) = 494,48762 \quad S_{\hat{\beta}_1} = 22,23707$

$S^2_{\hat{\beta}_2} = (7,1277)(0,007720861) = 0,05503 \quad S_{\hat{\beta}_2} = 0,23458$

$S^2_{\hat{\beta}_3} = (7,1277)(3,444691907) = 24,55273 \quad S_{\hat{\beta}_3} = 4,95507$

c) $R^2 = 1 - \dfrac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2$$

$$= 7516 - 8(30)^2 = 316$$

$$R^2 = 1 - \frac{35{,}63874}{316}$$

$$= 1 - 0{,}1127 = 0{,}8873$$

d) $Y_i = 25{,}9401 - 0{,}5726X_{i2} + 13{,}3066X_{i3}$

$$\hat{Y}_6 = 25{,}9401 - 0{,}5726(46) + 13{,}3066(3)$$

$$= 39{,}5203$$

e) $R_{Y.X_2X_3} = \sqrt{R^2} = \sqrt{0{,}8873} = 0{,}9419$

f) $r^2_{YX_3.X_2} = 1 - \dfrac{1 - R^2_{Y.X_2X_3}}{1 - r^2_{YX_2}}$

$$r^2_{YX_2} = \frac{(\sum X_{i2}Y_i - n\bar{X}_2\bar{Y})^2}{[\sum X_{i2}^2 - n(\bar{X}_2)^2][\sum Y_i^2 - n(\bar{Y})^2]}$$

$$= \frac{[12010 - 8(51)(30)]^2}{[21040 - 8(51)^2][7516 - 8(30)^2]}$$

$$= \frac{(-230)^2}{(232)(316)} = 0{,}7215$$

$$r^2_{YX_3.X_2} = 1 - \frac{1 - R^2_{Y.X_2X_3}}{1 - r^2_{YX_2}}$$

$$= 1 - \frac{1 - 0{,}8873}{1 - 0{,}7215}$$

$= 1 - 0,4046 = 0,5954$

$r^2_{YX_3.X_2} = 0,7716$

X_2 değişkeninin etkisi sabit tutulduğunda Y ve X_3 değişkenleri arasındaki ilişkinin derecesi 0,77 olarak bulunmuştur.

g) $Y_i = 25,9401 - 0,5726X_{i2} + 13,3066X_{i3}$
$(22,23707)\ (0,23458)\ \ (4,95507)$

olarak bulunmuştur.

β_1'in testi:

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
$t_{\alpha/2,n-k} = t_{0,05/2,5} = 2,571$

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{25,9401}{22,23707} = 1,166$$

$|1,166| < |2,571|$ olduğundan H_0 hipotezi kabul edilir.

β_2'in testi:

H_0: $\beta_2 = 0$
H_1: $\beta_2 < 0$
$t_{\alpha,n-k} = t_{0,05,5} = 2,015$

$$t = \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}} = \frac{-0,5726}{0,23458} = -2,440$$

$|2,440| > |2,015|$ olduğundan H_1 hipotezi kabul edilir.

β_3'in testi:

H_0: $\beta_3 = 0$
H_1: $\beta_3 > 0$

$t_{\alpha,n-k} = t_{0,05,5} = 2,015$

$$t = \frac{\hat{\beta}_3}{S_{\hat{\beta}_3}} = \frac{13{,}3066}{4{,}95507} = 2{,}685$$

$|2{,}685| > |2{,}015|$ olduğundan H_1 hipotezi kabul edilir.

h) $H_0 : \beta_2 = \beta_3 = 0$

$H_1 : \beta_2 \neq \beta_3 \neq 0$

$F_{\alpha,k-1,n-k} = F_{0{,}05,2,5} = 5{,}79$

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / (k-1)}{\sum (Y_i - \hat{Y})^2 / (n-k)}$$

$$\sum (\hat{Y}_i - \bar{Y})^2 = \sum (Y - \bar{Y})^2 - \sum (Y - \hat{Y})^2$$

$$= 316 - 35{,}63874 = 280{,}36126$$

$$F = \frac{280{,}36126 / (3-1)}{35{,}63874 / (8-2)} = 23{,}600$$

$23{,}600 > 5{,}79$ olduğundan H_1 hipotezi kabul edilir.

$R^2 = 0{,}8873$ olduğundan F test istatistiği

$$F = \frac{R^2}{1-R^2} . \frac{n-k}{k-1}$$

$$= \frac{0{,}8873}{1-0{,}8873} . \frac{8-2}{3-1} = 23{,}619$$

olarak da hesaplanabilir.

ÖRNEK 2: Örnek 1 için

$$\log Y_i = \log \hat{\beta}_1 + \hat{\beta}_2 \log X_{i2} + \hat{\beta}_3 \log X_{i3}$$

çoklu regresyon modelinin parametrelerini tahmin ediniz.

Çözüm:

Y_i	X_{i2}	X_{i3}	$\log Y_i$	$\log X_{i2}$	$\log X_{i3}$
24	53	2,0	1,38021	1,72427	0,30103
28	51	2,4	1,44715	1,70757	0,38021
33	47	2,6	1,51851	1,67209	0,41497
26	52	2,5	1,41497	1,71600	0,39794
21	60	2,3	1,32221	1,77815	0,36172
30	54	2,6	1,47712	1,73239	0,41497
37	51	2,7	1,56820	1,70757	0,43136
41	40	2,9	1,61278	1,60206	0,46239
240	408	20,0	11,74115	13,64010	3,16459

$(\log Y_i)^2$	$(\log X_{i2})^2$	$(\log X_{i3})^2$	$(\log X_{i2})(\log X_{i3})$	$(\log X_{i2})(\log Y_i)$	$(\log X_{i3})(\log Y_i)$
1,90497	2,97310	0,09061	0,51905	2,37985	0,41548
2,09424	2,91579	0,14455	0,64923	2,4711	0,55022
2,30587	2,79588	0,1722	0,69386	2,53908	0,63013
2,00214	2,94465	0,15835	0,68286	2,42808	0,56307
1,74823	3,16181	0,13084	0,64319	2,35108	0,47826
2,18188	3,00117	0,17220	0,71888	2,55894	0,61296
2,45925	2,91579	0,18607	0,73657	2,67781	0,67645
2,60105	2,56659	0,2138	0,74077	2,58377	0,74573
17,29763	23,27478	1,26862	5,38441	19,98971	4,67230

$$\begin{bmatrix} \log\hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} n & \log X_{i2} & \log X_{i3} \\ \log X_{i2} & \log X_{i2}^2 & (\log X_{i2})(\log X_{i3}) \\ \log X_{i3} & (\log X_{i3})(\log X_{i2}) & \log X_{i3}^2 \end{bmatrix}^{-1} \begin{bmatrix} \log Y_i \\ (\log Y_i)(\log X_{i2}) \\ (\log Y_i)(\log X_{i3}) \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 8 & 13,64010 & 3,16459 \\ 13,64010 & 23,27478 & 5,38441 \\ 3,16459 & 5,38441 & 1,26862 \end{bmatrix}^{-1}$$

D = (8.23,27478.1,26862) + (13,64010.5,38441.3,16459) +
(3,16459.13,64010.5,38441) – (3,16459.23,27478.3,16459) –
(5,38441.5,38441.8) – (1,26862.13,64010.13,64010)
= 0,0014366

$$(X'X)^{*} = \begin{bmatrix} \begin{vmatrix} 23,27478 & 5,38441 \\ 5,38441 & 1,26862 \end{vmatrix} & \begin{vmatrix} 13,64010 & 5,38441 \\ 3,16459 & 1,26862 \end{vmatrix} & \begin{vmatrix} 13,64010 & 23,27478 \\ 3,16459 & 5,38441 \end{vmatrix} \\ \begin{vmatrix} 13,64010 & 3,16459 \\ 5,38441 & 1,26862 \end{vmatrix} & \begin{vmatrix} 8 & 3,16459 \\ 3,16459 & 1,26862 \end{vmatrix} & \begin{vmatrix} 8 & 13,64010 \\ 3,16459 & 5,38441 \end{vmatrix} \\ \begin{vmatrix} 13,64010 & 3,16459 \\ 23,27478 & 5,38441 \end{vmatrix} & \begin{vmatrix} 8 & 3,16459 \\ 13,64010 & 5,38441 \end{vmatrix} & \begin{vmatrix} 8 & 13,64010 \\ 13,64010 & 23,27478 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \overset{+}{0,53498036} & \overset{-}{0,26465362} & \overset{+}{-0,2112452} \\ \overset{-}{0,26465362} & \overset{+}{0,13433014} & \overset{-}{-0,09004405} \\ \overset{+}{-0,2112452} & \overset{-}{-0,09004405} & \overset{+}{0,145912} \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{0,0014366} \begin{bmatrix} 0,53498036 & -0,26465362 & -0,2112452 \\ -0,26465362 & 0,13433014 & 0,09004405 \\ -0,2112452 & 0,09004405 & 0,145912 \end{bmatrix}$$

$$= \begin{bmatrix} 372,393401 & -184,2222052 & -147,0452457 \\ -184,2222052 & 93,50559654 & 62,67858137 \\ -147,0452457 & 62,67858137 & 101,5675901 \end{bmatrix}$$

$$\begin{bmatrix} \log \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 372,393401 & -184,2222052 & -147,0452457 \\ -184,2222052 & 93,50559654 & 62,67858137 \\ -147,0452457 & 62,67858137 & 101,5675901 \end{bmatrix} \begin{bmatrix} 11,74115 \\ 19,98971 \\ 4,67230 \end{bmatrix}$$

$$\begin{bmatrix} \log \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 2,73882 \\ -0,97765 \\ 1,00062 \end{bmatrix}$$

$$\log Y_i = 2,73882 - 0,97765 \log X_{i2} + 1,00062 \log X_{i3}$$

$$\sum (\log Y_i - \log \hat{Y}_i)^2 = \sum (\log Y_i)^2 - \log \hat{\beta}_1 \sum \log Y_i - \hat{\beta}_2 \sum (\log X_{i2})(\log Y_i) - \hat{\beta}_3 \sum (\log X_{i3})(\log Y_i)$$

$$= 17,29763 - 2,73882(11,74115) - (-0,97765)(19,98971) - (1,00062)(4,67230)$$

$$= 0,00847$$

$$\sum [\log Y_i - (\overline{\log Y})]^2 = \sum (\log Y_i)^2 - n(\overline{\log Y})^2$$

$$\overline{\log Y} = \frac{\sum \log Y_i}{n} = \frac{11{,}74115}{8} = 1{,}46764$$

$$\sum [\log Y_i - (\overline{\log Y})]^2 = 17{,}29763 - 8(1{,}46764)^2$$
$$= 0{,}06589$$

$$R^2 = 1 - \frac{\sum [\log Y_i - (\log \hat{Y}_i)]^2}{\sum [\log Y_i - (\overline{\log Y})]^2}$$

$$= 1 - \frac{0{,}00847}{0{,}06589} = 1 - 0{,}1285 = 0{,}8715$$

ÖRNEK 3: Aşağıda bir malın 6 yıllık satış miktarları (Y_i), yıllık ortalama satış fiyatı (X_{i2}) ve rakip malların yıllık satış fiyatlarının ortalaması (X_{i3}) verilmiştir.

Yıllar	Y_i (1=1000 Adet)	X_{i2} (1=1 TL)	X_{i3} (1=1 TL)
2004	181	16	18
2005	195	20	21
2006	218	25	30
2007	204	31	32
2008	233	40	38
2009	247	36	41
	1278	168	180

a) Çoklu doğrusal regresyon modelinin parametrelerini ortalamadan fark matrisleri ile tahmin ediniz.
b) Tam logaritmik regresyon modelinin ortalamadan fark matrisleri ile parametrelerini tahmin ediniz.
c) Her iki modelin belirlilik katsayılarını hesaplayınız.

Çözüm:

a)

Y_i	X_{i2}	X_{i3}	y_i	x_{i2}	x_{i3}	$y_i x_{i2}$	$y_i x_{i3}$	$x_{i2} x_{i3}$
181	16	18	-32	-12	-12	384	384	44
195	20	21	-18	-8	-9	144	162	72
218	25	30	5	-3	0	-15	0	0
204	31	32	-9	3	2	-27	-18	6
233	40	38	20	12	8	240	160	96
247	36	41	34	8	11	272	374	88
1278	168	180	0	0	0	998	1062	406

x_{i2}^2	x_{i3}^2	y_i^2
144	144	1024
64	81	324
9	0	25
9	4	81
144	64	400
64	121	1156
434	414	3010

$$\overline{Y} = \frac{\sum Y}{n} = \frac{1278}{6} = 213$$

$$\overline{X}_2 = \frac{\sum X_{i2}}{n} = \frac{168}{6} = 28$$

$$\overline{X}_3 = \frac{\sum X_{i3}}{n} = \frac{180}{6} = 30$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum x_{i2}^2 & \sum x_{i2}x_{i3} \\ \sum x_{i2}x_{i3} & \sum x_{i3}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i x_{i2} \\ \sum y_i x_{i3} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 434 & 406 \\ 406 & 414 \end{bmatrix}^{-1} \begin{bmatrix} 998 \\ 1062 \end{bmatrix}$$

$(\underline{X}' \underline{X})$ matrisinin determinantı

$D = 434.414 - (406)^2 = 14840$

$$(\underline{X}' \underline{X})^{-1} = \frac{1}{14840} \begin{bmatrix} \overset{+}{414} & \overset{-}{406} \\ \underset{-}{406} & \overset{+}{434} \end{bmatrix}$$

$$(\underline{X}'\,\underline{X})^{-1} = \begin{bmatrix} 0,0278975 & -0,0273584 \\ -0,0273584 & 0,0292452 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0,0278975 & -0,0273584 \\ -0,0273584 & 0,0292452 \end{bmatrix} \begin{bmatrix} 998 \\ 1062 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} -1,2129158 \\ 3,7547192 \end{bmatrix}$$

$\hat{\beta}_2 \cong -1,21$

$\hat{\beta}_3 \cong 3,75$

$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_2 - \hat{\beta}_3 \overline{X}_3$

= 213-(-1,21)(28) – (3,75)(30) = 134,38

$Y_i = 134,38 - 1,21X_{i2} + 3,75X_{i3}$

b)

Y_i	X_{i2}	X_{i3}	$\log Y_i$	$\log X_{i2}$	$\log X_{i3}$
181	16	18	2,25767	1,20411	1,25527
195	20	21	2,29003	1,30102	1,32221
218	25	30	2,33845	1,39794	1,47712
204	31	32	2,30963	1,49136	1,50514
233	40	38	2,36735	1,60205	1,57978
247	36	41	2,39269	1,5563	1,61278
1278	168	180	13,95582	8,55278	8,75230

$(\log Y_i)^2$	$(\log X_{i2})^2$	$(\log X_{i3})^2$	$(\log Y_i)(\log X_{i2})$	$(\log Y_i)(\log X_{i3})$	$(\log X_{i3})(\log X_{i2})$
5,09707	1,44988	1,5757	2,71848	2,83398	1,51148
5,24423	1,69265	1,74823	2,97937	3,0279	1,72022
5,46834	1,95423	2,18188	3,26901	3,45417	2,06492
5,33439	2,22415	2,26544	3,44448	3,47631	2,24470
5,60434	2,56656	2,49570	3,79261	3,73989	2,53088
5,72496	2,42206	2,60105	3,72374	3,85888	2,50996
32,47333	12,30953	12,86800	19,92769	20,39113	12,58216

$$(\overline{\log Y}) = \frac{\sum \log Y_i}{n}$$

$$= \frac{13{,}95582}{6} = 2{,}32597$$

$$(\overline{\log X_2}) = \frac{\sum \log X_{i2}}{n}$$

$$= \frac{8{,}55278}{6} = 1{,}42546$$

$$(\overline{\log X_3}) = \frac{\sum \log X_{i3}}{n}$$

$$= \frac{8{,}75230}{6} = 1{,}45871$$

$$\sum [\log X_{i2} - (\overline{\log X_2})]^2 = \sum (\log X_{i2})^2 - n(\overline{\log X_2})^2$$
$$= 12{,}30953 - 6(1{,}42546)^2$$
$$= 0{,}11791$$

$$\sum [\log X_{i3} - (\overline{\log X_3})]^2 = \sum (\log X_{i3})^2 - n(\overline{\log X_3})^2$$
$$= 12{,}86800 - 6(1{,}45871)^2$$
$$= 0{,}10099$$

$$\sum [\log X_{i2} - (\overline{\log X_2})][\log X_{i3} - (\overline{\log X_3})] = \sum (\log X_{i2})(\log X_{i3}) - n(\overline{\log X_2})(\overline{\log X_3})$$
$$= 12{,}58216 - 6(1{,}42546)(1{,}45871)$$
$$= 0{,}10616$$

$$\sum [\log Y_i - (\overline{\log Y})][\log X_{i2} - (\overline{\log X_2})] = \sum (\log Y_2)(\log X_{i2}) - n(\overline{\log Y})(\overline{\log X_2})$$
$$= 19{,}92769 - 6(2{,}32597)(1{,}42546)$$
$$= 0{,}03422$$

$$\sum [\log Y_i - (\overline{\log Y})][\log X_{i3} - (\overline{\log X_3})] = \sum (\log Y_2)(\log X_{i3}) - n(\overline{\log Y})(\overline{\log X_3})$$
$$= 20{,}39113 - 6(2{,}32597)(1{,}45871)$$
$$= 0{,}03363$$

$$\sum [\log Y_i - (\overline{\log Y})]^2 = \sum (\log Y_i)^2 - n(\overline{\log Y})^2$$
$$= 32{,}47333 - 6(2{,}32597)^2$$
$$= 0{,}01251$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}\sum[\log x_{i2}]^2 & \sum[(\log x_{i2})(\log x_{i3})]\\ \sum[(\log x_{i2})(\log x_{i3})] & \sum[\log x_{i3}]^2\end{bmatrix}\begin{bmatrix}\sum(\log y_i)(\log x_{i2})\\ \sum(\log y_i)(\log x_{i3})\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}0{,}11791 & 0{,}10616\\ 0{,}10616 & 0{,}10099\end{bmatrix}^{-1}\begin{bmatrix}0{,}03422\\ 0{,}03363\end{bmatrix}$$

$D = (0{,}11791)(0{,}10099) - (0{,}10616)^2 = 0{,}0006377$

$$\begin{bmatrix}0{,}11791 & 0{,}10616\\ 0{,}10616 & 0{,}10099\end{bmatrix}^{-1}=\frac{1}{0{,}0006377}\begin{bmatrix}\overset{+}{0{,}10099} & \overset{-}{0{,}10616}\\ \overset{-}{0{,}10616} & \overset{+}{0{,}11791}\end{bmatrix}$$

$$=\begin{bmatrix}158{,}36600 & -166{,}47326\\ -166{,}47326 & 184{,}89885\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}158{,}36600 & -166{,}47326\\ -166{,}47326 & 184{,}89885\end{bmatrix}\begin{bmatrix}0{,}03422\\ 0{,}03363\end{bmatrix}$$

$$\begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}-0{,}17921\\ 0{,}52143\end{bmatrix}$$

$$\log\hat{\beta}_1=\overline{\log Y}-\hat{\beta}_2\overline{\log X}_2-\hat{\beta}_3\overline{\log X}_3$$

$= 2{,}32597-(-0{,}17921)(1{,}42546) - (0{,}52143)(1{,}45871)$
$= 1{,}82081$

$\log Y_i = 1{,}82081 - 0{,}17921\log X_{i2} + 0{,}52143\ \log X_{i3}$

c) Doğrusal model

$$\sum e_i^2=\sum y_i^2-\hat{\beta}_2\sum y_i x_{i2}-\hat{\beta}_3\sum y_i x_{i3}$$

$= 3010 - (-1{,}21)(998) - (3{,}75)(1062)$

$= 235{,}08$

$$R^2=1-\frac{\sum e_i^2}{\sum y_i^2}$$

$$=1-\frac{235{,}08}{3010}$$

$= 1 - 0{,}078 = 0{,}922$

Logaritmik Model

$$\sum e_i^2 = \sum (\log y_i)^2 - \hat{\beta}_2 \sum (\log y_i)(\log x_{i2}) - \hat{\beta}_3 \sum (\log y_i)(\log x_{i3})$$

$$= 0{,}01251 - (-0{,}17921)(0{,}03422) - (0{,}52143)(0{,}03363)$$

$$= 0{,}001106$$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (\log y_i)^2}$$

$$= 1 - \frac{0{,}001106}{0{,}01251}$$

$$= 1 - 0{,}088 = 0{,}912$$

ÖRNEK 4: Aşağıda bir seri verilmiştir.

Yıllar	Y_i	X_{i2}	X_{i3}	X_{i4}
2001	5	12	18	4
2002	6	13	17	5
2003	4	10	16	3
2004	7	15	15	8
2005	8	18	16	6
2006	11	21	13	7
2007	9	22	12	9
2008	10	25	11	10
2009	12	26	8	11
	72	162	126	63

Serinin çoklu doğrusal regresyon modelinin parametrelerini ortalamadan farklar matrisi ile tahmin ediniz.

Çözüm:

Yıllar	Y_i	X_{i2}	X_{i3}	X_{i4}	y_i	x_{i2}	x_{i3}
2001	5	12	18	4	-3	-6	4
2002	6	13	17	5	-2	-5	3
2003	4	10	16	3	-4	-8	2
2004	7	15	15	8	-1	-3	1
2005	8	18	16	6	0	0	2
2006	11	21	13	7	3	3	-1
2007	9	22	12	9	1	4	-2
2008	10	25	11	10	2	7	-3
2009	12	26	8	11	4	8	-6
	72	162	126	63	0	0	0

x_{i4}	x_{i2}^2	x_{i3}^2	x_{i4}^2	$y_i x_{i2}$	$y_i x_{i3}$	$y_i x_{i4}$	$x_{i2} x_{i3}$
-3	36	16	9	18	-12	9	-24
-2	25	9	4	10	-6	4	-15
-4	64	4	16	32	-8	16	-16
1	9	1	1	3	-1	-1	-3
-1	0	4	1	0	0	0	0
0	9	1	0	9	-3	0	-3
2	16	4	4	4	-2	2	-8
3	49	9	9	14	-6	6	-21
4	64	36	16	32	-24	16	-48
0	272	84	60	122	-62	52	-138

$x_{i2} x_{i4}$	$x_{i3} x_{i4}$
18	-12
10	-6
32	-8
-3	1
0	-2
0	0
8	-4
21	-9
32	-24
118	-64

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{72}{9} = 8$$

$$\bar{X}_2 = \frac{\sum X_{i2}}{n} = \frac{162}{9} = 18$$

$$\bar{X}_3 = \frac{\sum X_{i3}}{n} = \frac{126}{9} = 14$$

$$\bar{X}_4 = \frac{\sum X_{i4}}{n} = \frac{63}{9} = 7$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} \sum x_{i2}^2 & \sum x_{i2} x_{i3} & \sum x_{i2} x_{i4} \\ \sum x_{i2} x_{i3} & \sum x_{i3}^2 & \sum x_{i3} x_{i4} \\ \sum x_{i2} x_{i4} & \sum x_{i3} x_{i4} & \sum x_{i4}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i x_{i2} \\ \sum y_i x_{i3} \\ \sum y_i x_{i4} \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 272 & -138 & 118 \\ -138 & 84 & -64 \\ 118 & -64 & 60 \end{bmatrix}^{-1}$$

D = [(272)(84)(60) + (-138)(-64)(118) + (118)(-138)(-64)]
– [(118)(84)(118) +(-64)(-64)(272) – (60)(-138)(-138)
= 28864

$$(\underline{X}'\underline{X})^{*} = \begin{bmatrix} \begin{vmatrix} 84 & -64 \\ -64 & 60 \end{vmatrix} & \begin{vmatrix} -138 & -64 \\ 118 & 60 \end{vmatrix} & \begin{vmatrix} -138 & 84 \\ 118 & -64 \end{vmatrix} \\ \begin{vmatrix} -138 & 118 \\ -64 & 60 \end{vmatrix} & \begin{vmatrix} 272 & 118 \\ 118 & 60 \end{vmatrix} & \begin{vmatrix} 272 & -138 \\ 118 & -64 \end{vmatrix} \\ \begin{vmatrix} -138 & 118 \\ 84 & -64 \end{vmatrix} & \begin{vmatrix} 272 & 118 \\ -138 & -64 \end{vmatrix} & \begin{vmatrix} 272 & -138 \\ -138 & 84 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \overset{+}{944} & \overset{-}{-728} & \overset{+}{-1080} \\ \overset{-}{-728} & \overset{+}{2396} & \overset{-}{-1124} \\ \overset{+}{-1080} & \overset{-}{-1124} & \overset{+}{3804} \end{bmatrix}$$

$$(\underline{X}'\underline{X})^{-1} = \frac{1}{D}(\underline{X}'\underline{X})^{*}$$

$$(\underline{X}'\underline{X})^{-1} = \frac{1}{28864}\begin{bmatrix} 944 & 728 & -1080 \\ 728 & 2396 & 1124 \\ -1080 & 1124 & 3804 \end{bmatrix}$$

$$= \begin{bmatrix} 0{,}0327050 & 0{,}0252217 & -0{,}0374168 \\ 0{,}0252217 & 0{,}0830099 & 0{,}0389412 \\ -0{,}0374168 & 0{,}0389412 & 0{,}1317904 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 0{,}0327050 & 0{,}0252217 & -0{,}0374168 \\ 0{,}0252217 & 0{,}0830099 & 0{,}0389412 \\ -0{,}0374168 & 0{,}0389412 & 0{,}1317904 \end{bmatrix}\begin{bmatrix} 122 \\ -62 \\ 52 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 0{,}4805 \\ -0{,}0446 \\ -0{,}1261 \end{bmatrix}$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2\overline{X}_2 - \hat{\beta}_3\overline{X}_3 - \hat{\beta}_4\overline{X}_4$$

= 8-(0,4805)(18) – (-0,0446)(14) – (-0,1261)(7) =0,8581

Y_i =0,8581 + 0,4805 X_{i2} –0,0446 X_{i3} – 0,1261 X_{i4}

BÖLÜM 15

ZAMAN SERİLERİNİN ANALİZİ

İkinci bölümde istatistiksel serileri açıklarken zaman serileri tanımlanmıştı. Zaman serileri pek çok faktör tarafından etkilenebilirler. Zaman serilerini geçmiş devrelerde etkileyen faktörler etkilerini gelecek devrelerde de sürdürebilirler. Seriyi geçmiş devrelerde etkileyen faktörlerin incelenmesi, etkinin gelecek devrelerde de değişmeden devam edeceği varsayımı ile gelecek devre değerlerinin tahmini için önemlidir.

Zaman serileri çeşitli zaman birimlerine göre düzenlenebilirlerse de, uygulamada genellikle aylık, üçer aylık veya yıllık olarak kullanılmaktadırlar. Örneğin, devlet ithalat ve ihracat istatistiklerini aylık, gayri safi milli hasıla ile ilgili istatistikleri ise üçer aylık olarak yayınlamaktadır. Zaman serileri yıllık olarak düzenlenmiş ise, yıllık değerler aylık değerlerin toplamından veya ortalamasından oluşabilir. Örneğin, yıllık ithalat, ihracat verileri veya bir işletmenin yıllık satışları ile ilgili veriler, genellikle aylık değerlerin toplamı olarak düzenlenirse de; yılların aylık ortalama ihracat, ithalat veya satış miktarlarını göstermek için, aylık değerlerin ortalamaları alınarak da zaman serileri düzenlenebilir.

Zaman serilerinin analizi ile gelecek devrelere ait tahminler ve planlar yapılabilir, yeni stratejiler geliştirilebilir. Bunlar ülke ekonomisi ile ilgili olabilecekleri gibi, işletmeler gibi daha küçük birimlerle ilgili de olabilirler. Yapılacak tahminlerde geçmiş devrelerdeki etkinin gelecekte de devam edeceği varsayılmaktadır. Gelecek devrelerde değişiklikler olacağı tahmin edilse bile, yapılacak çalışmalar en azından bu konuda bir fikir verecek veya yol gösterecektir.

15.1. ZAMAN SERİLERİNİ ETKİLEYEN FAKTÖRLER

Zaman serileri çok sayıda faktör tarafından etkilenirler. Hangi faktörler tarafından etkilendikleri ise serilerin ilgili oldukları konuya ve verilerin düzenlenme şekillerine göre değişecektir. Örneğin, iktisadi olaylar ile ilgili seriler ile, bir bölgenin sıcaklığı, bir şehirde belirli bir hastalıktan ölenlerin sayıları, bir karayolunda meydana gelen kazalar için düzenlenen serilerin aynı faktörler tarafından etkileneceğini söyleyemeyiz. Bu serileri etkileyen faktörlerin bazıları aynı olsa bile, bazıları farklı olacaktır. Ayrıca, yıllık olarak düzenlenmiş bir seride mevsim etkisinden söz edilemeyeceği de açıktır.

Burada iktisadi zaman serilerini etkileyen faktörler ele alınacaktır. İktisadi zaman serilerinin de konularına göre farklı faktörler etkileyebilirse de, bu faktörler genelleştirilebilir. İktisadi zaman serilerini genel olarak dört faktör tarafından etkilendiği kabul edilmektedir. Bu faktörler şöyle sıralanmaktadır.

1- Trend
2- Mevsimlik dalgalanmalar
3- Konjonktürel dalgalanmalar
4- Düzensiz dalgalanmalar

Zaman serilerinin uzun devre eğilimine trend adı verilmektedir. Seride devreden devreye dalgalanmalar, yani artışlar veya azalışlar olacaktır. Trend bu dalgalanmaların genel karakterini ifade etmektedir. Seriler uzun devrede artma veya azalma eğilimi gösterebilecekleri gibi, sabit de kalabilirler. Örneğin, bir işletmenin yıllık üretim miktarları incelendiğinde üretim çeşitli nedenlerle bazı yıllarda artış, bazı yıllarda ise azalış gösterebilir. Üretim miktarlarının trendi bu kısa devreli artış ve azalışların sonucunda serinin gösterdiği eğilimdir.

Mevsimlik dalgalanmalar, mevsimlerin zaman serisi üzerindeki etkilerini ifade etmektedir. Bazı olaylar mevsimlerden etkilenebilirler. Örneğin, işsizlik oranının mevsimlere göre farklılık göstermesi, meşrubat satışlarının yaz aylarında artması gibi. Mevsimlerin seriyi etkileyip etkilemediğini, etkiliyorsa etkinin ne kadar olduğunun belirlenmesi için mevsimlik dalgalanmaların incelenmesi gerekecektir. Mevsimlerin etkisinin incelenebilmesi için yildan daha kısa zaman birimlerine göre (aylık, üç aylık gibi) düzenlenmiş olması gerekecektir.

Konjonktürel dalgalanmalar ekonominin genel durumu ile ilgili dalgalanmalardır. Ekonomik durumun durgunluk, yükseliş, refah ve gerileme şeklinde gösterdiği dalgalanmalara konjonktürel dalgalanmalar adı verilmekte ve bu dalgalanmaların periyodik olarak tekrarlandığı ileri sürülmektedir. Yapılan araştırmalarda 3 yılda tekrarlanan konjonktürel dalgalanmalara rastlandığı gibi, yaklaşık 50 yılda tekrarlanan konjonktürel dalgalanmalara da rastlanmıştır. Bu dalgalanmalar ekonominin genel durumu ile ilgili olduklarından tüm ekonomik olayları etkilerler.

Düzensiz dalgalanmalar, tesadüfi hareketler olarak da adlandırılabilirler. Bu dalgalanmalar doğal, ekonomik veya siyasi nedenlerle ortaya çıkabilirler. Deprem, kuraklık, savaş, siyasi istikrarsızlık, grev gibi sebepler düzensiz dalgalanmalara neden olmaktadır. Bu tür olayların ne zaman olacakları belli olmadığından düzensiz hareketlerdir ve önceden tahmin edilemezler. Düzensiz hareketlerin ne zaman ortaya

çıkacakları belli olmadığı gibi, şiddetleri de farklı olacak ve etkileri şiddetlerine bağlı olarak artacak veya azalacaktır.

Zaman serilerinin analizinde zaman X ve serinin gerçek değerleri Y ile; zaman serisini etkileyen faktörlerden trend T, mevsimlik dalgalanmalar M, konjonktürel dalgalanmalar K ve düzensiz dalgalanmalar D harfleri ile ifade edilecektir. Bu faktörlerin değişken üzerindeki etkileri toplamsal, çarpımsal veya ikisinin çeşitli birleşimleri şeklinde açıklanabilmektedir.

Toplamsal ilişki,

$$Y_i = T_i + M_i + K_i + D_i$$

Çarpımsal ilişki ise,

$$Y_i = T_i \times M_i \times K_i \times D_i$$

şeklinde ifade edilebilir. Yapılan araştırmalar faktörlerin çarpımlarının daha doğru sonuç verdiğini ortaya koymuştur. Verilen ifadelerde T_i sayısal değerler, M_i, K_i ve D_i ise birer orandır. İncelenen zaman serisi yıllık ise mevsimlik dalgalanmalar söz konusu olamayacağından ilişki,

$$Y_i = T_i \times K_i \times D_i$$

şeklini alacaktır.

15.2. TRENDİN BELİRLENMESİ

Serilerin trendleri doğrusal veya eğrisel olabilir ve trendin belirlenmesi için farklı yöntemler kullanılabilir. Bu yöntemlerin kullanılması ile yapılacak analizlerde devre sayısının mümkün olduğunca çok olması, daha doğru sonuç alınmasını sağlayacaktır.

Bir serinin trendinin belirlenmesinin en basit yolu serinin grafiğini çizmektir. Bu şekilde trendin genel karakteri belirlenebilirse de kesin ve güvenilir sonuç elde edilemeyecektir. Bu bölümde trendin belirlenmesi için kullanılan diğer yöntemler açıklanacaktır.

15.2.1. Yarı Ortalamalar Yöntemi

Bu yöntemde seri iki eşit parçaya bölünerek, bölümlerin ayrı ayrı aritmetik ortalamaları hesaplanır. Hesaplanan aritmetik ortalamalar, grafik üzerinde uygun devrelerin karşılığı olarak işaretlenerek bu noktaların bir doğru ile birleştirilmesi ile trend doğrusu belirlenmiş olur.

Bu yöntemde devre sayısı çift sayı ise seri iki tam eşit parçaya bölünür, devre sayısı tek sayı ise ortadaki devre dikkate alınmaz ve bu şekilde bölünen serinin bö-

lümlerinin aritmetik ortalamaları hesaplanır. Serinin bölümlerinin aritmetik ortalamalarının grafik üzerinde gösterilmesi önemlidir. Aritmetik ortalama serinin yarısının orta devresinin karşılığı olarak işaretlenecektir. Serinin bölümlerinin çift sayıda olması durumunda aritmetik ortalama her yarının orta devresinin karşısında, serinin bölümlerinin tek sayıda olması durumunda ise aritmetik ortalama serinin yarılarının ortasında yer alan iki devrenin ortasının karşısında yer alacaktır.

ÖRNEK: Aşağıda bir malın 6 aylık satış miktarları verilmiştir (1 = 1000 adet). Serinin trendini yarı ortalamalar yöntemi ile belirleyiniz.

Aylar	Satış Miktarları
1	13
2	$17 \rightarrow \overline{Y}_1 = 15$
3	15
4	21
5	$18 \rightarrow \overline{Y}_2 = 22$
6	27

Satış miktarları Y ile ifade edilecektir.

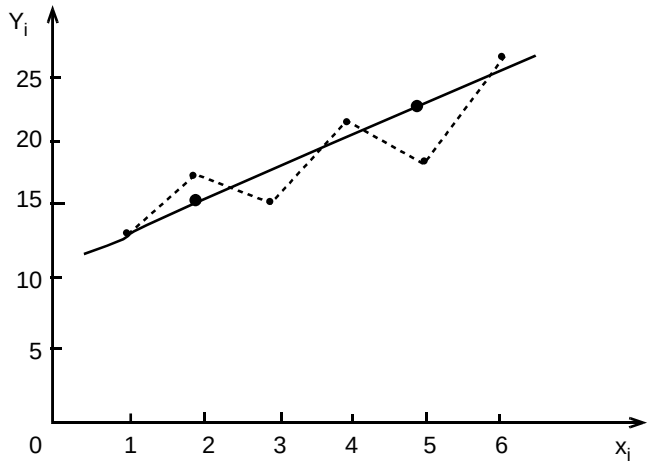
$$\frac{n}{2} = \frac{6}{2} = 3$$

$$\overline{Y}_1 = \frac{Y_1 + Y_2 + Y_3}{3}$$

$$= \frac{13 + 17 + 15}{3} = 15$$

$$\overline{Y}_2 = \frac{Y_4 + Y_5 + Y_6}{3}$$

$$= \frac{21 + 18 + 27}{3} = 22$$



ÖRNEK: Aşağıda bir işletmenin yıllık reel karları verilmiştir (1=1000 TL.). Serinin trendini yarı ortalamalar yöntemi ile belirleyiniz.

Yıllar	Reel Kar
2001	27
2002	29
	$\rightarrow \overline{Y}_1 = 25$
2003	24
2004	20
2005	22
2006	15
2007	16
	$\rightarrow \overline{Y}_2 = 13{,}25$
2008	12
2009	10

$$\frac{n}{2} = \frac{9}{2} = 4{,}5$$

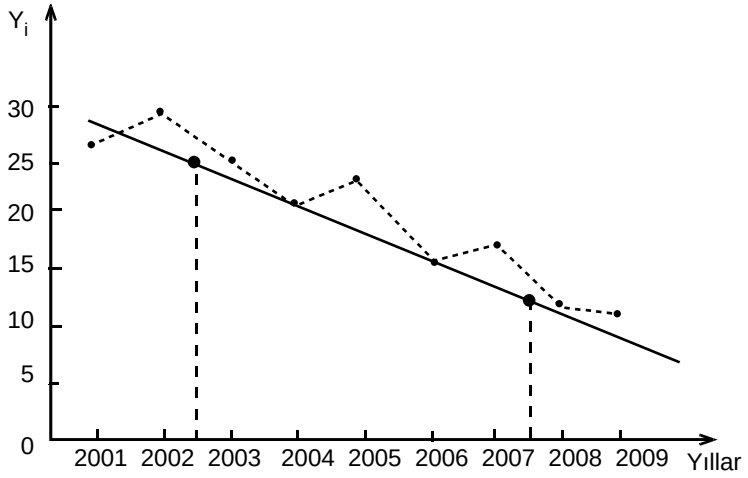
$$\overline{Y}_1 = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$$

$$= \frac{27 + 29 + 24 + 20}{4} = 25$$

Yıl sayısı tek olduğundan orta yıl, yani 2005 yılı dikkate alınmayacaktır.

$$\overline{Y}_2 = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4}$$

$$= \frac{15 + 16 + 12 + 10}{4} = 13{,}25$$



15.2.2. Hareketli Ortalamalar Yöntemi

Hareketli ortalamalar yönteminde serinin dalga boyu dikkate alınarak aritmetik ortalamalar hesaplanır. Dalga boyu serinin maksimum ve minimum değerleri arasındaki devre sayısıdır. Dalga boyunun devre sayısı, hesaplanacak aritmetik ortalamaların kapsayacağı devre sayısını belirler. Aritmetik ortalamaların hesaplamasına ilk devreden başlanarak, ilk birimin değerinin çıkarılıp bir sonraki birimin değerinin eklenmesi ile son birime kadar devam edilir. Yöntemden iyi sonuç alınması dalga boyunun ve şiddetinin eşit olmasına bağlıdır. Dalga boyları ve şiddetleri farklı ise bu yöntem iyi sonuç vermeyecektir. Uygulamada tüm dalga boylarının eşit olması durumuna az rastlanır. Dalga boylarının farklı olması durumunda dalga boylarının aritmetik ortalamaları alınacaktır.

Hesaplanacak hareketli ortalama sayısı, dalga boyunun tek veya çift sayı olmasına ve dalga boyunun uzunluğuna göre değişecektir. Dalga boyu tek sayıda ise dalga boyunun bir eksiği sayıda devre için hareketli ortalama hesaplanamaz. Hareketli ortalaması hesaplanamayacak devrelerin yarısı serinin başında, yarısı serinin sonunda yer alacaktır.

Dalga boyunu d harfi ile ifade edersek, d = 3 ise,

d-1 = 3-1 = 2

devre için hareketli ortalama hesaplanamaz. 2/2 = 1 olduğundan serinin ilk ve son devreleri için hareketli ortalamalar belirlenemeyecektir. d =3 için 3'erli hareketli ortalamalar şöyle hesaplanacaktır.

Devreler	Y_i	Hareketli Ortalamalar
1	Y_1	-
2	Y_2	$\bar{Y}_2 = \frac{Y_1 + Y_2 + Y_3}{3}$
3	Y_3	$\bar{Y}_3 = \frac{Y_2 + Y_3 + Y_4}{3}$
4	Y_4	$\bar{Y}_4 = \frac{Y_3 + Y_4 + Y_5}{3}$
⋮	⋮	⋮
n-1	Y_{n-1}	$\bar{Y}_{n-1} = \frac{Y_{n-2} + Y_{n-1} + Y_n}{3}$
n	Y_n	-

Dalga boyu d = 5 ise

d-1 = 5-1 = 4

devre için hareketli ortalama hesaplanamaz ve Y/2 = 2 olduğundan ilk ve son iki devre için hareketli ortalamalar belirlenemeyecektir. Bu durumda hareketli ortalamalar şöyle olacaktır.

Devreler	Y_i	Hareketli Ortalamalar
1	Y_1	-
2	Y_2	-
3	Y_3	$\bar{Y}_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5}$
4	Y_4	$\bar{Y}_4 = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5}$
5	Y_5	$\bar{Y}_5 = \frac{Y_3 + Y_4 + Y_5 + Y_6 + Y_7}{5}$
⋮	⋮	⋮
n-2	Y_{n-2}	$\bar{Y}_{n-1} = \frac{Y_{n-4} + Y_{n-3} + Y_{n-2} + Y_{n-1} + Y_n}{5}$
n-1	Y_{n-1}	-
n	Y_n	-

ÖRNEK: Aşağıda verilen serinin trendini hareketli ortalamalar yöntemi ile belirleyiniz.

Devreler	Y_i	$\overline{Y}$
1	2	-
2	3	3,66
3	6	4,66
4	5	5,66
5	6	6,33
6	8	7,00
7	7	8,00
8	9	8,66
9	10	-

Serinin değerleri 3, 6. ve 9. devrelerde maksimum yapmıştır. Bu nedenle dalga boyu d = 3'tür.

d –1 = 3-1 = 2

devrenin hareketli ortalama değeri hesaplanamayacaktır. Bu devreler
1. ve 9. devrelerdir. n –(d-1) = 9 –(3-1) = 7 hareketli ortalama hesaplanacaktır.

$$\overline{Y}_2 = \frac{Y_1 + Y_2 + Y_3}{3} = \frac{2+3+6}{3} = \frac{11}{3} = 3{,}66$$

$$\overline{Y}_3 = \frac{Y_2 + Y_3 + Y_4}{3} = \frac{3+6+5}{3} = \frac{14}{3} = 4{,}66$$

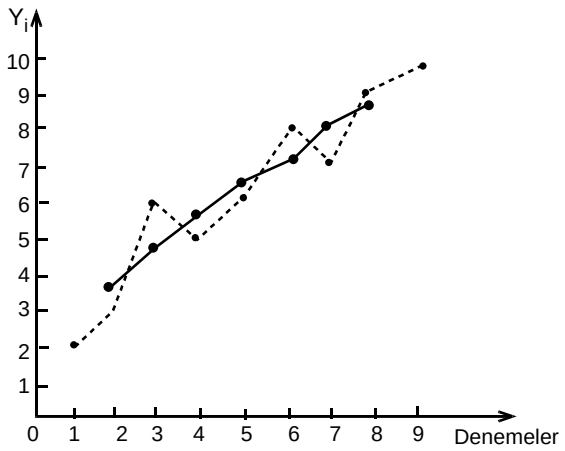
$$\overline{Y}_4 = \frac{Y_3 + Y_4 + Y_5}{3} = \frac{6+5+6}{3} = \frac{17}{3} = 5{,}66$$

$$\overline{Y}_5 = \frac{Y_4 + Y_5 + Y_6}{3} = \frac{5+6+8}{3} = \frac{19}{3} = 6{,}33$$

$$\overline{Y}_6 = \frac{Y_5 + Y_6 + Y_7}{3} = \frac{6+8+7}{3} = \frac{21}{3} = 7{,}00$$

$$\overline{Y}_7 = \frac{Y_6 + Y_7 + Y_8}{3} = \frac{8+7+9}{3} = \frac{24}{3} = 8{,}00$$

$$\overline{Y}_8 = \frac{Y_7 + Y_8 + Y_9}{3} = \frac{7+9+10}{3} = \frac{26}{3} = 8{,}66$$



Dalga boyu çift sayıda ise, yine dalga boyunun bir eksiği sayıda hareketli ortalama hesaplanamayacak ve bunların yarısı serinin başında yarısı sonunda yeralacaktır. Dalga boyu =4 ise,d –1 = 4-1 = 3 devre için hareketli ortalama hesaplanamaz ve 3/2 = 1,5 olduğundan serinin başından ve sonundan 1,5'er devrenin hareketli ortalama değerleri belirlenemeyecektir. Bu durumda hesaplanacak hareketli ortalamalar devrelerin ortalamalarında yer alacaklardır. Bu nedenle ardarda gelen hareketli ortalamaların ikişer ikişer ortalamaları alınarak düzeltme yapılır. Düzeltme yapıldıktan sonra dalga boyu sayıda devre için hareketli ortalama değeri hesaplanamamış olacaktır. Bunların yarısı serinin başında, yarısı serinin sonunda yer alacaktır. Dalga boyu d =4 ise hesaplanacak hareketli ortalamalar şöyle olacaktır.

Devreler	Y_i	Hareketli Ortalamalar	Düzeltilmiş Hareketli Ortalamalar
1	Y_1	-	
2	Y_2	$\overline{Y}_{2,5} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$	
3	Y_3	$\overline{Y}_{3,5} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4}$	$\overline{Y}_3 = \frac{Y_{2,5} + Y_{3,5}}{2}$
4	Y_4	$\overline{Y}_{4,5} = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4}$	$\overline{Y}_4 = \frac{Y_{3,5} + Y_{4,5}}{2}$
5	Y_5		
⋮	⋮	⋮	
⋮	⋮	⋮	
n-2	Y_{n-2}	$\overline{Y}_{n-2,5} = \frac{Y_{n-4} + Y_{n-3} + Y_{n-2} + Y_{n-1}}{4}$	$\overline{Y}_{n-3} = \frac{Y_{n-3,5} + Y_{n-2,5}}{2}$
			$\overline{Y}_{n-2} = \frac{Y_{n-2,5} + Y_{n-1,5}}{2}$

n-1	Y_{n-1}	$\bar{Y}_{n-1,5} = \frac{Y_{n-3} + Y_{n-2} + Y_{n-1} + Y_n}{4}$
n	Y_n	-

Dalga boyu uzunlukları 6, 8, 10,12 .... gibi çift sayıda olduğunda da aynı işlemlerin yapılması gerekecektir.

ÖRNEK: Aşağıda verilen serinin trendinin hareketli ortalamalar yöntemi ile belirleyiniz.

Devreler	Y_i	Hareketli Ortalamalar	Düzeltilmiş Hareketli Ortalamalar
1	3	-	-
2	4	-	-
		5,25	-
3	6	6,25	5,75
4	8	7,25	6,75
5	7	8,25	7,75
6	8	9,25	8,75
7	10	-	-
8	12	-	-

Serinin değerleri 4. ve 8. devrelerde maksimum yaptığı için dalga boyu 4'tür.

d –1 = 4-1= 3

devre için hareketli ortalama hesaplanamaz. Bunların yarısı seri başında yarısı seri sonunda yer alacaktır.

n –(d-1) = 8- (4-1) = 5

hareketli ortalama değeri hesaplanacaktır.

$$\bar{Y}_{2,5} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{3+4+6+8}{4} = \frac{21}{4} = 5{,}25$$

$$\bar{Y}_{3,5} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{4+6+8+7}{4} = \frac{25}{4} = 6{,}25$$

$$\bar{Y}_{4,5} = \frac{Y_3 + Y_4 + Y_5 + Y_6}{4} = \frac{6+8+7+8}{4} = \frac{29}{4} = 7{,}25$$

$$\bar{Y}_{5,5} = \frac{Y_4 + Y_5 + Y_6 + Y_7}{4} = \frac{8+7+8+10}{4} = \frac{33}{4} = 8{,}25$$

$$\bar{Y}_{6,5} = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4} = \frac{7+8+10+12}{4} = \frac{37}{4} = 9{,}25$$

Düzeltilmiş Hareketli Ortalamalar

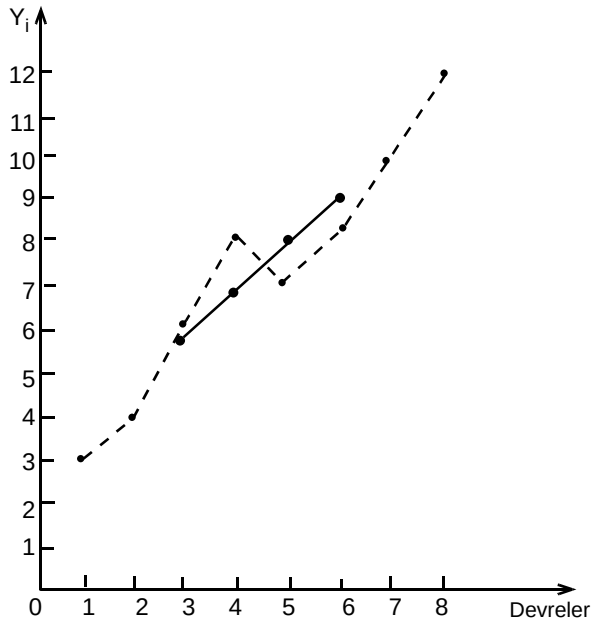
Düzeltme işleminden sonra dalga boyu sayısında, yani 4 devre için hareketli ortalama değeri olmayacaktır. Bunlar serinin ilk ve son iki devresidir.

$$\bar{Y}_3 = \frac{Y_{2,5} + Y_{3,5}}{2} = \frac{5,25 + 6,25}{2} = \frac{11,50}{2} = 5,75$$

$$\bar{Y}_4 = \frac{Y_{3,5} + Y_{4,5}}{2} = \frac{6,25 + 7,25}{2} = \frac{13,50}{2} = 6,75$$

$$\bar{Y}_5 = \frac{Y_{4,5} + Y_{5,5}}{2} = \frac{7,25 + 8,25}{2} = \frac{15,50}{2} = 7,75$$

$$\bar{Y}_6 = \frac{Y_{5,5} + Y_{6,5}}{2} = \frac{8,25 + 9,25}{2} = \frac{17,50}{2} = 8,75$$



Dalga boyu çift ise, hareketli ortalamalarda düzeltme işlemi ilk hesaplama sırasında da yapılabilir. Bunun için her hareketli ortalama, dalga boyundan bir fazla devre için hesaplanır. Hesaplamada hareketli ortalamanın hesaplanmasında kullanılan ilk ve son devre değerlerinin yarısı alınarak dalga boyuna bölünür.

Bu durumda dalga boyu d = 4 ise, 4. devre için, hareketli ortalama hesaplanamaz, bunlar serinin başında ve sonunda yer alan d/2 = 2'şer devredir. Hesaplanacak hareketli ortalamalar ise,

Devreler	Y_i	Hareketli Ortalamalar
1	Y_1	-
2	Y_2	-
3	Y_3	$\bar{Y}_3 = \dfrac{\frac{Y_1}{2} + Y_2 + Y_3 + Y_4 + \frac{Y_5}{2}}{4}$
4	Y_4	$\bar{Y}_4 = \dfrac{\frac{Y_2}{2} + Y_3 + Y_4 + Y_5 + \frac{Y_6}{2}}{4}$
⋮	⋮	⋮
n-2	Y_{n-2}	$\bar{Y}_{n-2} = \dfrac{\frac{Y_{n-4}}{2} + Y_{n-3} + Y_{n-2} + Y_{n-1} + \frac{Y_n}{2}}{4}$
n-1	Y_{n-1}	-
n	Y_n	-

olacaktır. Diğer çift sayıdaki dalga boyları için de aynı uygulama yapılacaktır.

ÖRNEK: Daha önce verilen örnek için hareketli ortalamaları hesaplayınız.

Devreler	Y_i	Hareketli Ortalamalar
1	3	-
2	4	-
3	6	5,75
4	8	6,75
5	7	7,75
6	8	8,75
7	10	-
8	12	-

$$\bar{Y}_3 = \frac{\frac{Y_1}{2} + Y_2 + Y_3 + Y_4 + \frac{Y_5}{2}}{4} = \frac{\frac{3}{2} + 4 + 6 + 8 + \frac{7}{2}}{4} = \frac{23}{4} = 5{,}75$$

$$\bar{Y}_4 = \frac{\frac{Y_2}{2} + Y_3 + Y_4 + Y_5 + \frac{Y_6}{2}}{4} = \frac{\frac{4}{2} + 6 + 8 + 7 + \frac{8}{2}}{4} = \frac{27}{4} = 6{,}75$$

$$\bar{Y}_5 = \frac{\frac{Y_3}{2} + Y_4 + Y_5 + Y_6 + \frac{Y_7}{2}}{4} = \frac{\frac{6}{2} + 8 + 7 + 8 + \frac{10}{2}}{4} = \frac{31}{4} = 7{,}75$$

$$\bar{Y}_6 = \frac{\frac{Y_4}{2} + Y_5 + Y_6 + Y_7 + \frac{Y_8}{2}}{4} = \frac{\frac{8}{2} + 7 + 8 + 10 + \frac{12}{2}}{4} = \frac{35}{4} = 8{,}75$$

15.2.3. Trendin Matematiksel Modeller ile Belirlenmesi

Regresyon analizinde değişkenler arasındaki ilişki matematiksel bir model ile açıklanıyordu. Trend, zaman serisinin değişkeninin uzun devre eğilimi olduğundan, regresyonda olduğu gibi matematiksel bir model ile belirlenebilir. Bu durumda değişkendeki değişmeler zaman değişkeni tarafından açıklanacağından, zamanın bağımsız değişken olduğu basit regresyon analizi sözkonusu olacaktır.

Trendin belirlenmesi için daha önce basit regresyonda verdiğimiz formüller kullanılacaktır. En küçük kareler yöntemi ile basit doğrusal regresyon için normal denklemler,

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

olarak elde edilmiş ve birinci denklemden $\hat{\alpha}$ 'in çekilerek ikinci denklemde yerine konulması ile,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}$$

bulunmuştu. Bu durumda sabit katsayı,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

olarak tahmin edilmekteydi.

Verilen bu formüller doğrusal trendin belirlenmesi için kullanılırken X_i, yani bağımsız değişken zamanı ifade edecektir. Eğer zaman serisi yıllık ise, yıllar büyük rakamlar olacağından bunların yerine değerler verilerek matematiksel işlemler kolaylaştırılabilir. Zaman değişkeni yerine değerler verilirken yılların birer birer artışı dikkate alınacaktır. X'e verilen değerler birer birer artmayabilir, fakat zaman gibi eşit miktarlarda artması gerekir. Ayrıca verilen değerler ne kadar küçük tamsayı değerler olursa matematiksel işlemler o kadar azalacaktır. Bu nedenle X'e 0 veya 1'den başla-

narak birer birer artacak şekilde değerler verilmektedir. (X=0,1,2,...,n-1 veya X= 1,2,3,....,n).

ÖRNEK: Aşağıda verilen serinin doğrusal trend modelinin parametrelerini tahmin ediniz.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2005	3	0	0	0
2006	6	1	6	1
2007	5	2	10	4
2008	7	3	21	9
2009	9	4	36	16
	30	10	73	30

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

$$\overline{X} = \frac{\sum X}{n} = \frac{10}{5} = 2$$

$$\overline{Y} = \frac{\sum Y}{n} = \frac{30}{5} = 6$$

$$\hat{\beta} = \frac{73 - 5(2)(6)}{30 - 5(2)^2} = \frac{13}{10} = 1{,}3$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 6 - 1{,}3(2) = 3{,}4$$

$Y_i = 3{,}4 + 1{,}3X_i$

Aynı parametreleri X_i değerlerini 1'den başlatarak tahmin edelim,

Y_i	X_i	X_iY_i	X_i^2
3	1	3	1
6	2	12	4
5	3	15	9
7	4	28	16
9	5	45	25
30	15	103	55

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

$$\overline{X} = \frac{\sum X}{n} = \frac{15}{5} = 3$$

$$\overline{Y} = \frac{\sum Y}{n} = \frac{30}{5} = 6$$

$$\hat{\beta} = \frac{103 - 5(3)(6)}{55 - 5(3)^2} = \frac{13}{10} = 1{,}3$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$= 6 - 1{,}3(3) = 2{,}1$

$Y_i = 2{,}1 + 1{,}3X_i$

Dikkat edilirse X_i'ye verilen farklı değerler sadece sabit katsayıyı değiştirmiştir.

Trendi belirlemek için kullanılan doğrusal modelde X_i'nin katsayısı zaman biriminin artması ile değişkenin değerinin ne kadar artacağını veya azalacağını açıklamaktadır. Yukarıda verilen örnekte görüldüğü gibi X_i'nin farklı değerler alması ile $\hat{\beta}$ katsayısı değişmemiştir. Yalnız X_i'nin değerleri birer birer artmadığı durumlarda ise, X_i'nin katsayısının değeri değişecektir.

ÖRNEK: Daha önce verilen örnek için X_i=0,2,.... şeklinde değerler vererek katsayıları tahmin ediniz.

Y_i	X_i	X_iY_i	X_i^2
3	0	0	0
6	2	12	4
5	4	20	16
7	6	42	36
9	8	72	64
30	20	146	120

$$\overline{Y} = 6$$

$$\overline{X} = \frac{\sum X}{n} = \frac{20}{5} = 4$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2}$$

$$\hat{\beta} = \frac{146 - 5(4)(6)}{120 - 5(4)^2} = \frac{26}{40} = 0{,}65$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$= 6 - 0{,}65(4) = 3{,}4$$

$$Y_i = 3{,}4 + 0{,}65X_i$$

Bu sonuç daha önce X_i =0,1,2,3,4 olarak alındığında elde edilen sonuç ile karşılaştırılırsa $\hat{\beta}$ katsayısının değiştiği $\hat{\alpha}$ katsayısının ise aynı kaldığı görülecektir. X_i, 2'şer 2'şer arttığında $\hat{\beta}$ katsayısı diğer katsayısının yarısına eşit bulunmuştur. Bu durumda $\hat{\beta}$ 6 aylık artış miktarını vermekte ve yıllık artış miktarının belirlenmesi için 2 ile çarpılması gerekmektedir.

Trendi belirlemek için kullanılan modellerde sabit katsayının matematiksel anlamı dışında bir anlamı olmayacaktır. Regresyon analizinde sabit katsayı bağımsız değişkenin sıfır değerini alması halinde bağımlı değişkenin alacağı değeri vermekteydi. Trend söz konusu olduğunda ise bağımsız değişken zaman olduğundan bu değişkenin sıfır değerini almasının bir anlamı olmayacaktır.

Trend modelini belirlemek için X_i değişkenine birer birer artan değerler verilerek eksen kaydırılması ve birden büyük değerler verilerek arttırılması ile ölçek değiştirilmesi yapılmaktadır. Değerler verilirken X_i'lerin toplamı sıfıra eşitlenirse katsayıların tahmini işlemlerinde basitleşme olacaktır. Bu durumda normal denklemlerden,

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}\sum_{i=1}^{n} X_i^2$$

$\sum_{i=1}^{n} X_i = 0$ olduğundan 1. Denklemden,

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha}$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

ve 2. denklemden,

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta} \sum_{i=1}^{n} X_i^2$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

olarak elde edilir.

X_i'lerin toplamının sıfıra eşitlenebilmesi için devre sayısı tek sayı ise ortadaki değere sıfır verilecektir. Serinin orta değerden önceki devreleri için birer birer veya daha farklı eşit aralıklı eksiltme, serinin orta değerinden sonraki devreleri için aynı aralıklı arttırma yapılacaktır. Bu durumda X_i'ler birer birer arttırılırsa, X_i = ...-3,-2,-1,0,1,2,3,... şeklinde, ikişer ikişer arttırılırsa ....,-6,-4,-2,0,2,4,6,.... olacaktır. Birer artış işlemleri daha basitleştireceğinden uygulamada tercih edilmektedir.

ÖRNEK: Daha önce verilen seri için $\sum X_i = 0$ olacak şekilde trend modelinin katsayılarını tahmin ediniz.

Y_i	X_i	X_iY_i	X_i^2
3	-2	-6	4
6	-1	-6	1
5	0	0	0
7	1	7	1
9	2	18	4
30	0	13	10

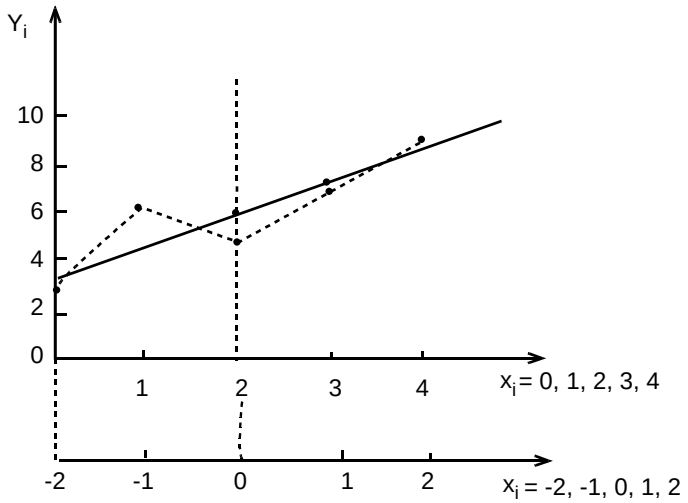
$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{30}{5} = 6$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{13}{10} = 1{,}3$$

$Y_i = 6 + 1{,}3X_i$

Görüldüğü gibi yıllık değişimi gösteren $\hat{\beta}$ katsayısı aynı kalmış, eksen kaydırılması nedeni ile sabit katsayı değişmiştir.

X= 0 için Y=6 ve X=1 için Y=7,3 alınarak çizilen trend grafiği Şekil 15.1.'de görülmektedir.



Şekil 15.1

Şekil 15.1'de iki X_i ekseni kullanılarak çizilen grafikte doğrunun eğiminin aynı kaldığı, orjinin 0'dan 2'ye kaydırıldığında kesim noktasının 3,4'ten 6'ya çıktığı görülmektedir.

Serinin devre sayısı çift sayıda olursa X_i'lere ....,-3,-2,-1,1,2,3,.... şeklinde değer verilmesi hatalı olacaktır. Çünkü X_i diğer devrelerde birer birer artarken –1,+1 aralığında 2 birim artmaktadır. Bu nedenle X_i'leri ikişer veya daha büyük aralıklar ile

arttırmak gerekecektir. Eğer X_i'ler ikişer ikişer artacaksa orta devrelere –1 ve +1 değerleri verilecek ve X_i = ...-5,-3,-1,1,3,5,... şeklinde değerler alacaktır.

ÖRNEK: Aşağıda verilen serinin doğrusal trend modelinin katsayılarını $\sum X_i = 0$ alarak belirleyiniz.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2004	8	-5	-40	25
2005	5	-3	-15	9
2006	7	-1	-7	1
2007	6	1	6	1
2008	4	3	12	9
2009	3	5	15	25
	33	0	-29	70

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{33}{6} = 5{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{-29}{70} = -0{,}414$$

$Y_i = 5{,}5 - 0{,}414X_i$

Devre sayısı çift sayı olduğundan X_i'ler ikişer ikişer artmıştır. Bu nedenle $\hat{\beta}$; 6 aylık artış miktarıdır. Aynı örneği X_i'ye 0'dan başlayarak değerler vererek çözelim.

Y_i	X_i	X_iY_i	X_i^2
8	0	0	0
5	1	5	1
7	2	14	4
6	3	18	9
4	4	16	16
3	5	15	25
33	15	68	55

$$\overline{X} = \frac{\sum X_i}{n} = \frac{15}{6} = 2{,}5$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{33}{6} = 5{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n(\bar{X})^2}$$

$$= \frac{68 - 6(2{,}5)(5{,}5)}{55 - 6(2{,}5)^2}$$

$$= \frac{-14{,}5}{17{,}5} = -0{,}828$$

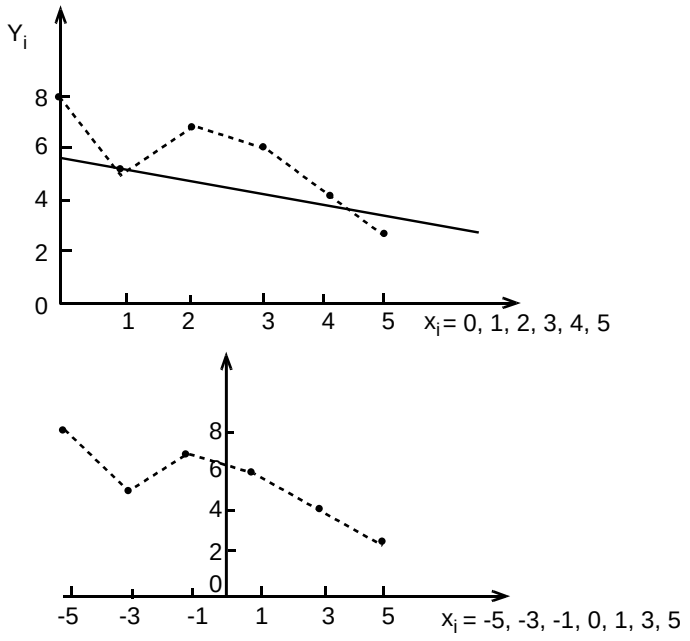
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$= 5{,}5 - (-0{,}828)(2{,}5) = 7{,}57$$

$$Y_i = 7{,}57 - 0{,}828X_i$$

Görüldüğü gibi $\hat{\alpha}$ değişmiş, $\hat{\beta}$ ise eski $\hat{\beta}$ 'in iki katına çıkmıştır.

1. modelde X=0 için Y=5,5 ve X=1 için Y=5,086 ve 2.modelde X=0 için Y=7,57 ve X=1 için Y=6,742 alınarak çizilen grafikler Şekil 15.2.'de görülmektedir.



Şekil 15.2

Şekil 15.2.'de de görüldüğü gibi hem eksen hem de ölçek değiştiğinde doğrunun eğimi de değişmiştir.

Parametreleri tahmin edilen trend modelleri için parametre tahminlerinin varyansları ve belirlilik katsayıları basit regresyonda verilen formüller ile hesaplanacaktır. Burada belirlilik katsayısı zaman değişkeninin değişkendeki değişmeleri açıklama oranını göstermektedir.

Trend modelleri için teorik değerler regresyon analizinde olduğu gibi X'e verilen değerlerin modelde tek tek yerine konması ile bulunacaktır. Ayrıca hata terimlerinin kareleri toplamının hesaplanması için verilen formüller burada da geçerlidir.

Trend modelleri doğrusal olduğu gibi eğrisel de olabilir. Eğrisel trend regresyon analizinde verilen formüller ile aynı şekilde belirlenebilir. Burada önemli olan tam logaritmik model ile X'in logaritmasının alındığı yarı logaritmik modelde X'e sıfır veya negatif değer verilmemesi gerektiğidir. Bu durumda hesaplama yapılamayacaktır.

15.2.3.1. Trend Modelleri İle Gelecek Devre Değerlerinin Tahmini

Basit regresyonda gelecek devrelere ait değerlerin tahmin edilebilmesi için bağımsız değişkenin değerlerinin tahmin edilmesi gerekiyordu. Trend söz konusu olduğunda bağımsız değişken zaman değişkeni olduğundan, değişkenin değeri, tahmin edilecek devrenin zaman değişkeni değerinin modelde yerine konması ile tahmin edilecektir. Eğer eksen ve ölçek değiştirilmesi yapılmış ise değeri tahmin edilecek devreye ait X_i değerinin belirlenerek modelde yerine konması gerekecektir. Örneğin, 2003-2009 yılları için X_i'lere $\sum X_i = 0$ olacak şekilde değer verilerek model belirlenmiş ise X_i değerleri,

Yıllar	2003	2004	2005	2006	2007	2008	2009
X_i	-3	-2	-1	0	1	2	3

olacaktır. 2010 yılı değerleri tahmin edilecek ise X_i =4, 2011 yılı değeri tahmin edilecek ise X_i=5 olarak alınacaktır. Aynı şekilde 2002-2009 yılları için $\sum X_i = 0$ olarak şekilde değer verilerek, model belirlenmiş ise X_i değerleri,

Yıllar	2002	2003	2004	2005	2006	2007	2008	2009	2010
X_i	-7	-5	-3	-1	1	3	5	7	9

olacaktır. Bu durumda 2010 yılı değerinin tahmini için X_i =9, 2011 yılı değerinin tahmini için X_i =11 olarak alınacaktır.

X_i'lere 0 veya 1'den başlanarak birer birer artan değerler verilmesi durumunda ise gelecek devreler için X_i'ler birer birim arttırılacaktır. Örneğin, 2003-2009 yılları için X_i= 0,1,2,3,4,5,6 olacağından 2010 yılı için X_i=7 ve 2011 yılı için X_i =8 olacaktır.

ÖRNEK: Daha önce verilen ve X_i'lere farklı şekillerde değerler verilerek parametreleri tahmin edilen örnekler için 2010 ve 2011 yılı Y_i değerlerini tahmin ediniz.

İlk verilen 5 yıllık örnekte,

X_i= 0,1,2,3,4 verilerek

$Y_i = 3{,}4 + 1{,}3X_i$

bulunmuştu.

2010 yılında X_i= 5 olacağından,

$\hat{Y}_{10} = 3{,}4 + 1{,}3(5) = 9{,}9$

ve 2011 yılında X_i= 6 olacağından,

$\hat{Y}_{11} = 3{,}4 + 1{,}3(6) = 11{,}2$

olacaktır. Aynı örnek için X_i= 1,2,3,4,5 verilerek

$Y_i = 2{,}1 + 1{,}3X_i$

bulunmuştu.

2010 yılında X_i= 6 olacağından,

$\hat{Y}_{10} = 2{,}1 + 1{,}3(6) = 9{,}9$

ve 2011 yılında X_i= 7 olacağından,

$\hat{Y}_{11} = 2{,}1 + 1{,}3(7) = 11{,}2$

olacaktır. Aynı örnek için X_i= 0,2,4,6,8 verilerek

$Y_i = 3{,}4 + 0{,}65X_i$

bulunmuştu.

2010 yılında X_i= 10 olacağından,

$\hat{Y}_{10} = 3{,}4 + 0{,}65(10) = 9{,}9$

ve 2011 yılında X_i= 12 olacağından,

$\hat{Y}_{11} = 3{,}4 + 0{,}65(12) = 11{,}2$

olacaktır. Aynı örnek için X_i'ler toplamı sıfıra eşitlenerek (X_i= -2,-1,0,1,2 verilerek)

$Y_i = 6 + 1{,}3X_i$

bulunmuştu.

2010 yılında X_i= 3 olacağından,

$\hat{Y}_{10} = 6 + 1{,}3(3) = 9{,}9$

ve 2011 yılında X_i= 4 olacağından,

$\hat{Y}_{11} = 6 + 1{,}3(4) = 11{,}2$

olacaktır.

İkinci olarak verilen örnekte

X_i= -5,-3,-1,1,3,5 alınarak,

$Y_i = 5{,}5 - 0{,}414X_i$

bulunmuştu.

2010 yılında X_i= 7 olacağından,

$\hat{Y}_{10} = 5{,}5 - 0{,}414(7) = 2{,}602$

ve 2011 yılında X_i= 9 olacağından,

$\hat{Y}_{11} = 5{,}5 - 0{,}414(9) = 1{,}774$

olacaktır.

Aynı örnek için X_i= 0,1,2,3,4,5 verilerek

$Y_i = 7{,}57 - 0{,}828X_i$

bulunmuştu.

2010 yılında X_i= 6 olacağından,

$\hat{Y}_{10} = 7{,}57 - 0{,}828(6) = 2{,}602$

ve 2011 yılında X_i= 7 olacağından,

$\hat{Y}_{11} = 7{,}57 - 0{,}828(7) = 1{,}774$

olacaktır.

Görüldüğü gibi her iki örnekte de X'e verilen farklı değerler ile aynı yıllar için yapılan tahmin sonuçları birbirine eşit olarak bulunmuştur.

15.2.3.2. Aylık Değerlerin Hesaplanması

Yıllık, altı aylık, üç aylık, aylık gibi devreler için düzenlenmiş zaman serilerinden aylık veya daha kısa zaman birimleri için teorik değerler hesaplanabilir. Burada yıllık serilerden aylık değerlerin hesaplanması açıklanacaktır.

Serinin trendine bağlı olarak aylık değerler yıl başından yıl sonuna doğru artarlar veya azalırlar. Serinin trendi artan ise, yani $\hat{\beta}$'nın işareti pozitif ise aylık değerler yıl başından yıl sonuna kadar artacak trend azalıyor ise, yani $\hat{\beta}$'nın işareti negatif ise aylık değerler yıl başından yıl sonuna doğru azalacaklardır.

Aylık değerlerin hesaplanmasında yıllık değerlerin aylık değerlerin toplamından veya aylık değerlerin ortalamasından oluşması ve zaman değişkenine verilen değerlerin birer birer veya ikişer ikişer artması önemlidir. Zaman serisi aylık değerlerin ortalamasından oluşuyorsa hesaplanacak değer aylık olacak ve bu yıl ortası değeri olarak kabul edilecektir. Zaman serisi aylık değerlerin toplamından oluşuyorsa bu durumda aylık ortalama değer alınarak yıl ortası değeri olarak kabul edilecektir. Zaman değişkenine verilen değerlerin artış miktarları ise aylık artış veya azalış miktarlarının belirlenmesi için önemlidir.

Yapılacak hesaplamalarda yıllık teorik değerlerin yıl ortasına, yani 30 Haziran'a aylık teorik değerlerin ise ay ortasına yani her ayın 15'ine ait oldukları kabul edilecektir. Yıl ortasının yılı, ay ortasının ayı temsil edeceği kabul edildiğinden 15 günlük artış veya azalış miktarlarının da hesaplanması gerekebilecektir.

Serinin aylık değerlerin toplamından veya ortalamasından oluşmasına ve zaman değişkenine verilen değerlerin birer birer veya ikişer ikişer artmasına göre yapılacak hesaplamalar şöyle oluşacaktır.

- Yıllık değerlerin aylık değerlerin ortalaması olması

Bu durumda X_i değerleri birer birer artıyor ise $\hat{\beta}$ yıllık değişim (artış veya azalış) miktarlarını verecektir. Aylık değişim miktarının belirlenmesi için yıllık değişim miktarı ay sayısına, yani 12'ye bölünecek ve aylık artış miktarı $\hat{\beta}$/12 olacaktır. Aylık teorik değerler aylık değişim miktarının ve trendin artan veya azalan olması da dikkate alınarak belirlenecektir.

ÖRNEK: Aşağıda verilen seride yıllık değerlerin aylık değerlerin ortalamasından oluştuğunu kabul ederek 2008 yılı Mart ve Aralık aylarına ait teorik değerleri hesaplayınız.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2003	5	-3	-15	9
2004	7	-2	-14	4
2005	8	-1	-8	1
2006	8	0	0	0
2007	9	1	9	1
2008	8	2	16	4
2009	11	3	33	9
	56	0	21	28

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{56}{7} = 8$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{21}{28} = 0{,}75$$

$Y_i = 8 + 0{,}75\ X_i$

Yıllık Artış Miktarı : $\hat{\beta} = 0{,}75$

Aylık Artış Miktarı : $\frac{\hat{\beta}}{12} = \frac{0{,}75}{12} = 0{,}0625$

$\hat{Y}_{2008} = 8 + 0{,}75(2) = 9{,}5$ 30 Haziran 2008 değeri.

Mart ayı değeri 15 Mart değeri olacağından, 30 Haziran ile 15 Mart arasında 3,5 ay vardır. Artan trend söz konusu olduğundan ve Mart ayı yılın ilk yarısında, yani 30 Hazirandan önce yeraldığından 15 Mart değeri 30 Haziran değerinden küçük olacaktır. Bu nedenle 3,5 aylık değişim miktarı hesaplanarak 30 Haziran değerinden çıkartılacaktır.

$0{,}0625(3{,}5) = 0{,}21875$

15 Mart 2008 değeri: $9{,}5 - 0{,}21875 = 9{,}28125$

Aralık ayı değeri 15 Aralık değeri olacağından, 30 Haziran ile 15 Aralık arasında 5,5 ay vardır. Artan trend söz konusu olduğundan ve Aralık ayı yılın ikinci yarısında, yani 30 Hazirandan sonra yeraldığından 15 Aralık değeri 30 Haziran değerinden büyük olacaktır. Bu nedenle 5,5 aylık değişim miktarı hesaplanarak 30 Haziran değerine eklenecektir.

$0{,}0625(5{,}5) = 0{,}34375$

15 Aralık 2008 değeri: $9{,}5 + 0{,}34375 = 9{,}84375$

X_i değerlerinin 0 veya 1'den başlayarak birer birer arttırılarak verilmesi durumunda da daha önce açıklandığı gibi teorik değerler ve $\hat{\beta}$ değeri değişmeyeceğinden aynı işlemler yapılarak aynı sonuçlar elde edilecektir.

Yıllık değerlerin aylık değerlerin ortalaması olması durumunda X_i değerleri ikişer ikişer artıyor ise $\hat{\beta}$, 6 aylık değişim miktarını vereceğinden yıllık değişim miktarı 2 $\hat{\beta}$ ve aylık değişim miktarı ise 2 $\hat{\beta}$ /12 yani $\hat{\beta}$ /6 olacaktır.

ÖRNEK: Aşağıda verilen seride yıllık değerlerin aylık değerlerin ortalamasından oluştuğunu kabul ederek 2005 yılı Şubat ve Eylül aylarına ait teorik değerleri hesaplayınız.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2004	11	-5	-55	25
2005	8	-3	-24	9
2006	9	-1	-9	1
2007	10	1	10	1
2008	8	3	24	9
2009	5	5	25	25
	51	0	-29	70

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{51}{6} = 8{,}5$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{-29}{70} = -0{,}41$$

$Y_i = 8{,}5 - 0{,}41\ X_i$

Yıllık Azalış Miktarı : $2\hat{\beta} = 2(-0{,}41) = -0{,}82$

Aylık Azalış Miktarı : $\frac{\hat{\beta}}{6} = \frac{-0{,}41}{6} = -0{,}0683$

$\hat{Y}_{2005} = 8{,}5 - 0{,}41(-3) = 9{,}73$

15 Şubat - 30 Haziran arası 4,5 aydır ve azalan trend sözkonusu olduğundan Şubat ayı değeri 30 Haziran değerinden küçük olacaktır.

(-0,0683)(4,5) = -0,30735

15 Şubat 2005 değeri : 9,73 – (-0,30735) = 10,03735

15 Eylül - 30 Haziran arası 2,5 aydır ve azalan trend sözkonusu olduğundan Eylül ayı değeri 30 Haziran değerinden küçük olacaktır.

(-0,0683)(2,5) = 0,17075

15 Eylül 2005 değeri : 9,73 + (-0,17075) = 9,55925

- Yıllık değerlerin aylık değerlerin toplamı olması

Bu durumda X_i değerleri birer birer artıyor ise $\hat{\beta}$ yıllık değişim miktarını verecektir. Aylık değişim miktarı ise $\hat{\beta}$/12'nin ortalaması, yani $\hat{\beta}$/144 olacaktır. Burada yıllık teorik değeri aylık değere dönüştürmek için yıllık teorik değerin ortalamasının alınması, yani 12'ye bölünmesi gerekecektir.

ÖRNEK: Aşağıda verilen seride yıllık değerlerin aylık değerlerin toplamından oluştuğunu kabul ederek 2009 yılı Nisan ve Kasım aylarına ait teorik değerleri hesaplayınız.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2005	12	-2	-24	4
2006	9	-1	-9	1
2007	10	0	0	0
2008	8	1	8	1
2009	6	2	12	4
	45	0	-13	10

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{45}{5} = 9$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{-13}{5} = -1{,}3$$

$Y_i = 9 - 1{,}3\ X_i$

Yıllık Azalış Miktarı : $\hat{\beta} =$ -1,3

Aylık Azalış Miktarı : $\frac{\hat{\beta}}{144} = \frac{-1{,}3}{144} = -0{,}00902$

$\hat{Y}_{2009} = 9 - 1{,}3(2) = 6{,}4$

30 Haziran 2009 $= \frac{6{,}4}{12} = 0{,}53333$

15 Nisan - 30 Haziran arası : 2,5 aydır

(-0,00902) (2,5) = -0,02255

15 Eylül 2009 değeri : 0,5333 - (-0,02255) = 0,55588

15 Kasım - 30 Haziran arası : 4,5 aydır

(-0,00902)(4,5) = -0,04059

15 Kasım 2009 değeri : 0,5333 + (-0,04059) = 0,49274

Yıllık değerlerin aylık değerlerin toplamı olması durumunda X_i değerleri ikişer ikişer artıyor ise $\hat{\beta}$, 6 aylık değişim miktarını vereceğinden, yıllık değişim miktarı 2 $\hat{\beta}$ ve aylık değişim miktarı ise 2 $\hat{\beta}$ /144 = $\hat{\beta}$ / 72 olacaktır. Burada da yıllık teorik değerlerin aritmetik ortalaması alınarak aylık teorik değer bulunacaktır.

ÖRNEK: Aşağıda verilen seride yıllık değerlerin aylık değerlerin toplamından oluştuğunu kabul ederek 2006 yılı Mayıs ve Ekim aylarına ait teorik değerleri hesaplayınız.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2006	8	-3	-24	9
2007	9	-1	-9	1
2008	10	1	10	1
2009	14	3	42	9
	41	0	19	20

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{41}{4} = 10{,}25$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{19}{20} = 0{,}95$$

$Y_i = 10{,}25 + 0{,}95\ X_i$

Yıllık Azalış Miktarı : 2 $\hat{\beta}$ = 2(0,95) = 1,9

Aylık Azalış Miktarı : $\frac{\hat{\beta}}{72} = \frac{0{,}95}{72} = 0{,}01319$

$\hat{Y}_{2006}$ = 10,25 + 0,95(-3) = 7,4

30 Haziran 2006 = $\frac{7{,}4}{12} = 0{,}61666$

15 Mayıs – 30 Haziran arası : 1,5 ay

(0,01319)(1,5) = 0,01978

15 Mayıs değeri : 0,61666 – 0,01978 = 0,59688

15 Ekim – 30 Haziran arası : 3,5 ay

(0,01319)(3,5) = 0,04616

15 Ekim değeri : 0,61666 + 0,04616 = 0,66282

15.3. MEVSİMLİK DALGALANMALARIN BELİRLENMESİ

Mevsimlik dalgalanmalar mevsim indekslerinin hesaplanması ile belirlenebilmektedir. Mevsim indeksleri, daha önce verilen çarpımsal modellerde değişkenin değerlerinin trende oranlanması ile hesaplanabilir. Bu durumda,

$$Y = T \times M \times K \times D$$

$$\frac{Y}{T} = M \times K \times D$$

olacaktır. Bu şekilde hesaplanan mevsim indeksleri özellikle konjoktürün etkisini taşıdıklarından, değişkenin değerlerinin trende bölünmesi yeterli olmayacaktır. Hareketli ortalamalar trend ile birlikte konjonktürün de etkisini taşırlar. Değişkenin değerlerinin hareketli ortalamalara oranlanması, trend ve konjonktürün etkisini ortadan kaldıracaklarından daha iyi sonuç verecektir. Hareketli ortalamaların hesaplanması, uzun işlemlere ve seride devre kayıplarına neden olmasına rağmen, yukarıda sayılan nedenlerle burada sadece hareketli ortalamalara oranlama yöntemine yer verilecektir.

Aylık serilerden mevsim indeksinin hesaplanabilmesi için, değişkenin değerleri, hesaplanacak 12'şerli hareketli ortalamalara oranlanarak 100 ile çarpılacaktır. Bu durumda,

$$H \times O = T \times K$$

olduğundan

$$\frac{Y}{TxK}100 = \frac{TxKxMxD}{TxK}100 = M.D.100$$

olacaktır. Hesaplanacak hareketli ortalama çift sayıda olduğundan, serinin ilk ve son 6 ayı için hareketli ortalama hesaplanamayacaktır. Mevsim indeksleri, hesaplanan aylık oranların aylara göre alınan aritmetik ortalamalarıdır. Aylık seriler için mevsim indekslerinin toplamı 1200 olacaktır. Daha önce yapılan hesaplamalarda sonuçların ondalıkların tam alınamaması nedeni ile toplam 1200'den farklı çıkabilir. Bu durumda hesaplanan aylık ortalamalar,

1200 /Aylık Ortalamalar Toplamı

ile çarpılarak sonuçlar düzenlenecektir. Zaman serisi 2'şer aylık düzenlemiş ise 6'şarlı, 3'er aylık düzenlenmiş ise 4'erli hareketli ortalama vb. hesaplanacak ve diğer işlemler aynı şekilde yapılacaktır. Mevsim indeksleri toplamı seri 2'şer aylık düzenlenmiş ise 600, 3'er aylık düzenlenmiş ise 400, 4'er aylık düzenlenmiş ise 300 vb. olacaktır. Mevsim indeksleri toplamında hesaplamalardan doğan sapma varsa, düzeltme benzer şekilde yapılacaktır.

ÖRNEK: Aşağıda verilen seri için mevsim indekslerini hesaplayınız.

	Yıllar					
Aylık	2004	2005	2006	2007	2008	2009
1	42	49	61	58	33	67
2	45	61	70	60	52	72
3	59	64	82	69	50	83
4	51	72	85	77	68	81
5	77	85	92	85	73	94
6	83	93	95	97	96	97
7	88	90	99	86	92	95
8	75	81	88	88	98	90
9	63	85	76	76	81	82
10	48	64	65	58	77	76
11	40	53	52	41	68	70
12	37	46	49	35	54	53

Burada 12'şerli hareketli ortalamalar hesaplanacağından ilk 6 ve son 6 ay için hareketli ortalama değerleri bulunamayacak ve seride toplam 72 ay olduğundan 72 – 12 = 60 adet hareketli ortalamanın hesaplanması gerekecektir.

$$\bar{Y}_7 = \frac{\frac{Y_1}{2} + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + Y_9 + Y_{10} + Y_{11} + Y_{12} + \frac{Y_{13}}{2}}{12}$$

$$= \frac{\frac{42}{2} + 45 + 59 + 51 + 77 + 83 + 88 + 75 + 63 + 48 + 40 + 37 + \frac{49}{2}}{12}$$

$$= \frac{711{,}5}{12} = 59{,}29$$

$$\bar{Y}_8 = \frac{\frac{Y_2}{2} + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + Y_9 + Y_{10} + Y_{11} + Y_{12} + Y_{13} + \frac{Y_{14}}{2}}{12}$$

$$= \frac{\frac{45}{2} + 59 + 51 + 77 + 83 + 88 + 75 + 63 + 48 + 40 + 37 + 49 + \frac{61}{2}}{12}$$

$$= \frac{723}{12} = 60{,}25$$

$$\vdots$$

$$\vdots$$

$$\bar{Y}_{66} = \frac{\frac{Y_{60}}{2} + Y_{61} + Y_{62} + Y_{63} + Y_{64} + Y_{65} + Y_{66} + Y_{67} + Y_{68} + Y_{69} + Y_{70} + Y_{71} + \frac{Y_{72}}{2}}{12}$$

$$= \frac{\frac{54}{2} + 67 + 72 + 83 + 81 + 94 + 97 + 95 + 90 + 82 + 76 + 70 + \frac{53}{2}}{12}$$

$$= \frac{960,5}{12} = 80,04$$

			Hesaplanan	Hareketli	Ortalamalar	
Aylık	2004	2005	2006	2007	2008	2009
1	-	64,66	75,70	71,87	63,25	80,04
2	-	65,00	76,37	71,33	63,91	80,25
3	-	66,16	76,29	71,33	64,54	79,95
4	-	67,75	75,95	71,04	65,54	79,95
5	-	68,95	75,95	70,29	67,45	80,00
6	-	69,87	76,04	69,25	69,37	80,04
7	59,29	70,75	76,04	67,62	71,58	-
8	60,25	71,62	75,50	66,25	73,83	-
9	61,12	72,75	74,54	65,12	76,04	-
10	62,20	74,04	73,66	63,95	77,95	-
11	63,41	74,87	73,04	63,08	79,37	-
12	64,16	75,25	72,58	62,79	80,29	-

Hesaplanan bu değerlerden yaralanılarak mevsim indeksleri gerçek Y_i değerlerinin hareketli ortalama değerlerine bölünüp 100 ile çarpılması ile belirlenecektir.

$$I_i = \frac{Y_i}{\bar{\bar{Y}}} 100$$

$$I_{2004,7} = \frac{88}{59,29} 100 = 148,42$$

$$I_{2004,8} = \frac{75}{60,25} 100 = 124,48$$

$$\vdots$$

$$\vdots$$

$$I_{2009,6} = \frac{97}{80,04} 100 = 121,18$$

	Yıllar						Mevsimlik İndeksleri
Aylık	2004	2005	2006	2007	2008	2009	(Aylık Ortalamalar)
1	-	75,78	80,58	80,70	52,17	83,70	74,58
2	-	93,84	91,65	84,11	81,36	89,71	88,13
3	-	96,73	107,48	96,73	77,47	103,81	96,44
4	-	106,27	111,91	108,38	103,75	101,31	106,32
5	-	123,27	121,13	120,92	108,22	117,50	118,20
6	-	133,10	124,93	131,40	138,38	121,18	129,79
7	148,42	127,20	130,19	127,18	128,52	-	132,30
8	124,48	113,09	116,55	132,83	132,73	-	123,93
9	103,07	116,83	101,95	116,70	106,52	-	109,01
10	77,17	86,43	88,24	90,69	98,78	-	88,26
11	63,08	70,78	71,19	64,99	85,67	-	71,14
12	57,66	61,12	67,51	55,74	67,25	-	61,85
							1199,95

Aylık ortalamalar aylık oranların ortalamasıdır ve

1. ay için : $\frac{75,78+80,58+80,70+52,17+83,70}{5}=74,58$

2. ay için : $\frac{93,84+91,65+84,11+81,36+89,71}{5}=88,13$

$\vdots$

$\vdots$

12. ay için : $\frac{57,66+61,12+67,51+55,74+67,25}{5}=61,85$

olarak hesaplanmıştır. Aylık ortalamaların toplamı 1199,95 olarak bulunmuştur. Bu değer 1200'e çok yakın olduğundan aylık ortalamalarda düzeltme yapılmayabilir ve hesaplanan ortalamalar mevsim indekslerini vermektedir. Düzeltme yapılması gerekseydi,

1. ay için : $74,58\frac{1200}{1199,95}$

2. ay için : $88,13\frac{1200}{1199,95}$

$\vdots$

$\vdots$

12. ay için : $61,85\frac{1200}{1199,95}$

olarak yapılacaktı.

15.4. KONJONKTÜREL DALGALANMALARIN BELİRLENMESİ

Konjonktürel dalgalanmaların etkisi zaman serisini etkileyen diğer faktörlerin etkilerinin giderilmesi ile belirlenebilir. Seri yıllık ise trend ile konjonktürel ve düzensiz dalgalanmaların etkisi söz konusu olacaktır.

$Y = T \times K \times D$

Değişken değerleri trende oranlanırsa,

$$\frac{Y}{T} = K \times D$$

olacaktır. Elde edilen değerlerin 100 ile çarpımı ile konjoktürün etkisini gösteren oranlar elde edilmiş olacaktır.

Seri yıldan daha kısa zaman birimleri için düzenlenmiş ise mevsimlerin etkisinin de dikkate alınması gerekecektir. Bu durumda

$Y = T \times M \times K \times D$

olacağından, trend ve birlikte mevsimlerin etkisinin de giderilmesi gerekecektir.

$$\frac{Y}{TxM} = K \times D$$

Buluna sonuçlar yine 100 ile çarpılacaktır.

ÖRNEK: Aşağıda bir ülkenin yıllık ihracat miktarları verilmiştir. Konjonktürel dalgalanmaların etkisini belirleyiniz.

Yıllar	Y_i	X_i	X_iY_i	X_i^2	$\hat{Y}$	K.D
2001	18	-4	-72	16	19,2	93,75
2002	21	-3	-63	9	21,4	98,13
2003	29	-2	-58	4	23,6	122,88
2004	22	-1	-22	1	25,8	85,27
2005	27	0	0	0	28,0	96,42
2006	30	1	30	1	30,2	99,33
2007	35	2	70	4	32,4	108,02
2008	33	3	99	9	34,6	95,37
2009	37	4	148	16	36,8	100,54
	252	0	132	60		

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{252}{9} = 28$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{132}{60} = 2{,}2$$

$Y_i = 28 + 2{,}2X_i$

$\hat{Y}_{01} = 28 + 2{,}2(-4) = 19{,}2$
$\hat{Y}_{02} = 28 + 2{,}2(-3) = 21{,}4$
$\hat{Y}_{03} = 28 + 2{,}2(-2) = 23{,}6$
$\hat{Y}_{04} = 28 + 2{,}2(-1) = 25{,}8$
$\hat{Y}_{05} = 28 + 2{,}2(0) = 28$
$\hat{Y}_{06} = 28 + 2{,}2(1) = 30{,}2$
$\hat{Y}_{07} = 28 + 2{,}2(2) = 32{,}4$
$\hat{Y}_{08} = 28 + 2{,}2(3) = 34{,}6$
$\hat{Y}_{09} = 28 + 2{,}2(4) = 36{,}8$

$$KxD = \frac{Y}{T}100$$

$2001 \rightarrow \frac{18}{19{,}2}100 = 93{,}75$

$2002 \rightarrow \frac{21}{21{,}4}100 = 98{,}13$

$2003 \rightarrow \frac{29}{23{,}6}100 = 122{,}88$

$2004 \rightarrow \frac{22}{25{,}8}100 = 85{,}27$

$2005 \rightarrow \frac{27}{28{,}0}100 = 96{,}42$

$2006 \rightarrow \frac{30}{30{,}2}100 = 99{,}33$

$2007 \rightarrow \frac{35}{32{,}4}100 = 108{,}02$

$$2008 \rightarrow \frac{33}{34,6}100 = 95,37$$

$$2009 \rightarrow \frac{37}{36,8}100 = 100,54$$

15.5. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Aşağıda bir işletmenin 8 yıllık karları verilmiştir. (1= 1000 TL)

Yıllar	Yıllık Karlar
2002	13
2003	18
2004	21
2005	17
2006	26
2007	35
2008	52
2009	48

Bu serinin trendini yarı ortalamalar yöntemi ile belirleyiniz.

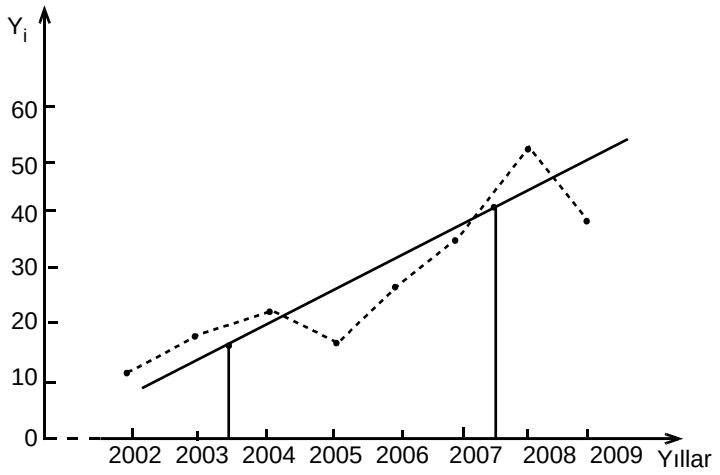
Çözüm:

$$\overline{Y}_1 = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$$

$$\overline{Y}_1 = \frac{13 + 18 + 21 + 17}{4} = \frac{69}{4} = 17,25$$

$$\overline{Y}_2 = \frac{Y_5 + Y_6 + Y_7 + Y_8}{4}$$

$$\overline{Y}_2 = \frac{26 + 35 + 52 + 48}{4} = \frac{161}{4} = 40,25$$



ÖRNEK 2: Aşağıda bir işletmenin aylık televizyon satışları verilmiştir.

Aylar	Satışlar (Adet)
Ocak	73
Şubat	65
Mart	67
Nisan	58
Mayıs	51
Haziran	53
Temmuz	42

Bu serinin trendini yarı ortalamalar yöntemi ile belirleyiniz.

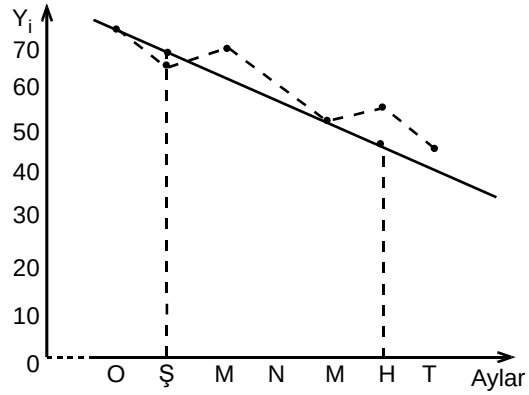
Çözüm:

$$\overline{Y_1} = \frac{Y_1 + Y_2 + Y_3}{3}$$

$$\overline{Y_1} = \frac{73 + 65 + 67}{3} = \frac{205}{3} = 68{,}33$$

$$\overline{Y_2} = \frac{Y_5 + Y_6 + Y_7}{3}$$

$$\overline{Y_2} = \frac{51 + 53 + 42}{3} = \frac{146}{3} = 48{,}66$$



ÖRNEK 3: Aşağıda bir ülkenin yıllık ithalatı verilmiştir. (1= 1 milyar $)

Yıllar	İthalat
2000	7,9
2001	8,9
2002	9,0
2003	9,2
2004	10,7
2005	9,1
2006	9,8
2007	11,1
2008	12,3
2009	13,2

Bu serinin trendini hareketli ortalamalar yöntemi ile belirleyiniz.

Çözüm:

Yi	H.O
7,9	-
8,9	-
9,0	9,14
9,2	9,38
10,7	9,56
9,1	9,98
9,8	10,60
11,1	11,10
12,3	-
13,2	-

d = 15 $\qquad \frac{5-1}{2} = 2$

Serinin başından ve sonundan 2'şer devre için hareketli ortalama hesaplanamaz.

$$\bar{Y}_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5} = \frac{7{,}9 + 8{,}9 + 9{,}0 + 9{,}2 + 10{,}7}{5} = 9{,}14$$

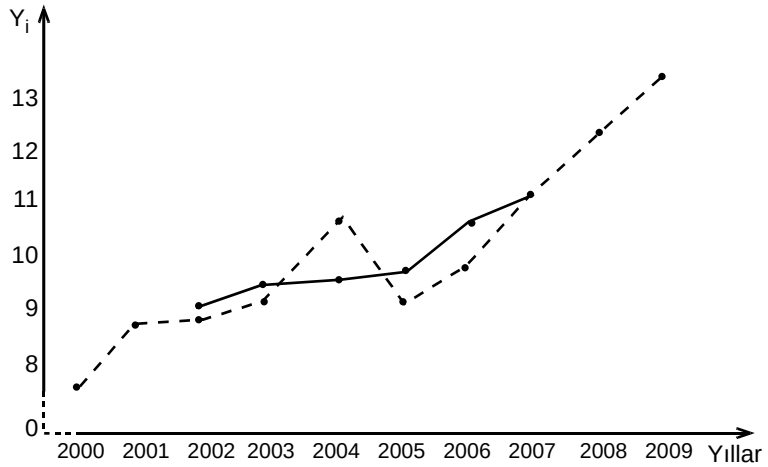
$$\bar{Y}_4 = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5} = \frac{8{,}9 + 9{,}0 + 9{,}2 + 10{,}7 + 9{,}1}{5} = 9{,}38$$

$$\bar{Y}_5 = \frac{Y_3 + Y_4 + Y_5 + Y_6 + Y_7}{5} = \frac{9{,}0 + 9{,}2 + 10{,}7 + 9{,}1 + 9{,}8}{5} = 9{,}56$$

$$\bar{Y}_6 = \frac{Y_4 + Y_5 + Y_6 + Y_7 + Y_8}{5} = \frac{9{,}2 + 10{,}7 + 9{,}1 + 9{,}8 + 11{,}1}{5} = 9{,}98$$

$$\bar{Y}_7 = \frac{Y_5 + Y_6 + Y_7 + Y_8 + Y_9}{5} = \frac{10{,}7 + 9{,}1 + 9{,}8 + 11{,}1 + 12{,}3}{5} = 10{,}60$$

$$\bar{Y}_8 = \frac{Y_6 + Y_7 + Y_8 + Y_9 + Y_{10}}{5} = \frac{9{,}1 + 9{,}8 + 11{,}1 + 12{,}3 + 13{,}2}{5} = 11{,}10$$



ÖRNEK 4: Aşağıda bir ülkenin aylık kömür üretim miktarları verilmiştir. (1=bin ton)

Aylar	Satışlar (Adet)
Ağustos	22
Eylül	23
Ekim	24
Kasım	29
Aralık	33
Ocak	41
Şubat	32
Mart	38
Nisan	40
Mayıs	43
Haziran	45
Temmuz	48

Bu serinin trendini hareketli ortalamalar yöntemi ile belirleyiniz.

Çözüm:

Y_i	HO
22	-
23	-
24	-
29	29,50
33	31,58
41	34,16
32	36,66
38	38,83
10	40,41
43	-
45	-
48	-

d = 6 $\qquad \frac{6}{2} = 3$

Serinin başından ve sonundan 3'er devre için hareketli ortalama hesaplanamaz.

$$\bar{Y}_4 = \frac{\frac{Y_1}{2} + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + \frac{Y_7}{2}}{6}$$

$$= \frac{\frac{22}{2} + 23 + 24 + 29 + 33 + 41 + \frac{32}{2}}{6} = \frac{177}{6} = 29{,}50$$

$$\bar{Y}_5 = \frac{\frac{Y_2}{2} + Y_3 + Y_4 + Y_5 + Y_6 + Y_7 + \frac{Y_8}{2}}{6}$$

$$= \frac{\frac{23}{2} + 24 + 29 + 33 + 41 + 32 + \frac{38}{2}}{6} = \frac{189{,}5}{6} = 32{,}58$$

$$\bar{Y}_6 = \frac{\frac{Y_3}{2} + Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + \frac{Y_9}{2}}{6}$$

$$= \frac{\frac{24}{2} + 29 + 33 + 41 + 32 + 38 + \frac{40}{2}}{6} = \frac{205}{6} = 34{,}16$$

$$\bar{Y}_7 = \frac{\frac{Y_4}{2} + Y_5 + Y_6 + Y_7 + Y_8 + Y_9 + \frac{Y_{10}}{2}}{6}$$

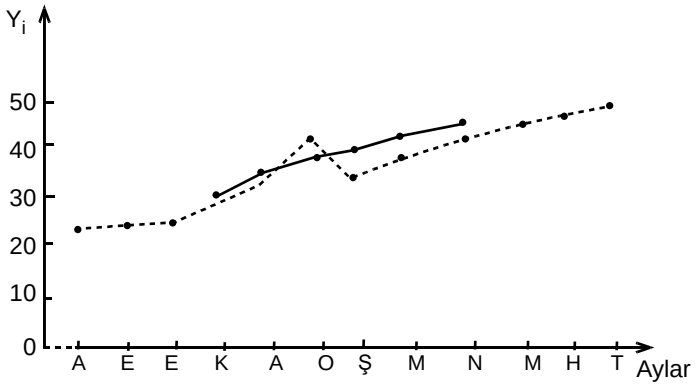
$$= \frac{\frac{29}{2} + 33 + 41 + 32 + 38 + 40 + \frac{43}{2}}{6} = \frac{220}{6} = 36{,}66$$

$$\bar{Y}_8 = \frac{\frac{Y_5}{2} + Y_6 + Y_7 + Y_8 + Y_9 + Y_{10} + \frac{Y_{11}}{2}}{6}$$

$$= \frac{\frac{33}{2} + 41 + 32 + 38 + 40 + 43 + \frac{45}{2}}{6} = \frac{233}{6} = 38{,}83$$

$$\bar{Y}_9 = \frac{\frac{Y_6}{2} + Y_7 + Y_8 + Y_9 + Y_{10} + Y_{11} + \frac{Y_{12}}{2}}{6}$$

$$= \frac{\frac{41}{2} + 32 + 38 + 40 + 43 + 45 + \frac{48}{2}}{6} = \frac{242{,}5}{6} = 40{,}41$$



ÖRNEK 5: Aşağıda bir ülkenin yıllık milli gelirleri verilmiştir.

Yıllar	Milli Gelir
2001	2,3
2002	2,8
2003	2,5
2004	2,9
2005	3,2
2006	3,1
2007	3,6
2008	3,7
2009	3,8

Serinin doğrusal trend modelinin parametrelerini ve 2010 yılı milli gelirini tahmin ediniz.

Çözüm:

Yıllar	Y_i	X_i	X_iY_i	X_i^2
2001	2,3	-4	-9,2	16
2002	2,8	-3	-8,4	9
2003	2,5	-2	-5,0	4
2004	2,9	-1	-2,9	1
2005	3,2	0	0	0
2006	3,1	1	3,1	1
2007	3,6	2	7,2	4
2008	3,7	3	11,1	9
2009	3,8	4	15,2	16
	27,9		11,1	60

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{27,9}{9} = 3,1$$

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{11,1}{60} = 0,185$$

$Y_i = 3,1 + 0,185X_i$

$\hat{Y}_{10} = 3,1 + 0,185(5) = 4,025$ milyon dolar.

ÖRNEK 6: Aşağıda bir bölgenin yıllık nüfusu verilmiştir. (1=100000 kişi)

Yıllar	Nüfus
2000	18
2001	21
2002	19
2003	16
2004	17
2005	15
2006	13
2007	14
2008	13
2009	12

Serinin doğrusal trend modelinin parametrelerini ve 2010 yılı nüfusunu tahmin ediniz.

Yıllar	Y_i	X_i	X_iY_i	X_i^2
1995	18	-9	-162	81
1996	21	-7	-147	49
1997	19	-5	-95	25
1998	16	-3	-48	9
1999	17	-1	-17	1
2000	15	1	15	1
2001	13	3	39	9
2002	14	5	70	25
2003	13	7	91	49
2004	12	9	108	81
	158		-146	330

$$\hat{\alpha} = \frac{\sum Y_i}{n} = \frac{158}{10} = 15,8 \qquad \hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{-146}{330} = -0,44$$

$Y_i = 15,8 - 0,44X_i$

$\hat{Y}_{10} = 15,8 - 0,44(11) = 10,96$

BÖLÜM 16

PARAMETRİK OLMAYAN TESTLER

Önceki bölümlerde anlatılan hipotez testlerinde parametreler test edildiği için sözkonusu testler parametrik testler olarak adlandırılmaktadır. Ancak parametrik testlerin uygulanabilmesi için bazı varsayımların geçerli olması gerekmektedir. Bu varsayımların geçerli olmadığı durumlarda parametrik olmayan testlerden yararlanılabilir. Bu bölümde parametrik olmayan testlerin başlıcaları açıklanacaktır.

16.1. PARAMETRİK OLMAYAN TEST KAVRAMI

Parametrik testlerle ilgili açıklamalarda görüldüğü gibi bu testler, anakütlenin dağılımı ve parametrelerine ilişkin bazı varsayımlara dayanılarak yapılmaktadır; fakat uygulamada bazı durumlarda anakütle ile ilgili varsayımların yerine getirilmesi mümkün olmamaktadır. Çünkü, anakütle ile ilgili varsayımların yerine getirilmesi genellikle araştırıcının elinde değildir.

Araştırmacı anakütle ile ilgili parametrik test varsayımlarını yerine getiremiyorsa, diğer bir deyişle, örneklerin seçildiği anakütlenin dağılımının normal dağılım olduğuna ve dağılımın parametrelerine ilişkin bir varsayımda bulunamıyorsa, parametrik testlerin uygulanması yanlış sonuç verebilir. Böyle durumlarda parametrik olmayan testlerin kullanılması uygun olur. Çeşitli sorunların çözümünde, araştırmacının analizlerde kullanılacak dağılım hakkında bilgi sahibi olmadan ve dağılımın türünü gözönünde bulundurmadan kullanılabileceği testlere gereksinmesi olabilir. Bu testlere parametrik olmayan testler adı verilir.

Diğer bir tanıma göre parametrik olmayan testler; örneğin seçildiği anakütlenin dağılma türünü gözönünde bulundurmadan, diğer bir deyişle, anakütlenin sahip

olduğu dağılımın normal dağılım olduğu varsayımına bağlı olmadan, yalnız örneklerin bağımsız ve tesadüfi olarak seçildiğini varsayan testlerdir.

Parametrik olmayan testlerin uygulanabilmesi için gerekli olan varsayımlar:

-Örneklerin anakütleden tesadüfi olarak seçilmesi ve

-Örneklerin birbirinden bağımsız olmasıdır.

Bunlar örnekle ilgili olan varsayımlardır. Bu nedenle bu varsayımların yerine getirilmesi araştırmacının elindedir. Diğer taraftan bu varsayımlar hem parametrik ve hem de parametrik olmayan testlerde kullanılan ortak varsayımlardır. Yani, araştırmacı ister parametrik, isterse parametrik olmayan testleri kullansın bu varsayımları yerine getirmesi gerekir. Parametrik ve parametrik olmayan testler için gerekli olan varsayımlar Tablo 16.1.'de görülmektedir.

Tablo 16.1

Varsayımlar	Parametrik	Parametrik olmayan
Anakütle ile ilgili	- Normal dağılıma sahip olacak - Dağılımın parametresi $T_1 \leq \theta \leq T_2$ olacak	- Genellikle hiçbir varsayım aranmaz.
Örnekle ilgili	- Örnekler tesadüfi olarak seçilecek - Örnekler birbirinden bağımsız olacak	

Tabloda görüldüğü gibi parametrik olmayan testlerde anakütle ile ilgili varsayım aranmaz; fakat örnekle ilgili varsayımların hem parametrik hem de parametrik olmayan testlerde yerine getirilmesi gerekir.

Parametrik olmayan testler aynı zamanda serbest dağılım metodları olarak da isimlendirilir. Çünkü yukarıdaki tanımlarda da değinildiği gibi bu testler anakütle dağılımına bağlı olmayan testlerdir. Anakütlenin dağılımına bağlı olmayan diğer bir deyişle, genellikle anakütle ile ilgili hiçbir varsayımın aranmadığı bu testler aşağıdaki durumların mevcut olması halinde kullanılabilmektedir.

-İncelenecek problemde sayımla belirtilen karakterlerin bulunması durumunda,

-Parametrik test varsayımlarının yerine getirilmediği durumlarda,

-Testte kullanılacak değerler yerine bu değerlerin sıralarının verildiği durumlarda, diğer bir deyişle, değerler yerine belirli bir düzene göre sıralanmış değerlerin sıra numaralarının verildiği durumlarda,

-Testte kullanılacak örneklerin küçük hacimli olması durumunda.

Yukarıda belirtilen durumların gerçekleşmesi halinde, çeşitli parametrik olmayan testlerden biri seçilerek analiz yapılabilmektedir.

Daha önceki açıklamalarda da değinildiği gibi parametrik testlere karşılık çeşitli parametrik olmayan testler vardır. Bu testlerin parametrik testlere göre birçok olumlu ve olumsuz yönleri olduğu karşılaştırmalar sonucu ortaya konulmuştur. Parametrik olmayan testlerin bu olumlu ve olumsuz yönleri aşağıda görülmektedir.

Parametrik olmayan testlerin olumlu yönleri:

- Testlerin yapılabilmesi için birçok varsayıma gerek yoktur,

- Testlerin anlaşılması ve uygulanması kolaydır,

- Test sonuçlarının elde edilmesi güç işlemleri gerektirmez,

- Testlerin küçük hacimli bir örnek üzerinden yapılması mümkündür.

Parametrik olmayan testlerin olumsuz yönleri:

- Örnek hacminin küçük olması halinde, uygulanması oldukça kolay ve basit olan bu testlerin örnek hacminin büyümesi halinde uygulanması güçleşir.

- Bu tür testlerin uygulanmasıyla elde edilen sonuçlar, parametrik testlerin uygulanmasıyla elde edilen sonuçlardan daha az güvenilirdir.

Parametrik olmayan testler hakkında genel bir açıklama verdikten sonra parametrik olmayan testleri ayrı ayrı inceleyelim.

16.2. İŞARET TESTİ

Parametrik olmayan testlerin en eskisi olan işaret testi normal dağılmayan bir olasılık dağılımının merkezi eğilimi hakkındaki hipotezleri test etmek için kullanılan parametrik olmayan bir testtir. İşaret testi, dağılımın merkezi eğilimi hakkındaki hipotezleri test ederken ortalama yerine medyanı alarak test işlemlerini yürütür. Çünkü, bilindiği gibi merkezi eğilim ölçülerinden biri olan medyan dağılımın eğikliğinden ve uç değerlerden daha az etkilenmektedir. İşaret testinde test işlemleri yapılırken gerçek değerler yerine gözlem değerlerinin (+) veya (-) değerleri alınarak test istatistiği hesaplanır.

İşaret Testinin Aşamaları,

- Hipotezlerin oluşturulması:

Tek Taraflı:

$H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

$H_1 : \mu < \mu_0$

Çift Taraflı :

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

-Güven olasılığı veya hata payının belirlenmesi:

-Test İstatistiğinin hesaplanması :

Test istatistiğinin hesaplanabilmesi için herbir gözlem değerinden kabul edilen medyan değeri çıkarılarak (+) veya (-) işaretli sayılar bulunur. Bulunan bu sayılara göre test istatistiği, hipotezin tek veya çift taraflı olmasına göre iki farklı şekilde hesaplanır. Sol taraflı testte, test istatistiği, pozitif işaretli farkların sayısı, sağ taraflı testte test istatistiği negatif işaretli farkların sayısı iki taraflı teste ise pozitif ve negatif işaretli fark sayılarının küçük olanıdır.

-İşaret testinde karar alma:

Çift taraflı hipotez testinde n birimden oluşan bir örnekte daha az olarak gözlenen fark işaretlerinin olasılığı $\alpha/2$'ye eşit veya daha küçükse H_0 hipotezi α hata payı ile reddedilir.

Sol taraflı testte n birimlik bir tesadüfi örnekte (+) işaretli sayılar daha az ve bunların olasılığı α'ya eşit veya daha küçükse H_0 hipotezi α hata payı ile reddedilir.

Sağ taraflı testte n birimlik bir tesadüfi örnekte (-) işaretli sayılar daha az ve bunların olasılığı α'ya eşit veya daha küçükse H_0 hipotezi α hata payı ile reddedilir.

H_0 hipotezi doğru olduğu zaman belirlenen hipoteze göre fark sayılarının örnekleme dağılımı p=0,50 olan bir Binom dağılımıdır.

Karar aşamasında binom tablosundan bulunan değer ile belirlenen hata payı karşılaştırılarak karar alınır; ancak örnek hacmi n>30 için Binom dağılımı Normal dağılıma yaklaşacağından test istatistiği aşağıda görüldüğü gibi hesaplanır.

$$Z = \frac{(k + 0{,}5) - 0{,}5n}{0{,}5\sqrt{n}} \qquad k < \frac{n}{2}$$

$$Z = \frac{(k - 0{,}5) - 0{,}5n}{0{,}5\sqrt{n}} \qquad k > \frac{n}{2}$$

Hesaplanan bu değer normal dağılım tablosundan hesaplanan değer ile karşılaştırılarak karar verilir.

ÖRNEK: Bir işletmede çalışan işçilerin 1 aylık dönemde gördükleri kurstan sağladıkları yararı ölçmek için 10 kişilik bir gruba bir test yapılmış ve test sonucunda şu puanlar elde edilmiştir.

10, 9, 6, 4, 2, 7, 6, 8, 3, 4

Bu işçilerin aldıkları puanların ortalamasının 5'den büyük olup olmadığını α=0,05 hata payına göre test ediniz.

- H_0 : ortalama = 5
  H_1 : ortalama > 5
-Hata payı α =0,05
-Test istatistiği

10-5, 9-5, 6-5, 4-5, 2-5, 7-5, 6-5, 8-5, 3-5, 4-5
\+ + + - - + + + - -

k= 4 tablo değeri $\rightarrow$ 0,377 tablo değeri > α

0,377>0,05 H_0 kabul edilir.

İşçilerin aldıklarının ortalaması 5'ten büyük bulunmamıştır.

16.3. WILCOXON TESTİ

Bu test, Hipotez testleri bölümünde ortalamalar arasındaki farkların testinde olduğu gibi, bağımsız ve bağımlı örnekler için ayrı ayrı uygulanabilmektedir. Farklı birimlerden oluşmuş iki örnek sözkonusu olduğunda uygulanan teste Wilcoxon Sıra Toplamı Testi, aynı birimlerden elde edilmiş iki farklı ölçüm sözkonusu olduğunda ise uygulanan teste Wilcoxon İşaretli Sıralar Testi adı verilmektedir.

16.3.1. Wilcoxon Sıra Toplamı Testi

İki bağımsız örnekten yararlanarak, iki anakütle ortalamasının karşılaştırılmasının t testi ile yapıldığından daha önce bahsedilmişti. Ancak, anakütlenin dağılımına ilişkin bilgi olmadığı ve örnek birimlerinin değerleri yerine, onların büyüklük sıralarının hesaplanabildiği durumlarda Wilcoxon sıra toplamı testinin kullanılması uygun olmaktadır.

Wilcoxon Sıralar Toplamı Testinde Gerekli Olan Aşamalar,

- Hipotezlerin Belirlenmesi

H_0 : Örneklenen iki anakütle benzer olasılık dağılımlarına sahiptir.

H_1 : Anakütle dağılımı diğerinin sağına (soluna) kaymıştır.

- Testin hata payının belirlenmesi

- Test istatistiğinin hesaplanması

Test istatistiğinin hesaplanabilmesi için, gözlem değerlerinin derlenerek bir tablo halinde düzenlenmesi, düzenlenen bu gözlem değerlerinin çeşitli gruplara ayrılması ve ilgili örneklere dağıtılması, her gruptaki örneklerin gözlem değerlerinin küçükten büyüğe doğru sıralanması, sıralanan bu gözlem değerlerine sıra numaralarının verilmesi gerekir. Bu işlemlerin yapılmasından sonra her tabaka içindeki küçük olan sıra numaralarının toplamı alınarak test istatistiği değeri,

$$n_1 < n_2 \quad ise \quad T = T_1 = \sum_{i=1}^{n} t_{1.i}$$

veya

$$n_1 > n_2 \quad ise \quad T = T_2 = \sum_{i=1}^{n} t_{2.i}$$

olarak hesaplanır.

Yukarıdaki ifadede görülen $t_{1.i}$ ve $t_{2.i}$ notasyonu 1 ve 2 no'lu örneklerdeki birimlere verilen sıra numaralarını göstermektedir.

Bu testle her iki örneğin birim sayılarının birbirine eşit olması durumunda toplamlardan birisi test istatistiği olarak kullanılır. Ancak incelenen örneklerde aynı gözlem değerlerini alan birimlerin bulunması durumunda bu birimler sıra numaralarının ortalama değerini alırlar. Her iki örneğin sıra numaraları toplamı her zaman

$$T_1 + T_2 = \frac{n(n+2)}{2}.n = n_1 + n_2$$

olacaktır.

Diğer taraftan her iki örneğin birim sayıları $n_1 \geq 30$ ve $n_2 \geq 30$ ise bu durumda T_2'nin örnekleme dağılımı ortalaması ve varyansı,

$$E(T_2) = \frac{n_1(n_1 + n_2 + 1)}{2} \qquad ve \qquad \sigma^2_{T_2} = \frac{n(n+1)(2n+1)}{24}$$

olan normal dağılıma yaklaşacağından test istatistiği,

$$Z = \frac{T_2 - E(T_2)}{\sigma_{T_2}}$$

şeklinde hesaplanır.

$$\left.\begin{array}{l} |Z| < Z_{\alpha} \\ |Z| < Z_{\alpha/2} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ kabul}$$

$$\left.\begin{array}{l} |Z| > Z_{\alpha} \\ |Z| > Z_{\alpha/2} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ reddedilir.}$$

ÖRNEK: Bir bankanın iki ayrı şubesinin 2009 yılı teknik destek masrafları aylar itibariyle hesaplanmıştır. 0,01 anlam düzeyinde şubelerin teknik destek masrafları arasında anlamlı bir fark olup olmadığını test ediniz. (1 = 1 TL.)

A Şubesi	B Şubesi
10	18
15	16
22	12
18	24
30	15
17	20
20	12
25	28
45	38
12	18
34	36
26	20

A Şubesi	B Şubesi	A Şubesi Sıra No	B Şubesi Sıra No
10	18	1	10
15	16	5,5	7
22	14	15	4
18	24	10	16
30	15	20	5,5
17	20	8	13
20	12	13	2
25	28	17	19
45	38	24	23
13	18	3	10
34	36	21	22
26	20	18	13
		155,5	

H_0 : İki şubenin masrafları arasında fark yoktur.

H_1 : İki şubenin masrafları arasında fark vardır.

w : 155,5

$$\overline{X}_w = \frac{n_A(n_A + n_B + 1)}{2} = \frac{12(12 + 12 + 1)}{12} = 150$$

$$\sigma_w = \sqrt{\frac{n_A(n_B)(n_A + n_B + 1)}{12}} = 17{,}321$$

$$Z = \frac{w - \overline{X}_w}{\sigma_w} = \frac{155{,}5 - 150}{17{,}321} = 0{,}3175$$

$Z_{0,05} = 2{,}58$

0,3175 < 2,58 H_0 kabul, anlamlı bir fark yoktur.

16.3.2. Wilcoxon İşaretli Sıralar Testi

Wilcoxon işaretlenmiş sıralar toplamı testi parametrik olmayan testler içinde sık kullanılan bir test türüdür. Bu test, birbirleri ile eşleştirilmiş gözlem değerleri arasındaki farkın anlamlı olup olmadığını araştıran bir testtir.

Wilcoxon işaretlenmiş sıralar toplamı testi birbirleri ile eşleştirilmiş gözlem değerleri arasındaki farklılığın anlamlılığını ölçmekle kalmayıp aynı zamanda değişmelerin yönünü ve büyüklüğünü de dikkate alan bir testtir.

Yukarıda açıklandığı üzere gözlem değerleri arasındaki farklılığın hem büyüklüğünü ve hem de yönünü gözönüne alan bu testin uygulanabilmesi için her bir eşlendirilen değer arasındaki farkların hesaplanması gereklidir. Hesaplanan bu farkların değerlerinin bazıları pozitif, bazıları negatif olacaktır. Fakat analizde, önce bu işaretler dikkate alınmadan, herbir farklılığın mutlak değerlerine birer sıra numaraları verilir. Bu numaralama işleminden sonra farkların işaretleri sıra numaralarının önlerine yazılarak işaretli sıra numaraları elde edilir.

Değerler arasında bir farkın olmaması halinde, pozitif işaretli sıra numaralarının toplamının, yaklaşık olarak negatif işaretli sıra numaralarının toplamına eşit olması gerekir. Bu düşünceden hareketle farklılığın olması halinde sıfır hipotezinin reddedilmesi gereklidir. Bu testte negatif farkların toplamı T_- ve pozitif farkların toplamı T_+ olarak belirlenir.

Yukarıda genel olarak açıklanan Wilcoxon işaretlenmiş sıralar testi aşağıdaki aşamalardan oluşur.

- Hipotezlerin Oluşturulması

Tek Taraflı Testte :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$
$$H_1 : \theta < \theta_0$$

Çift Taraflı Testte:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

H_0 : İki anakütle benzer olasılık dağılımlarına sahiptir.

H_1 : Anakütle dağılımı diğerinin soluna kaymıştır.

-Testin hata payının belirlenmesi

-Test istatistiğinin hesaplanması

Test istatistiğinin hesaplanabilmesi için, önce eşlendirilmiş gözlem değerleri arasındaki farklara (+) ve (-) işaretleri, daha sonra bu farkların işaretlerine bakılmadan yani mutlak değerlerine göre uygun olan sıra numaraları verilir. Sıra numaraları verilirken aynı değeri alan birden çok fark varsa bu farklara karşılık gelen sıra numaralarının ortalaması bu değerlerin sıra numarası olur. Sıra numaraları verilen bu farklara (+) veya (-) işaretler verilir ve her örnekteki (+) veya (-) işaretli sıra numaraları toplanarak, küçük olan işaretlenmiş sıra numaraları toplamını gösteren test istatistiği değeri (T′) hesaplanmış olur.

-Karar alma

Karar alma işlemi için hesaplanan test istatistiği değeri α ve n değerlerine göre Wilcoxon işaretlenmiş sıralar tablosundan hesaplanan, T_α değeri ile karşılaştırılır ve karar aşağıda görüldüğü gibi alınır.

Tek taraflı testin uygulanması durumunda karar işlemi,

$$\left.\begin{array}{l} T' < -T_\alpha \\ T' > T_\alpha \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} T' > -T_\alpha \\ T' < T_\alpha \end{array}\right\} \quad H_0 \text{ kabul}$$

İki taraflı testin uygulanması durumunda karar işlemi,

$$\left.\begin{array}{l} T' < -T_{\alpha/2} \\ T' > T_{\alpha/2} \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} T' > -T_{\alpha/2} \\ T' < -T_{\alpha/2} \end{array}\right\} \quad H_0 \text{ kabul}$$

Açıklanan bu test işlemleri $n_1+n_2<20$ olması hali için geçerli olmaktadır. Diğer taraftan $n_1+n_2>20$ olması halinde T′nin örnekleme dağılımı ortalaması,

$$m = \frac{n(n+1)}{4}$$

ve standart sapması

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

olan bir normal dağılıma yaklaşır. Bu durumda test istatistiği,

$$Z' = \frac{T' - \{n(n+1)/4\}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

ifadesine göre hesaplanır.

Burada,

T′ = küçük olan işaretlenmiş sıra numaraları toplamını

n = örnekteki birimlerin sayısını

gösterir. Bu durumda karar işlemi aşağıdaki gibi olacaktır.

İki taraflı testin uygulanması halinde karar işlemi,

$$\left.\begin{array}{l} Z' < -Z_{\alpha/2} \\ Z' > Z_{\alpha/2} \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} Z' > -Z_{\alpha/2} \\ Z' < Z_{\alpha/2} \end{array}\right\} \quad H_0 \text{ kabul}$$

Tek taraflı testin uygulanması halinde karar işlemi,

$$\left.\begin{array}{l} Z' < -Z_{\alpha} \\ Z' > Z_{\alpha} \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} Z' > -Z_{\alpha} \\ Z' < -Z_{\alpha} \end{array}\right\} \quad H_0 \text{ kabul}$$

olacaktır.

ÖRNEK: Bir fabrikada iş kazalarının önlenmesi konusunda bir iş güvenliği semineri düzenlenmiştir. Bu seminer 10 grup üzerinden yürütülmüştür. Aşağıdaki tabloda seminerden sonra meydana gelen iş kazaları miktarı verilmiştir.

Gruplar	**Seminerden Önce(Y_1)**	**Seminerden Sonra(Y_2)**
1	61	82
2	42	74
3	74	89
4	37	35
5	49	64
6	41	35
7	35	45
8	64	71
9	78	74
10	34	46

α=0,05 hata payına göre kaza miktarlarındaki azalmanın seminerden dolayı meydana geldiği hipotezini Wilcoxon işaretlenmiş sıralar toplamı testine göre inceleyiniz.

$H_0 : \theta_1 = \theta_2$ Kaza miktarındaki azalma seminerden dolayıdır.
$H_1 : \theta_1 \neq \theta_2$ Kaza miktarındaki azalma seminerden dolayı değildir.
$n = 10$ $\alpha = 0,05$

Farklar d=Y1-Y2	**Farkların Sıralandırılması**	**Farkların İşaretlendirilmesi**
-21	9	-9
-32	10	-10
-15	7,5	-7,5
2	1	1
-15	7,5	-7,5
6	3	3
-10	5	-5
-7	4	-4
4	2	2
-12	6	-6

$T' = 1+2+3 = 6$
$T_\alpha = 8$
$n = 10$
$\alpha = 0,05$

$T' < T_\alpha$ H_0 kabul

H_0 kabul edildiğinden kaza miktarındaki azalmanın seminerden olduğu sonucuna varılır.

16.4. MANN-WHITNEY U TESTİ

Parametrik olan hipotez testlerinde iki bağımsız örneğin aynı anakütleden gelip gelmediğini test etmek için t testi kullanılır. Parametrik hipotez testlerinde gerekli olan varsayımların yerine getirilemediği durumlarda t testi yerine Mann-Whitney U Testi kullanılır. Diğer taraftan değişken değerleri sayısal ve kesikli ise örnekteki birim sayısı, 30'dan az ise bu gibi durumlarda da Mann-Whitney U testi kullanılır.

Mann-Whitney U Testiyle t testinde olduğu gibi iki bağımsız örneğin aynı anakütlelerden alınıp alınmadığı test edilmektedir. Ancak t testinde anakütle ile ilgili varsayım önemli olduğundan bu kolaylık nedeniyle Mann-Whitney U Testi t testine göre daha fazla tercih edilen bir test olmaktadır.

Mann-Whitney U Test'inde gerekli olan safhalar,
-Hipotezlerin oluşturulması
Tek Taraflı Testte :

$H_0 : \theta = \theta_0$
$H_1 : \theta > \theta_0$
$H_1 : \theta < \theta_0$

Çift Taraflı Testte:

$H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$

-Hata payının belirlenmesi
-Test istatistiğinin hesaplanması

Test istatistiğinin hesaplanabilmesi için önce her iki örnekteki gözlem değerleri küçükten büyüğe doğru sıralanır ve sıralanmış bu değerlere birer sıra numarası verilir. Sıra numaralarının verilmesinde dikkat edilecek nokta iki veya daha fazla gözlem değerinin aynı olması halinde, numaralama işleminin farklı bir şekilde yapılacağıdır. Yani, iki veya daha fazla gözlem değeri aynı ise herbir değere ortalama sıra değeri verilerek sıralama işlemine devam edilir. Sıralama işleminden sonra her örnekteki sıra numaralarının toplamı alınarak, küçük numaralar toplamını gösteren t test istatistiği değeri olan T′ değeri hesaplanır.

-Karar alma

T′ değerinin hesaplanmasından sonra α anlamlılık düzeyine göre Mann-Whitney U tablosundan T_α değeri bulunarak aşağıda görüldüğü şekilde karar alınır. Karar işlemi,

Tek taraflı testin uygulanması durumunda,

$$\left.\begin{array}{l} T' < T_\alpha \\ T' > -T_\alpha \end{array}\right\} \quad H_0 \text{ kabul}$$

$$\left.\begin{array}{l} T' > -T_\alpha \\ T' < -T_\alpha \end{array}\right\} \quad H_0 \text{ red}$$

İki taraflı testin uygulanması durumunda karar işlemi,

$$\left.\begin{array}{l} T' < T_{\alpha/2} \\ T' > -T_{\alpha/2} \end{array}\right\} \quad H_0 \text{ kabul}$$

$$\left.\begin{array}{l} T' > T_{\alpha/2} \\ T' < -T_{\alpha/2} \end{array}\right\} \quad H_0 \text{ red}$$

olacaktır. Karar işleminden sonra alınan karara göre yorum yapılır.

Açıklanan bu test işlemleri $n_1+n_2<20$ olması hali için geçerli olmaktadır. Diğer taraftan $n_1+n_2>20$ olması halinde T′nin örnekleme dağılımı ortalaması,

$$m = \frac{n_1 n_2}{2}$$

ve standart sapması

$$\sigma = \sqrt{\frac{n_1 n_2 (n+1)}{12}}$$

olan bir normal dağılıma yaklaşır. Bu durumda test istatistiği,

$$Z' = \frac{T' - \{n_1 n_2 / 2\}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}}$$

ifadesine göre hesaplanır.

Burada,

T′ = örnek hacmi küçük olan örneğin sıra numaraları toplamını

n_1 = birinci örnekteki birim sayısını

n_2 = ikinci örnekteki birim sayısını

gösterir.

Bu durumda karar işlemi aşağıdaki gibi olacaktır.

İki taraflı testin uygulanması halinde karar işlemi,

$$\left.\begin{array}{l} Z' < -Z_{\alpha/2} \\ Z' > Z_{\alpha/2} \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} Z' > -Z_{\alpha/2} \\ Z' < Z_{\alpha/2} \end{array}\right\} \quad H_0 \text{ kabul}$$

Tek taraflı testin uygulanması halinde karar işlemi,

$$\left.\begin{array}{l} Z' < -Z_{\alpha} \\ Z' > Z_{\alpha} \end{array}\right\} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} Z' > -Z_{\alpha} \\ Z' < Z_{\alpha} \end{array}\right\} \quad H_0 \text{ kabul}$$

olacaktır.

ÖRNEK: Bir fabrikada çalışan teknik personele yeni gelişmeleri öğretmek amacıyla bir kurs düzenlenmiştir. Düzenlenen kursa katılan personel iki ayrı gruba ayrılmıştır. Birinci gruba ayrılan sekiz personel 1.kursa, ikinci gruba ayrılan dokuz personel 2.kursa katılmıştır. Kurs sonunda yapılan sınavda alınan notlar aşağıda verilmiştir.

Sınav Sonuçları

Kurs 1	81	83	67	65	75	87	61	80	
Kurs 2	79	75	91	88	67	93	85	95	77

Sıralanmış Gözlem Değerleri	**Ait Olduğu Örnekler**	**Gözlem Değerlerine Sıra Numaralarının Verilmesi**	**I.Örneğin Sıra Numaraları**	**II.Örneğin Sıra Numaraları**
61	1	1	1	-
65	1	2	2	-
67	2	3,5	-	3,5
67	1	3,5	3,5	-
75	1	5,5	5,5	-
75	2	5,5	-	5,5
77	2	7	-	-
79	2	8	-	8
80	1	9	9	-
81	1	10	10	-
83	1	11	11	-
85	2	12	-	12
87	1	13	13	14
88	2	14	-	14
91	2	15	-	15
93	2	16	-	16
95	2	17	-	17

α=0,05 anlamlılık seviyesine göre iki ayrı gruba verilen kurslar arasında bir farkın olup olmadığını Mann Whitney U testine göre inceleyiniz.

$H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$

n_1 =8 n_2=9 α=0,05

Mann-Whitney U testi tablosundan görüleceği gibi,

n_1 =8 n_2=9
α=0,05'a göre T_α=51

$T'=55$ $T_\alpha=51$
$T'>T_\alpha$ H_0 red.

1. ve 2. kurs arasında bir fark vardır.

16.5. KRUSKAL-WALLIS H TESTİ

Kruskal-Wallis H Testi, iki veya daha çok örneğin seçildikleri anakütlelerin olasılık dağılımlarının benzerliğini test eden parametrik olmayan bir testtir. F testinin kullanıldığı varyans analizinde gerekli olan varsayımlar yerine getirilemediği durumlarda Kruskal-Wallis H Testi uygulanabilir. Kruskal-Wallis H Testi için gerekli olan aşamalar aşağıda görüldüğü gibidir.

-Hipotezlerin Oluşturulması

H_0 : k adet olasılık dağılımına eşittir.
H_1 : k adet olasılık dağılımından en az iki tanesi birbirlerinden farklı yerleşmiştir.

-Hata payının belirlenmesi

-Test istatistiğinin hesaplanması

Test istatistiği H harfi ile ifade olunur ve aşağıda görülen formül ile hesaplanır.

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

veya

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} n_j \overline{R}_j - \overline{R}^2$$

Bu formüllerde;

n_j : j. Örneğin hacmini
R_j : j. Örneğin sıra numarasını
k : olasılık dağılımları karşılaştırılan örnek sayısı
n : örnek hacmini
gösterir.

-Karar alma

Yukarıda görülen formülle hesaplanan test istatistiği hata payı ve serbestlik derecesine göre ki-kare tablosundan elde edilen değer ile karşılaştırılarak karar alınır.

Çünkü sıfır hipotezi doğru iken test istatistiği dağılımının tekrarlanan örneklemede ki-kare dağılımına yaklaşmasıdır.

Böylece,

$H < \chi^2_{\alpha,k-1}$ ise H_0 kabul

$H > \chi^2_{\alpha,k-1}$ ise H_0 reddedilir

olarak karar alınır. Kararın alınmasından sonrada elde edilen sonuç yorumlanır.

ÖRNEK: Bir kargo şirketinin dört ayrı bölge müdürlüğüne bağlı şubelerde, yılın ilk altı ayındaki hasarlı karga maliyetleri karşılaştırılmak isteniyor. Toplam hasar maliyetleri aşağıdaki gibidir. 0,05 hata payı ile farklılık olup olmadığını test ediniz.

	1.	**2.**	**3.**	**4.**
Ocak	60	22	53	30
Şubat	44	24	38	19
Mart	32	40	29	20
Nisan	55	52	45	27
Mayıs	21	46	33	18
Haziran	48	25	36	39

	1.	**2.**	**3.**	**4.**
Ocak	60**(24)**	22**(5)**	53**(22)**	30**(10)**
Şubat	44**(17)**	24**(6)**	38**(14)**	19**(2)**
Mart	32**(11)**	40**(16)**	29**(9)**	20**(3)**
Nisan	55**(23)**	52**(21)**	45**(18)**	27**(8)**
Mayıs	21**(4)**	46**(19)**	33**(12)**	18**(1)**
Haziran	48**(20)**	25**(7)**	36**(13)**	39**(15)**
w_i	99	74	88	39
w_i^2 / n_i	1633,5	912,67	1290,67	253,5

H_0 : Şubelerin hasarlı kargo maliyetleri arasında fark yoktur.
H_1 : Şubelerin hasarlı kargo maliyetleri arasında fark vardır.

$$H = \left[\frac{12}{n(n+1)}\sum\frac{w_i^2}{n_i}\right] - 3(n+1) = \left[\frac{12}{24(25)}(4090,34)\right] - 3(25) = 6,8068$$

$$\chi^2_{0,05,3} = 7,815$$

6,8068 < 7,815 olduğundan H_0 kabul. Şubelerin hasarlı kargo maliyetleri arasında anlamlı bir fark yoktur.

16.6. SPEARMAN SIRA KORELASYON KATSAYISI VE TESTİ

Korelasyon konusunda değinildiği gibi istatistiğin önemli amaçlarından birisi değişkenler veya olaylar arasındaki ilişkilerin incelenmesidir. Değişkenler arasındaki ilişki korelasyon katsayısı ile ölçülmektedir. İlişkinin ölçülmesinde kullanılan korelasyon katsayısının anlamlılığı test edilirken X ve Y değerlerinin iki değişkenli normal dağılıma sahip bir anakütleden seçilmiş bir örnek olduğu varsayılır. Ancak gerçek hayatta ve özellikle ekonomik olaylarda bu varsayımın sağlanması zordur. Bu nedenle korelasyona, Spearman sıra korelasyonu katsayısı adı verilen bir yaklaşım getirilmiştir.

Spearman sıra korelasyon katsayısı ilişkilerin derecesi ve yönünün ölçülmesinde kullanılan parametrik olmayan bir yöntemdir.

Spearman sıra korelasyon katsayısı iki farklı şekilde hesaplanabilir.

Birinci yöntem,

$$r_s = \frac{SS_{UV}}{\sqrt{SS_{UU}SS_{VV}}}$$

$$SS_{UV} = \Sigma U_i - \overline{U}V_i - \overline{V} = \Sigma U_i V_i - \frac{\Sigma U_i \Sigma V_i}{n}$$

$$SS_{UU} = \Sigma U_i - \overline{U}U_i - \overline{U} = \Sigma U_i^2 - \frac{\Sigma U_i^2}{n}$$

$$SS_{VV} = \Sigma V_i - \overline{V}V_i = \Sigma V_i^2 - \frac{\Sigma V_i^2}{n}$$

Bu ifadelerde,

U_i : Birinci örneğin sıra numaraları
V_i : İkinci örneğin sıra numaraları
n : Her örneğin gözlem sayısını gösterir.

İkinci yöntem,

$$r_S = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$$

$d_i = u_i - v_i$

Bu ifadede

d : sıralar arasındaki farkları
n : terim sayısını

gösterir.

Her iki yöntem ile yapılan hesaplamada verilerin değerleri yerine sıraları esas alınmaktadır. Eldeki veriler sıraları gösterdiğinden hesaplama doğrudan doğruya yapılır. Ancak iki değişkenin gerçek değerleri biliniyorsa, önce terimlere küçükten büyüğe veya büyükten küçüğe sıra numaraları verilir. Ancak eğer iki terimin değerleri birbirine eşitse bu durumda eşit terimlere bulundukları sıraların ortalaması verilir ve daha sonra hesaplama yapılır. Hesaplama sonucunda bulunan sıra korelasyonu katsayısı şöyle yorumlanır.

$r_s = 1$ tam pozitif ilişki

$0 < r_s < 1$ pozitif ilişki

$r_s = 0$ ilişki yok

$-1 < r_s < 0$ negatif ilişki

$r_s = -1$ tam negatif ilişki

Örnekten hesaplanan Spearman sıra korelasyon katsayısının istatistiksel olarak sıfırdan anlamlı bir farkı olup olmadığı test edilebilir. Bu teste Spearman Sıra Korelasyon Katsayısı Testi adı verilir. Bu testin aşamaları şöyledir.

- Hipotezlerin Oluşturulması

Tek taraflı

$$H_0 : \rho_S = 0$$
$$H_1 : \rho_S > 0$$

veya

$$H_1 : \rho_S < 0$$

Çift taraflı

$$H_0 : \rho_S = 0$$
$$H_1 : \rho_S \neq 0$$

- Hata payının belirlenmesi
- r tablosundan kritik değerlerin bulunması
- Karar işlemi

Yapılan hesaplamalar sonucunda elde edilen değer ile Spearman sıra korelasyon katsayısı kritik değerleri karşılaştırılarak karar alınır.

-Karar işlemleri şöyle yapılır.

$$\left.\begin{array}{l} r_S < r_{S\alpha} \\ r_S > -r_{S\alpha} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ kabul}$$

$$\left.\begin{array}{l} r_S > r_{S\alpha} \\ r_S < -r_{S\alpha} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ red}$$

$$\left.\begin{array}{l} r_S < r_{S\alpha/2} \\ r_S > -r_{S\alpha/2} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ kabul}$$

$$\left.\begin{array}{l} r_S > r_{S\alpha/2} \\ r_S < -r_{S\alpha/2} \end{array}\right\} \quad \text{ise} \quad H_0 \text{ red}$$

Karar işleminden elde edilen sonuca göre yorum yapılır.

ÖRNEK: İnşaat sektöründe yapılan bir incelemede, çimento fiyatları ne demir fiyatları arasında bir ilişki olup olmadığı araştırılmak isteniyor. Bu amaçla, 12 ay itibariyle saptanan çimento ve demir fiyatları aşağıda verilmiştir. 0,95 güvenle test ediniz.

Çimento Fiyatları	Demir Fiyatları
100	101
110	107
109	110
120	117
121	118
121	118
125	143
165	168
183	172
184	172
262	292
279	295

Çimento Fiyatları	Demir Fiyatları	Sıra No Çimento	Sıra No Demir	D^2
100	101	1	1	0
110	107	3	2	1
109	110	2	3	1
120	117	4	4	0
121	118	5,5	5,5	0
121	118	5,5	5,5	0
125	143	7	7	0
165	168	8	8	0
183	172	9	9,5	0,25
184	172	10	9,5	0,25
262	292	11	11	0
279	295	12	12	0
				2,5

$H_0 : \rho_s = 0$

$H_1 : \rho_s \neq 0$

$r_S = 1 - \frac{6\Sigma D^2}{n(n^2-1)} = 1 - \frac{6(2,5)}{12(12^2-1)} = 0,99 \rightarrow$ çok kuvvetli aynı yönlü ilişki

$\alpha = 0,05$ ve n =12 için r tablo değeri 0,497'dir.

0,99 > 0,497 olduğundan H_0 red edilir, ilişki vardır.

16.7. Kİ-KARE TESTİ

Önceki bölümlerde olasılık dağılımları incelenirken ki-kare dağılımına değinilmişti. Bu bölümde de parametrik olmayan testlerden biri olan ki-kare testi incelenecektir.

Aralarında ilişki olup olmadığı araştırılan değişkenlerin değişimleri rakamlarla belirlenebildiği gibi bazı durumlarda ise bu değişimler rakamlarla belirlenemez. Bu gibi durumlarda değişkenler nitelik sınıflarına ayrılarak ve her sınıfa düşen birim sayısı hesaplanarak kontenjans tabloları ve bu tablolardan yararlanarak da ki-kare testi yapılır.

Ki-kare testi parametrik olmayan testler arasında yer almakta ve ki-kare bağımsızlık testi, ki-kare homojenlik testi ve ki-kare uygunluk testi olmak üzere üç ayrı şekilde uygulanmaktadır.

16.7.1. Ki-Kare Bağımsızlık Testi

Ki-kare bağımsızlık testi ile herhangi iki olay arasında bağımsızlığın olup olmadığı incelenir. Parametrik olmayan testlerden biri olan ki-kare testi ile, iki olay arasında bağımsızlığın olup olmadığı, teorik frekanslar ile gerçek frekanslar yani, gözlemlenen frekanslar arasındaki farklara göre incelenerek belirlenir. Frekans değerlerinin birbirine yakın olması halinde bağımsızlık olduğu, değerlerin farklı olması halinde ise bağımsızlık olmadığı sonucuna varılır.

Ki-kare Bağımsızlık Testi'nin aşamaları,

-Hipotezin düzenlenmesi

H_0 : Bağımsızlık var.

H_1 : Bağımsızlık yok.

-Hata payının belirlenmesi

Bu testte hata payı 0,05 veya 0,01 olabilir. Bu hata payına ve $\nu=n-1$ serbestlik derecesine göre χ^2 tablosundan tablo değeri bulunur. Ancak χ^2 tablosu serbestlik derecesi yani $\nu \leq 30$ değerleri için kullanılır. Serbestlik derecesinin değerleri otuzdan büyük ise yani $\nu > 30$ ise tablo değeri normal dağılım tablosundan bulunur.

-Test istatistiğinin hesaplanması

Test istatistiğinin hesaplanmasında gerçek frekanslar ile teorik frekanslar ele alınır. Bu frekanslar arasındaki farklılık az ise incelenen iki olay arasında bağımsızlığın olduğu sonucuna varılır. Aksi halde ise olaylar arasında bağımsızlık yoktur. Test istatistiği,

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - Np_i)^2}{Np_i} = \sum \frac{(f_i - f_i')^2}{f_i'}$$

ifadesine göre hesaplanır. Bu ifadede,

f_i : gerçek frekanslar

f_i' : teorik frekansları

göstermektedir.

İfadede önemli olan nokta Np_i değerinin 5'den büyük olmasıdır. Yani χ^2 değerinin hesaplanabilmesi için,

$Np_i \geq 5$

şartının gerçekleşmesi gerekir. Eğer bu şart gerçekleşmiyorsa, bazı sınıflar birleştirilerek şartın gerçekleşmesi sağlanır. Fakat uygulamalarda sınıf birleştirilmesinin uygun olmadığı durumlar ile karşılaşılır. Bu gibi durumlarda χ_c^2 değeri Yates düzeltmesi adı verilen, yani

$$\chi^2 = \sum \frac{(f_i - f_i')^2 - 0{,}5}{f_i'}$$

ifadesi ile hesaplanır.

-Karar işlemi

Karar alınırken hesaplanan test istatistiği değeri ile serbestlik derecesi ve anlamlılık düzeyine göre hesaplanan tablo değeri karşılaştırılarak sonuca varılır. Önceki açıklamalarda da değinildiği gibi $\nu \leq 30$ ise tablo değeri χ^2 tablosundan, $\nu > 30$ ise tablo değeri normal dağılım tablosundan yani Z tablosundan elde edilir. Özetle,

$\nu \leq 30$ ise

$\chi^2 < \chi^2_\alpha$ ise H_0 kabul edilir.

$\chi^2 > \chi^2_\alpha$ ise H_0 reddedilir.

$\nu > 30$ ise

$\chi^2 < Z_\alpha$ ise H_0 kabul edilir.

$\chi^2 > Z_\alpha$ ise H_0 reddedilir.

ÖRNEK: Yapılan araştırmada çeşitli dinler ile coğrafi bölgeler arasındaki ilişki incelenmek istenmektedir. Bu araştırma için doğu ve batı olmak üzere iki bölgeden iki ayrı grup tesadüfi olarak seçilmiş ve gruptaki herbir kişi protestan, katolik ve yahudi olarak sınıflandırılmıştır. Gerçek frekanslar aşağıdaki tabloda görüldüğü gibidir.

Bölgeler	Protestan	Katolik	Yahudi	Toplam
Doğu	182	215	203	600
Batı	154	136	110	400
Toplam	336	351	313	1000

Bu verilere göre bölgeler ile seçilen dinler arasında α=0,05 hata payına göre bir ilişki var mıdır, inceleyiniz.

H_0 : Bağımsızlık var.
H_1 : Bağımsızlık yok.

$\nu = (r-1)(k-1)$ r: sıra k: kolon
$= (2-1)(3-1)$
$= 2$

$\chi^2_{0,05} = 5{,}991$ ν=2 için

dir.

$$\chi^2 = \sum \frac{(f_i - f_i')^2}{f_i'}$$

İfadede görülen teorik frekansların (f_i') hesabı şöyle yapılır.

K : seçilen kişi protestandır.
C : seçilen kişi katoliktir.
J : seçilen kişi yahudidir.
E : seçilen kişi doğu bölgesindendir.
W: seçilen kişi batı bölgesindendir.

Bu olayların meydana gelme olasılıkları şöyledir:

$$P(K) = \frac{336}{1000} \qquad P(C) = \frac{351}{1000} \qquad P(J) = \frac{313}{1000}$$

$$P(E) = \frac{600}{1000} \qquad P(W) = \frac{400}{1000}$$

H_0 hipotezinin doğru ve iki değişkenin birbirinden bağımsız olduğu varsayımına göre aşağıdaki olasılık değerleri hesaplanabilir.

$$P(K \cap E) = P(K).P(E) = \left(\frac{336}{1000}\right).\left(\frac{600}{1000}\right)$$

$$P(K \cap W) = P(K).P(W) = \left(\frac{336}{1000}\right).\left(\frac{400}{1000}\right)$$

$$P(C \cap E) = P(C).P(E) = \left(\frac{351}{1000}\right).\left(\frac{600}{1000}\right)$$

$$P(C \cap W) = P(C).P(W) = \left(\frac{351}{1000}\right).\left(\frac{400}{1000}\right)$$

$$P(J \cap E) = P(J).P(E) = \left(\frac{313}{1000}\right).\left(\frac{600}{1000}\right)$$

$$P(J \cap W) = P(J).P(W) = \left(\frac{313}{1000}\right).\left(\frac{400}{1000}\right)$$

değerinin hesabı için gerekli olan teorik frekanslar yukarıda hesaplanan olasılık değerleri ile gözlem sayısının çarpımından bulunur. Örneğin, doğu bölgesinde yaşayan Protestanların teorik frekansları

$$\left(\frac{336}{1000}\right).\left(\frac{600}{1000}\right).(1000) = \frac{(336)(600)}{1000} = 202$$

olacaktır.

Diğer frekansları aynı şekilde hesaplayarak aşağıdaki tabloyu düzenleyebiliriz.

Bölgeler	Protestan	Katolik	Yahudi	Toplam
Doğu	182(202)	215(211)	203(187)	600
Batı	154(134)	136(140)	110(126)	400
Toplam	336	351	313	1000

Bu tablo değerinden yararlanarak χ^2 değeri,

$$= \frac{(182-202)^2}{202} + \frac{(215-211)^2}{211} + \frac{(203-187)^2}{187}$$

$$+ \frac{(154-134)^2}{134} + \frac{(136-140)^2}{140} + \frac{(110-126)^2}{126} = 8{,}556$$

olarak hesaplanır.

$$\chi^2 > \chi^2_{0,05}$$

olduğundan H_0 reddedilir. Yani seçilen dinler ile bölgeler arasında bir ilişki olmadığı diğer bir ifadeyle bu iki olayın birbirinden bağımsız olduğu sonucuna varılır.

16.7.2. Ki-Kare Homojenlik Testi

Araştırma sonucunda elde edilen gözlem değerlerinde tespit edilen farkların ne dereceye kadar tesadüfi olarak meydana gelebilecekleri incelenmek istendiğinde ki-kare homojenlik testinden yararlanılır. Bu test sonucunda homojenliğin olup olmadığı sonucuna varılır.

Ki-Kare Homojenlik Testindeki Safhalar,

-Hipotezin Düzenlenmesi

H_0: homojenlik var
H_1:homojenlik yok.

-Hata payının belirlenmesi

Bu testte anlamlılık düzeyi 0,05 veya 0,01 olabilir.

-Test istatistiğinin hesaplanması

Test istatistiğinin hesaplanmasında gerçek frekanslar ile teorik frekanslar hesaplanır. Test istatistiği,

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - Np_i)^2}{Np_i} = \sum \frac{(f_i - f_i')^2}{f_i'}$$

ifadesine göre hesaplanır. Burada önemli olan noktalardan biri test işlemlerinin yapılabilmesi için teorik frekansların her birinin değerinin en az 5 olmasıdır. Teorik değerler arasında 5'ten küçük değerlerin olması durumunda birbirine yakın olan sınıflar birleştirilir veya Yates düzeltmesine başvurulur.

-Karar işlemi

Karar alınırken test istatistiği değeri ile serbestlik derecesi ve anlamlılık düzeyine göre hesaplanan tablo değeri karşılaştırılarak sonuca varılır. Ancak $\nu \leq 30$ ise tablo değeri ki-kare tablosundan, $\nu > 30$ ise tablo değeri normal dağılım tablosundan bulunur. Böylece karar işlemi,

$\nu \leq 30$ ise

$\chi^2 < \chi^2_{\alpha,\nu}$ ise H_0 kabul

$\chi^2 > \chi^2_{\alpha,\nu}$ ise H_0 reddedilir.

$\nu > 30$ ise

$\chi^2 < Z_\alpha$ ise H_0 kabul

$\chi^2 > Z_\alpha$ ise H_0 reddedilir.

ÖRNEK: Elektronik aletler imal edilen bir fabrikada, müzikle üretimin işçilerin verimliliğini arttırıp arttırmayacağı incelenecektir. Bu amaçla bir deneme süresi sonunda tespit edilen kusurlu ve kusursuz mamul sayılan aşağıdaki kontenjans tablosunda verilmiştir. 0,05 hata payı ile testi yapınız.

	Kusurlu Mamüller	**Kusursuz Mamüller**	**Toplam**
Müzikle Üretim	102	8	110
Müziksiz Üretim	95	15	110
Toplam	197	23	220

Çözüm:

H_0 : Müzikle ve müziksiz üretimde kusurluluk açısından fark yoktur.
H_1 : Müzikle ve müziksiz üretimde kusurluluk açısından fark vardır.

	Kusurlu Mamüller	**Kusursuz Mamüller**	**Toplam**
Müzikle Üretim	102 **/98,5**	8 **/11,5**	110
Müziksiz Üretim	95 **/98,5**	15 **/11,5**	110
Toplam	197	23	220

$$\chi^2_{hes} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(102-98{,}5)^2}{98{,}5} + \frac{(95-98{,}5)^2}{98{,}5}$$

$$+ \frac{(8-11{,}5)^2}{11{,}5} + \frac{(15-11{,}5)^2}{11{,}5}$$

$$= 0{,}124 + 0{,}124 + 1{,}065 + 1{,}065 = 2{,}378$$

$$\chi^2_{0{,}05,1} = 3{,}841$$

$2{,}378 < 3{,}841$ H_0 kabul.

Müzikle üretim bir fark yaratmamıştır.

16.7.3. Ki-Kare Uygunluk Testi

χ^2 testinin uygulama biçimlerinden biri olan χ^2 ki-kare uygunluk testi ile, tahmin edilen frekanslara yani teorik frekanslara göre hesaplanan dağılımın, gerçek frekans yani gözlemlenen frekanslara göre hesaplanan dağılıma uygun olup olmadığı test edilir. Eğer gerçek frekans değerleri ile teorik frekans değerleri birbirine yakınsa χ^2 değeri küçük olacak ve böylece seçilen dağılımın teorik dağılıma uygun olduğu sonucuna varılacaktır. Diğer taraftan gerçek frekanslar ile teorik frekanslar arasında oldukça önemli farklar olduğunda ς^2 değeri büyüyecek ve bu durumda seçilen dağılımın teorik dağılıma uymadığı sonucuna varılacaktır.

Ki-Kare Uygunluk Testinin Aşamaları;

-Hipotezin düzenlenmesi
H_0 : Uygunluk var.
H_1 : Uygunluk yok.

-Hata payının belirlenmesi

Anlamlılık düzeyi 0,05 veya 0,01 olabilir.

-Test istatistiğinin hesaplanması

Test istatistiği yani ki-kare değeri aynı şekilde hesaplanır, yalnız burada işlemler bir kontenjans tablosu için değil de 2 sütundan oluşan bir bileşik seri için yapılır. Bu testte de önemli olan nokta test işlemlerinin yapılabilmesi için teorik frekansların her birinin değerinin en az 5 olmasıdır.

-Karar işlemi

Karar, test istatistiği, α ve serbestlik derecesine göre bulunan ki-kare tablosundan bulunan değerin karşılaştırılması ile alınır. Bu testte serbestlik derecesi yani $n = m - 1$'dir. m burada teorik frekansların hesaplanmasında kullanılan parametre sayısıdır.

$\chi^2 < \chi^2_{\alpha,\varsigma}$ ise H_0 kabul

$\chi^2 > \chi^2_{\alpha,\varsigma}$ ise H_0 reddedilir.

Karar işleminden sonra elde edilen sonuç yorumlanır.

ÖRNEK: Kusursuz olduğu kabul edilen bir zar 120 defa atılmış ve elde edilen sonuçlar kaydedilmiştir. Eğer zar kusursuz ise teorik olarak her bir yüz 20 defa görünecektir. Sonuçlar aşağıda tabloda görüldüğü gibidir.

	Zarın Yüzleri					
	1	2	3	4	5	6
Gerçek	20	22	17	18	19	24
Teorik	20	20	20	20	20	20

Gerçek ve teorik değerleri karşılaştırarak zarın kusurlu veya kusursuz olduğuna 0,05 hata payına göre karar veriniz.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i')^2}{f_i'}$$

$$= \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(17-20)^2}{20}$$

$$+ \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(24-20)^2}{20} = 1{,}7$$

Bu örnekte gözlemlenen verilerden elde edilen parametre toplam frekans olduğundan serbestlik derecesi,

$\nu = 6\text{-}1 = 5$

dir. Hata payı 0,05 olduğundan χ^2 tablosuna göre değer,

$\chi^2_{0,05}(5) = 11,070 \quad \nu = 5$ için

dir. Böylece 1,7 değeri kritik değerden küçük olduğundan yani,

$$\chi^2 > \chi^2_\alpha$$

olduğundan sıfır hipotezi kabul edilir. Yani zarın kusursuz olduğu sonucuna varılır.

16.8. ÇÖZÜMLÜ ÖRNEKLER

ÖRNEK 1: Herhangi bir pil imalatında, pillerin dayanma sürelerinin incelenmesi için aşağıda görülen tablo düzenlenmiştir. Tabloda görülen değerlere göre dayanma süreleri dağılımının yaklaşık olarak normal dağılıma uyduğu hipotezi test edilmek istenmektedir.

Sınıflar	fi	Fi'
1,45-1,95	2	0,6
1,95-2,45	1	2,6
2,45-2,95	4	6,8
2,95-3,45	15	10,7
3,45-3,95	10	10,3
3,95-4,45	5	6,1
4,45-4,95	3	2,2

Yukarıdaki tabloda teorik frekanslar normal dağılıma göre hesaplanmıştır. Bu frekansların hesaplanabilmesi için 40 birimlik pilden oluşan bir örnek düzenlenmiş ve bu örneğin ortalaması $\overline{X} = 3,4125$ ve standart sapması S=0,697 bulunmuştur. Bulunan bu değerler Z değerlerinin hesaplanmasında μ ve σ olarak kullanılacaktır. Örneğin 4. sınıfa karşılık gelen Z değerinin hesabını inceleyelim.

$$Z_1 = \frac{X_1 - \mu}{\sigma}$$

$$= \frac{2,95 - 3,4125}{0,697} = -0,664$$

$$Z_2 = \frac{X_2 - \mu}{\sigma}$$

$$= \frac{3{,}45 - 3{,}4125}{0{,}697} = 0{,}054$$

Hesaplanan bu Z değerlerine karşılık gelen değerler normal dağılım tablosuna göre,

$$Z_1 = 0{,}664 \rightarrow 0{,}2533$$

$$Z_2 = 0{,}054 \rightarrow 0{,}5215$$

Alan = P(-0,664 < Z < 0,054)

= P(Z<0,054) – P(Z<-0,664)

= 0,5215 – 0,2533

= 0,2682

dir. Alanın hesaplanmasından sonra 4. sınıf için teorik frekans

$f_4' = (0{,}2682)(40) = 10{,}7$

olarak bulunur.

Diğer sınıflar içinde teorik frekanslar aynı şekilde hesaplanır. Ancak önceki açıklamalarda değinildiği gibi teorik frekansların 5'den az olmaması gerekir. Bu nedenle tabloda da görüldüğü gibi 5'den az olan teorik frekanslar birbirleriyle birleştirilmiştir. Bu işlemlerden sonra χ^2 değeri hesaplanarak karara varılır. Yani,

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i')^2}{f_i'}$$

$$= \frac{(7-10)^2}{10} + \frac{(15-10{,}7)^2}{10{,}7} + \frac{(10-10{,}3)^2}{10{,}3} + \frac{(8-8{,}3)^2}{8{,}3}$$

$$= 2{,}639$$

$\chi^2_{0{,}05} = 3{,}841$ $\nu = 4\text{-}3 = 1$ için

olarak hesaplanır. Böylece

$$\chi^2 > \chi_\alpha^2$$

olduğundan sıfır hipotezi kabul edilir. Yani pillerin dayanma süreleri dağılımın normal dağılıma uyduğu sonucuna varılır.

ÖRNEK 2: Meteorolojik koşulların tarımsal üretim üzerindeki etkisinin araştırıldığı bir bölgede, Haziran ayı sıcaklığı 17^0 kabul edilerek inceleme yapılacaktır. Bu varsayımın geçerliliğini sınamak amacı ile Haziran ayı için 12 yıllık ortalama sıcaklık değerleri saptanmıştır. 0,05 hata payı ile iddiayı test ediniz.

18,4 19,6 15,2 19,7 17,0 18,3 17,8 16,8 18,3 16,9 17,2 17,4

Çözüm:

$$H_0 : M_0 = 17{,}0$$
$$H_1 : M_A \neq 17{,}0$$

Gözlemler :	18,4	19,6	15,2	19,7	17,0	18,3	17,8	16,8	18,3	16,9	17,2	17,4
Farklar :	1,4	2,6	-1,8	2,7	0	1,3	0,8	-0,2	1,3	-0,1	0,2	0,4
Farkların İşaretleri :	+	+	-	+	0	+	+	-	+	-	+	+

(-) farkların sayısı, (+) farkların sayısından daha küçük olduğundan K=3 alınır.

$$Z = \frac{K - n/2}{(n/4)^{1/2}} = \frac{3 - (12/2)}{(12/4)^{1/2}} = \frac{3-6}{\sqrt{3}} = -1{,}73$$

P(Z<-1,73) = 0,5 – 0,4582

= 0,0418 > 0,025

olduğundan H_0 red edilir. Sözkonusu bölgede Haziran ayı sıcaklığının 17^0 kabul edilmesi hatalara yol açacaktır.

ÖRNEK 3: Sigara kullanımının yol açtığı kalsiyum eksikliği üzerinde yapılan bir incelemede, günde en fazla bir paket sigara içen 20 kişi ile günde en az bir paket sigara içen 20 kişinin vücutlarındaki kalsiyum miktarı tespit edilmiştir. Gruplar arasında fark olup olmadığını 0,05 anlam düzeyinde test ediniz.

A En Fazla Bir Paket İçenler	B En Az Bir Paket İçenler
80	62
78	64
65	52
90	48
55	50
87	62
66	45
98	54
79	65
70	55
80	49
74	57
69	51
72	60
58	56
80	55
75	46
72	52
92	58
74	59

Çözüm:

A En Fazla Bir Paket İçenler	B En Az Bir Paket İçenler	A Sıra No	B Sıra No
80	62	35	19,5
78	64	32	21
65	52	22,5	7,5
90	48	38	3
55	50	11	5
87	62	37	19,5
66	45	24	1
98	54	40	9
79	65	33	22,5
70	55	26	11
80	49	35	4
74	57	29,5	14

69	51	25	6
72	60	27,5	18
58	56	15,5	13
80	55	35	11
75	46	31	2
72	52	27,5	7,5
92	58	39	15,5
74	59	29,5	17
		593	227

H_0 : İki grubun kalsiyum miktarı arasında fark yoktur.
H_1 : İki grubun kalsiyum miktarı arasında fark vardır.

$$\overline{X}_w = \frac{n_A(n_A + n_B + 1)}{2} = \frac{20(20 + 20 + 1)}{2} = 410$$

$$\sigma_w = \sqrt{\frac{n_A(n_B)(n_A + n_B + 1)}{12}} = 36{,}968$$

$$Z = \frac{w - \overline{X}_w}{\sigma_w} = \frac{593 - 410}{36{,}968} = 4{,}95$$

$Z_{0,025} = 1{,}96$ $\quad$ $4{,}95 > 1{,}96$ H_0 red.
İki grup arasında anlamlı bir fark vardır.

ÖRNEK 4: 2009 Avrupa Buz Pateni Şampiyonası'nda çiftler kategorisinde yarışan 10 çifte, iki hakem tarafından verilen teknik puanlar karşılaştırılmak isteniyor. 0,05 anlam düzeyinde anlamlı bir fark olup olmadığını test ediniz.

1. Hakemin Verdiği Puanlar	**2. Hakemin Verdiği Puanlar**
4,8	4,6
5,2	4,8
5,4	5,5
5,8	5,9
4,4	4,7
4	4,4
5,3	5,1
5	5,2
4,6	4,8
5,4	5,2

Çözüm:

1.Hakemin Verdiği Puanlar	2. Hakemin Verdiği Puanlar	Fark Değerleri	Mutlak Farkların Sıraları	İşaretli Farkların Sıraları
4,8	4,6	0,2	5	5
5,2	4,8	0,4	9,5	9,5
5,4	5,5	-0,1	1,5	-1,5
5,8	5,9	-0,1	1,5	-1,5
4,4	4,7	-0,3	8	-8
4	4,4	-0,4	9,5	-9,5
5,3	5,1	0,2	5	5
5	5,2	-0,2	5	-5
4,6	4,8	-0,2	5	-5
5,4	5,2	0,2	5	5
				-6

T = -6

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{6}} = \sqrt{\frac{10(11)(21)}{6}} = 19{,}62$$

$$Z = \frac{T}{\sigma_T} = \frac{-6}{19{,}62} = -0{,}30$$

$\alpha = 0{,}05\ Z_{0{,}025} = 1{,}96$

0,30 < 1,96 olduğundan H_0 kabul edilir. İki hakemin değerlendirmeleri arasında anlamlı bir fark yoktur.

ÖRNEK 5: Türkiye'de, protein tüketimi açısından Marmara, Akdeniz ve Karadeniz bölgeleri arasında fark olup olmadığı incelenecektir. Bu amaçla her bölgeden seçilen 2000'er ailenin, aylık tüketimleri kg. cinsinden tespit edilmiştir. 0,05 hata payı ile test ediniz.

	Marmara	Akdeniz	Karadeniz
Kırmızı Et	3200	2960	2600
Balık	2800	3250	3800
Tavuk	2400	2500	2200

Çözüm:

	Marmara	Akdeniz	Karadeniz
Kırmızı Et	3200**(7)**	2960**(6)**	2600**(4)**
Balık	2800**(5)**	3250**(8)**	3800**(9)**
Tavuk	2400**(2)**	2500**(3)**	2200**(1)**
w_i	14	17	14
w_i^2 / n_i	65,33	96,33	65,33

H_0 : Bölgelerin protein tüketimi arasında fark yoktur.
H_1 : Bölgelerin protein tüketimi arasında fark vardır.

$$H = \left[\frac{12}{n(n+1)} \sum \frac{w_i^2}{n_i} \right] - 3(n+1) = \left[\frac{12}{9(10)} (226,99) \right] - 3(10) = 0,265$$

$$\chi^2_{0,05,2} = 5,991$$

$0,265 < 5,991$ olduğundan H_0 kabul.

Bölgelerin protein tüketimi arasında anlamlı bir fark yoktur.

ÖRNEK 6: Bir kolej yöneticisi, öğrencilerinin üniversiteye giriş sınavında aldıkları sözel ve sayısal puanlar arasındaki ilişkiyi incelemektedir. Bu amaçla seçilen 8 öğrencinin sözel ve sayısal puanları aşağıda verilmiştir. 0,01 hata payı ile test ediniz.

Sözel Puan	Sayısal Puan
143	139,3
156,2	164,2
170,8	176,4
146,4	141,2
120	118,9
138,2	127,6
145,7	135,7
168,3	164,5

Çözüm:

Sözel Puan	Sayısal Puan	Sözel Sıra No	Sayısal Sıra No	D^2
143	139,3	3	4	1
156,2	164,2	6	7	1
170,8	176,4	8	8	0
146,4	141,2	5	5	0
120	118,9	1	1	0
138,2	127,6	2	2	0
145,7	135,7	4	3	1
168,3	164,5	7	6	1
				4

$$H_0 : \rho_s = 0$$
$$H_1 : \rho_s \neq 0$$
$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(4)}{8(64 - 1)} = 0{,}95$$

çok kuvvetli doğru yönlü ilişki

$\alpha = 0{,}01$ ve n =8 için r tablo değeri 0,8333'tür.

0,95 > 0,833 olduğundan H_0 red edilir. Sayısal ve sözel puanlar arasında ilişki vardır.

ÖRNEK 7: Bir hukuk fakültesinde, 2. sınıf öğrencilerinin yabancı dil dersleri iki farklı öğretim üyesi tarafından yürütülmektedir. Öğretim üyelerinin eğitim verdiği iki ayrı şubenin başarılarının aynı olup olmadığı test edilecektir. Her iki şubeden seçilen 10'ar öğrencinin aşağıdaki notlarını kullanarak 0,05 hata payı ile testi yapınız.

A Şubesi	B Şubesi
45	50
65	45
70	85
50	40
80	70
55	65
75	90
60	60
80	55
60	65

Çözüm:

A Şubesi	B Şubesi	Sıra No A	Sıra No B
45	50	2,5	4,5
65	45	12	2,5
70	85	14,5	19
50	40	4,5	1
80	70	17,5	14,5
55	65	6,5	12
75	90	16	20
60	60	9	9
80	55	17,5	6,5
60	65	9	12
		109	101

H_0 : Sınıfların başarıları arasında fark yoktur.
H_1 : Sınıfların başarıları arasında fark vardır.

$$U = n_1 n_2 + \frac{n_1(n_2+1)}{2} - R_1 = 10(10) + \frac{(10)(11)}{2} - 109 = 46$$

$$\overline{X}_U = \frac{n_1 n_2}{2} = 50$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13{,}23$$

$$Z = \frac{U - \overline{X}_U}{\sigma_u} = \frac{46 - 50}{13{,}23} = -0{,}3023$$

$Z_{0,025} = 1{,}96$ $\qquad$ $0{,}3023 < 1{,}96$ H_0 kabul edilir.

Sınıfların başarıları arasında fark yoktur.

ÖRNEK 8: Pancar üretimi yapılan bir bölgede üç tür gübrenin verime etkileri araştırılacaktır. Bir ekim dönemi sonucunda gözlenen hasat sonuçları aşağıdaki kontenjans tablosunda verilmiştir. 0,05 anlam düzeyinde testi yapınız.

	Ürün		
Gübre	**Sağlam**	**Kusurlu**	**Toplam**
I	430	20	450
II	375	25	400
III	455	45	500
Toplam	1260	90	1350

Çözüm:

	Sağlam	Kusurlu	Toplam
I	430 **/420**	20 **/30**	450
II	375 **/373,3**	25 **/26,7**	400
III	455 **/466,7**	45 **/33,3**	500
Toplam	1260	90	1350

H_0 : Gübrelerin verime etkileri arasında fark yoktur.
H_1 : Gübrelerin verime etkileri arasında fark vardır.

$$\chi^2_{hes} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(430-420)^2}{420} + \frac{(20-30)^2}{30} + \frac{(375-373{,}3)^2}{373{,}3}$$

$$+ \frac{(25-26{,}7)^2}{26{,}7} + \frac{(455-466{,}7)^2}{466{,}7} + \frac{(45-33{,}3)^2}{33{,}3}$$

$$= 0{,}238 + 3{,}33 + 0{,}008 + 0{,}108 + 0{,}293 + 4{,}111 = 8{,}088$$

$\chi^2_{0,05,2} = 5{,}991$

$8{,}088 > 5{,}991$ H_0 red.

Gübrelerin verimleri birbirinden farklıdır.

ÖRNEK 9: Bir fast food işletmesinde müşteri gelme sıklığının dakikada ortalama 2 müşteri ile Poisson dağıldığı bilinmektedir. Bir hafta içinde 150 farklı zamanda gözlem yapılmış ve dakikada gelen müşteri sayıları kaydedilmiştir. Örneklem dağılımının Poisson dağılımına uygunluğunu %95 güvenle test ediniz.

Müşteri Sayısı Xi	Gözlenen Zaman fi
0	12
1	27
2	48
3	31
4	15
5	11
6 ve üstü	6
	150

Çözüm:

Müşteri Sayısı X_i	Gözlenen Zaman f_i	Beklenen Frekans $E(f_i)$	$P(X_i)$	A $\left(\lvert f_i - E(f_i)\rvert - 0{,}5\right)^2$	A/ $E(f_i)$
0	12	20,25	0,135	45,5625	2,25
1	27	40,65	0,271	172,9225	4,25
2	48	40,65	0,271	46,9225	1,15
3	31	27	0,18	12,25	0,46
4	15	13,5	0,09	1	0,07
5	11	5,4	0,036	26,01	4,82
6 ve üstü	6	2,55	0,017	8,7025	3,41
	150	150	1		16,41

$$p(x = X_i) = \frac{e^{-\lambda}\lambda^x}{X!} = \frac{e^{-2}2^x}{X!}$$

$P(X \geq 6) = 1 - P(X < 6)$

$E(f_i) = n.p(x = X_i)$

$$\chi_h^2 = \sum_{i=1}^{r} \frac{[f_i - E(f_i)]^2}{E(f_i)} = \left[\sum_{i=1}^{r} \frac{f_i^2}{E(f_i)}\right] - n$$

Yates düzeltmesi ile,

$$\chi_h^2 = \sum_{i=1}^{r} \frac{[\,|f_i - E(f_i)| - 0{,}5]^2}{E(f_i)} = 16{,}41$$

H_0 : Örneklem dağılımı Poisson dağılımına uygundur.
H_1 : Örneklem dağılımı Poisson dağılımına uygun değildir.

Serbestlik derecesi $v = r-k-1 = 7-1-1 = 5$

$\chi^2_{5,0,05} = 11{,}07 < \chi_h^2 \Rightarrow H_0$ red

Örneklem dağılımı Poisson dağılımına uygun değildir.

ÖRNEK 10: Bir fabrikada işçilere üretim teknikleri eğitimi verilecektir. İşçiler, bir birimi üretim sürelerine göre sınıflandırılmıştır. Eğitim alan ve almayan işçilerden 50'şer kişilik iki örnek seçilerek aşağıdaki tablo hazırlanmıştır. ($\alpha = 0{,}01$)

Süre (Dk)	(A) Eğitimli	(B) Eğitimsiz	Toplam
10'un Altı	18(12)	6(12)	24
10-20	18(16)	14(16)	32
20-30	8(14)	20(14)	28
30 Üzeri	6(8)	10(8)	16
	50	50	100

$$\chi_h^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{\left(f_{ij}\right)^2}{E\left(f_{ij}\right)} - n = 112{,}643 - 100 = 12{,}643$$

$$= \frac{(18-12)^2}{12} + \frac{(6-12)^2}{12} + \frac{(18-16)^2}{16} + \frac{(14-16)^2}{16} + \frac{(8-14)^2}{14} +$$

$$\frac{(20-14)^2}{14} + \frac{(6-8)^2}{8} + \frac{(10-8)2}{8} = 12{,}643$$

$$E\left(f_{ij}\right) = \frac{n_i.n_j}{n}$$

v = (r-1).(c-1) = (4-1).(2-1) = 3

$$\chi^2_{3,0,01} = 11{,}34$$

H_0 : Eğitime göre üretim süreleri dağılımları homojendir.
H_1 : Eğitime göre üretim süreleri dağılımları homojen değildir.

$$\chi^2_{3,0,01} < \chi^2_h \Rightarrow H_0 \text{ red}$$

Eğitim alanlar ve almayanların üretim sürelerinin dağılımları homojen (benzer) değildir.

ÖRNEK 11: Bir fabrikada üretilen oto lastiklerinin dayanıklılık süresinin, lastiğin yapımında kullanılan X mamülünün katkı derecesi ile ilişkisi olup olmadığı araştırılmak isteniyor. X mamülü, lastiğe 1-3 br., 3-5 br., ya da 5-7 br. katılabiliyor. Lastiklerin dayanıklılıkları ise 50.000 km'den az, 50.000-60.000 km ve 60.000 km'den çok olarak belirlenmiş ve incelenen 200 lastikte aşağıdaki sonuçlar bulunmuştur.

Lastiğin Dayanıklığı	X Mamülünün 1-3 br.	Katkı 3-5 br.	Derecesi 5-7 br.	Toplam
>50.000	35	25	10	70
50.000-60.000	40	30	10	80
60.000<	25	15	10	50
Toplam	100	70	30	200

f'_{11} = 70.100 / 200 = 35 f'_{12} = 70.70 / 200 = 24,5 f'_{13} = 70.30 / 200 = 10,5

f'_{21} = 80.100 / 200 = 40 f'_{22} = 80.70 / 200 = 28 f'_{23} = 80.30 / 200 = 12

f'_{31} = 50.100 / 200 = 25 f'_{32} = 50.70 / 200 = 17,5 f'_{33} = 50.30 / 200 = 7,5

$$\chi^2_{hesap} = \sum_{ij=1}^{n} \frac{\left(f_{ij} - f'_{ij}\right)}{f'_{ij}} = \frac{(35-35)^2}{35} + ... + \frac{(10-7{,}5)^2}{7{,}5} = 1{,}70$$

$$\chi^2_{0,05,4} = 9{,}488$$

$\chi^2_{hesap} < \chi^2_{0,05;4}$ H_0 kabul

ÖRNEK 12: Yazı tura denemesinde kullanılan bir paranın hileli olup olmadığı araştırılmak istenmektedir. Para hilesiz ise Y ve T sonuçlarının aynı olasılıkla gelmesini gösterir. Bunun için para 100 defa atılıyor ve 65Y 35 defa da T geliyor. Paranın hileli olup olmadığını test ediniz.

H_0 : Para hilesizdir. (Y gelme olasılığı 0,50'dir.)
H_1 : Para hilelidir. (Y gelme olasılığı 0,50'den farklıdır.)

Sonuçlar	Gözlenen f_i	Teorik f_i'
Y	65	50
T	35	50

$$\chi^2_{hesap} = \sum_{ij=1}^{n} \frac{\left(f_{ij} - f'_{ij}\right)}{f'_{ij}} = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} = 9$$

$$\chi^2_{0,05,1} = 3{,}84$$

$\chi^2_{hesap} = 9 > \chi^2_{0,05;1} = 3{,}84 \Rightarrow$ H_0 red

Para hilelidir. Dağılımı düzgün dağılıma uymaz.

ÖRNEK 13: Ekonometri bölümü öğrencilerinin, istatistik dersindeki başarılarının, matematik dersindeki başarılarına bağlı olup olmadığı 0,05 anlamlılık düzeyinde araştırılmak isteniyor. Bu amaçla 100 öğrencinin istatistik ve matematik derslerinden başarılı-başarısızlık durumları aşağıdaki tabloda belirtilmiştir.

Kontenjans Tablosu

Matematik Dersinden	İstatistik Dersinden **Başarılı**	**Başarısız**	Toplam
Başarılı	35	13	48
Başarısız	30	22	52
Toplam	65	35	100

H_0 : İstatistik dersinden başarı durumu, matematik dersinden başarı durumundan bağımsızdır.
H_1 : İstatistik dersinden başarı durumu, matematik dersinden başarı durumuna bağlıdır.

f'_{11} = 48.65 / 100 = 31,2 f'_{12} = 48.35 / 100 = 16,8
f'_{21} = 52.65 / 100 = 33,8 f'_{22} = 52.35 / 100 = 18,2

$$\chi^2_{hesap} = \sum_{ij=1}^{n} \frac{\left(f_{ij} - f'_{ij}\right)}{f'_{ij}} = \frac{(35-31{,}2)^2}{31{,}2} + \frac{(30-33{,}8)^2}{33{,}8}$$

$$+ \frac{(13-16{,}8)^2}{16{,}8} + \frac{(22-18{,}2)^2}{18{,}2} = 2{,}54$$

$\chi^2_{0,05,1} = 3{,}84$

$\chi^2_{hesap} < \chi^2_{0,05;1} \Rightarrow$ H_0 kabul

İstatistik dersinden başarı durumu, matematik dersinden bağımsızdır.

ÖRNEK 14: Motor güçleri ile motor hacimleri teknolojik gelişme nedeniyle bağımsız olmaya başladıkları iddia edilmektedir. Yeni model arabalardan 200 farklı model tesadüfi olarak seçilmiş ve sınıflandırılmıştır. İki özelliğin bağımsızlığını %10 anlam seviyesinde test ediniz.

Motor Hacmi(ccm)	Motor Gücü(hp) **50-80**	**80-110**	Sınıfları **110 üstü**	Toplam
1400 altı	5,5(24)	5(24)	0(12)	60
1400-2000	15(32)	45(32)	20(16)	80
2000 üstü	10(24)	30(24)	20(12)	60
Toplam	80	80	40	200

$$\chi_h^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{[f_i - E(f_i)]^2}{E(f_i)} = 97{,}396$$

v = (c-1).(r-1) = (3-1).(3-1) = 4

$$\chi_{4,0,1}^2 = 7{,}78$$

H_0 : Motor gücü ve motor hacimleri birbirlerinden bağımsızdır.
H_1 : Motor gücü ve motor hacimleri birbirlerinden bağımsız değil.

$$\chi_h^2 > \chi_{v,0,1}^2 \Rightarrow H_0 \text{ red}$$

İlgilenilen özellikler birbirlerinden bağımsız değillerdir.