



İSTATİSTİK

ORTAK DERS

EDİTÖR

DOÇ.DR. SEMA ULUTÜRK AKMAN

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

ISBN 978-605-07-0686-4



İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ



İSTATİSTİK

ORTAK DERS

EDİTÖR

DOÇ.DR. SEMA ULUTÜRK AKMAN

Yazar Notu

Elinizdeki bu eser, İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi'nde okutulmak için hazırlanmış **bir ders notu niteliğindedir.**

ÖNSÖZ

Pek çoğumuz günlük yaşamımız içinde istatistik sözcüğünü duymuş ve hattâ zaman zaman kullanmışızdır. Örneğin, “İstatistiklere göre, son 50 yılın en kurak yazını yaşıyoruz.” gibi bir haber okuduğumuzda son 50 yılın yağış verileri üzerinden bir değerlendirme yapıldığını anlıyoruz. Ya da hafta sonunda oynanan bir derbi maçı ile ilgili spor haberlerine göz gezdirirken, takımların topa sahip olma oranı, top kaybı oranı, ikili mücadelede kazanma oranı, isabetli pas oranı gibi istatistikler ile karşılaşılıyor ve bu sayede maçın görünen skoru yanında maçın 90 dakikalık süreci ile ilgili fikir sahibi olabiliyoruz. Yine benzer şekilde, gazetede “Boşanma istatistiklerine bakıldığında son yıllarda bir artış gözleniyor.” gibi bir ifade ile boşanan çift sayısının artmakta olduğunu anlıyoruz.

Günlük hayatımız içinde karşılaşabileceğimiz bu tür örnekleri arttırabiliriz. Bu örneklerin ortak noktası bir olaya ilişkin sayısal gözlem değerleri, sayısal veriler ifade ediyor olmasıdır. Dolayısıyla, günlük yaşam içinde istatistik; gözlem ya da ölçüm yoluyla oluşturulan sayısal veriler olarak algılanmaktadır.

Özünde sayısal veriler içermekle beraber, istatistik veri toplama ve toplanan verileri sunma dışında fonksiyonlara sahiptir. O halde istatistik nedir?

İstatistik; veri toplama, verileri düzenleme, özetleme ve sunmanın yanı sıra; tahmin, modelleme gibi pek çok yöntemi bünyesinde barındıran ve çok yaygın kullanım alanına sahip bir bilim dalıdır. Bünyesinde, enflasyon hesaplarından rating ölçümlerine, kamuoyu araştırmalarından hayat tablolarına kadar pek çok teknik ve çözümlemeyi barındırır.

Sayıları bir alfabe olarak düşünürsek, istatistik sayıların dilini anlamak ve onları okuyabilmektir. Sayılar ve rakamlar bizlere bir şeyler anlatır, bilgi verir. Dolayısıyla istatistik sayıların dilini anlamaktır.

İstatistik günümüzde sağlık bilimlerinden mühendisliğe, sosyal bilimlerden fen bilimlerine kadar hemen her alanda yer alan bir disiplin olarak karşımıza çıkmaktadır. Akademik eğitim içinde gerek lisans gerekse lisansüstü düzeyde verilmekte olan istatistik eğitiminin iş hayatında da karşılığı olduğunu ve pek çok alanda kullanıldığını görmekteyiz.

İstatistiğe gönül vermiş bir istatistikçinin kaleme aldığı bu kitap, sadece benim değil içinde yetiştiğim kürsü geleneğinin de bir sonucu olarak, İstanbul Üniversitesi İktisat Fakültesi İstatistik Enstitüsü’nün Ord.Prof.Dr. Ömer Celal Sarç ile başlayan ve benim kıymetli Hocam, idolüm Prof.Dr. Haluk Cillov ile devam eden kürsü misyonuyla oluşan, Prof.Dr. Durmuş Dünder, Doç.Dr. Kayıhan Özoğuz ve Prof.Dr. Fazıl Güler’in katkılarıyla şekillenen akademik hayatımın bir yansıması olarak ortaya çıktı. Dolayısıyla, her satırında sonsuz saygı ve sevgi duyduğum hocalarımın izleri var. Hepsine ayrı ayrı müteşekkirim.

Geçmişte öğrencim, bugün ise meslektaşım olan sevgili Hakan Bektaş da bahsettiğim kürsü geleneğinin benden sonraki halkası olarak kitaba bölüm yazarı düzeyinde katkı sağladı, kendisine de teşekkür ediyorum. Onunla çalışmak her zaman çok keyifli...

Ayrıca, kitabı büyük bir titizlikle okuyarak, eleştiri ve önerileriyle katkı sağlayan eski öğrencim, çiçeği burnunda meslektaşım sevgili Ege Yeşilaltay’a da teşekkür etmek isterim.

Kitabın, başta İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi öğrencileri olmak üzere, istatistiğe ilgi duyanlara ve istatistiği iş hayatında kullananlara faydalı olması dileğiyle...

DOÇ.DR. SEMA ULUTÜRK AKMAN

İÇİNDEKİLER

ÖNSÖZ.....	I
KISALTMALAR	VIII
YAZAR/YAZARLAR NOTU	IX
1. İSTATİSTİĞİN TANIMI KONUSU VE TEMEL KAVRAMLAR.....	1
1.1. İstatistiğin Tanımı ve Konusu	6
1.1.1. Tarifsel (Tanımlayıcı / Betimsel) İstatistik.....	6
1.1.2. Tümevarım (Çıkarsama) İstatistiği	7
1.2. Verilerin Toplanması ve Rölöve (Derleme) Kavramı	8
1.3. İstatistik Metodolojisine Yönelik Temel Kavramlar	10
1.3.1. Anakütle, Örnek Kütle, Parametre ve İstatistik.....	10
1.3.2. Rölöve Hataları.....	13
2. BİRİM, VASIF, ŞIK KAVRAMLARI VE VERİNİN ÖLÇÜM DÜZEYİ.....	19
2.1. İstatistikte Birim Kavramı ve Çeşitleri.....	24
2.2. Vasıf ve Şık Kavramı	25
2.3. Verinin Ölçüm Düzeyi	27
4. VERİLERİN DÜZENLENMESİNDE TASNİF VE GRUPLAMA	35
3.1. Tasnif.....	41
3.2. Gruplama	43
4. VERİLERİN DÜZENLENMESİNDE SERİLER VE GRAFİKLER.....	56
4.1. Seriler	61
4.1.1. Zaman Serileri	61
4.1.2. Mekân Serileri	63
4.1.3. Panel Veriler.....	64
4.1.4. Bileşik Seriler	65
4.1.5. Bölünme (Frekans) Serileri	66
4.1.5.1. Basit Seri	66

4.1.5.2. Tasnif Edilmiş Seri	68
4.1.5.3. Sınıflanmış Frekans Serisi	70
4.2. Sınıf Genişliğini Tayin Etmede Sturge Kuralı	74
4.3. Verinin Grafik Yardımıyla Sunulması	76
4.3.1. Nitel Veriler için Grafikler	77
4.3.2. Nicel Veriler için Grafikler	80
5. ANALİTİK ORTALAMALAR – ARİTMETİK ORTALAMA	90
5.1. Ortalamalar	98
5.1.1. Analitik Ortalamalar	98
5.1.2. Analitik Olmayan Ortalamalar	98
5.2. Aritmetik Ortalama.....	99
5.2.1. Basit Serilerde Aritmetik Ortalamanın Hesabı.....	100
5.2.2. Tasnif Edilmiş Serilerde Aritmetik Ortalamanın Hesabı	101
5.2.3. Sınıflanmış Seride Aritmetik Ortalamanın Hesabı.....	104
5.2.4. Ağırlıklı Aritmetik Ortalama	107
5.2.5. Aritmetik Ortalamanın Özellikleri	109
6. ANALİTİK ORTALAMALAR – GEOMETRİK VE KARELİ ORTALAMALAR	116
6.1. Geometrik Ortalama	121
6.1.1. Basit Seride Geometrik Ortalamanın Hesabı	121
6.1.2. Tasnif Edilmiş Seride Geometrik Ortalamanın Hesabı	124
6.1.3. Sınıflanmış Seride Geometrik Ortalamanın Hesabı	125
6.1.4. Geometrik Ortalamanın Özellikleri	127
6.2. Kareli Ortalama	127
6.2.1. Basit Seride Kareli Ortalama Hesabı.....	128
6.2.2. Tasnif Edilmiş Serilerde Kareli Ortalamanın Hesabı	129
6.2.3. Sınıflanmış Frekans Serilerinde Kareli Ortalamanın Hesabı	129
7. VERİLERİN ÖZETLENMESİ: ANALİTİK OLMAYAN ORTALAMALAR	137

7.1. Mod (Egemen Değer)	143
7.1.1. Basit Seride Mod'un Belirlenmesi	143
7.1.2. Tasnif Edilmiş Seride Modun Belirlenmesi	144
7.1.3. Sınıflanmış Frekans Serilerinde Modun Belirlenmesi	145
7.1.4. Mod'un Özellikleri ve Kullanım Alanı	148
7.2. Medyan	148
7.2.1. Basit Seride Medyan Hesabı	149
7.2.2. Tasnif Edilmiş Seride Medyan Hesabı	151
7.2.3. Sınıflanmış Seride Medyan Hesabı	153
7.2.4. Medyanın Özellikleri	155
7.3. Ortalamalar Arasındaki İlişkiler ve Bir Dağılımın Simetri Durumunun Değerlendirilmesi	155
8. DAĞILMA ÖLÇÜLERİ	165
8.1. Dağılma Ölçüleri	172
8.2. Değişim Aralığı	172
8.2.1. Tasnif Edilmiş ve Sınıflanmış Frekans Serilerinde Değişim Aralığı	174
8.2.2. Değişim Aralığının Özellikleri	174
8.3. Ortalama Sapma	175
8.3.1. Basit Seride Ortalama Sapmanın Hesabı	176
8.3.2. Tasnif Edilmiş Seride Ortalama Sapmanın Hesabı	177
8.3.3. Sınıflanmış Seride Ortalama Sapmanın Hesabı	178
8.3.4. Ortalama Sapmanın Özellikleri	180
8.4. Standart Sapma ve Varyans	180
8.4.1. Basit Seride Standart Sapma ve Varyansın Hesabı	181
8.4.2. Tasnif Edilmiş Serilerde Standart Sapma ve Varyansın Hesabı	183
8.4.3. Sınıflanmış Serilerde Standart Sapma ve Varyansın Hesabı	184
8.4.4. Standart Sapma ve Varyans Hesabında Başka Bir Formül	186
8.4.5. Standart Sapmanın Özellikleri	188

9. NİSPİ DAĞILMA ÖLÇÜLERİ	197
9.1. Değişim Katsayısı.....	203
9.2. Standart Değer	206
10. OLASILIK – TEMEL KAVRAMLAR	214
10.1. Olasılık Kavramı	220
10.1.1. Sübjektif Olasılık.....	220
10.1.2. Objektif Olasılık - Frekans Olasılığı	221
10.2. Olasılık Kuramına Yönelik Temel Kavramlar:	222
10.2.1. Olasılık Bir Limit Durumdur	223
10.2.2. Olasılığın Sınırları	223
10.3. Basit ve Birleşik Olaylar	224
10.3.1. Basit Olaylar ve Basit Olaylarda Olasılık Hesabı	224
10.3.2. Birleşik Olaylar ve Birleşik Olaylarda Olasılık Hesabı	225
10.3.2.1. Bağımsız Olaylar.....	226
10.3.2.2. Bağımlı Olaylar.....	229
10.4. Koşullu Olasılık.....	231
10.5. Bağımsızlık Kavramı ve Bağımsız Olaylar	234
10.6. Bağımlı Olaylar	236
10.7. Permütasyon ve Kombinasyon	238
10.7.1. Permütasyon	238
10.7.2. Kombinasyon.....	239
10.8. Bağımlı Olaylar Olasılığı Hesabına Kombinasyon Yaklaşımı:.....	239
11. OLASILIK - TESADÜFİ DEĞİŞKEN, OLASILIK DAĞILIMI VE BEKLENEN DEĞER	246
11.1. Tesadüfi (Rastlantısal) Değişken.....	252
11.2. Olasılık Dağılımı	253
11.3. Beklenen Değer (Matematik Ümit)	256
12. NORMAL DAĞILIM	266

12.1. Neden “Normal” Dağılım?	272
12.2. Normal Dağılım: Kuramsal Çerçeve	273
12.3. Standart Normal Dağılım	275
12.4. Standart Normal Dağılım Tablosu Kullanılarak Olasılıkların Belirlenmesi	277
12.5. Normal Dağılım Olasılıklarını Kullanarak Beklenen Değer Hesabı	290
13. BASİT İNDEKSLER	300
13.1. İndeks Türleri	307
13.2. Basit İndeks	307
13.2.1. Sabit Esaslı Basit İndeks	307
13.2.2. Değişik Esaslı Basit İndeks	309
13.2.3. Sabit Esaslı İndeksten Değişik Esaslı İndekse Geçiş	310
13.2.4. Değişik Esaslı İndeksten Sabit Esaslı İndekse Geçiş	311
13.2.5. Sabit Esaslı İndeks Serisinde Esas Dönemin Değiştirilmesi	311
14. BİLEŞİK İNDEKSLER	320
14.1. Bileşik İndeksler	326
14.1.1. Tartısız Bileşik İndeksler	327
14.1.2. Tartılı Bileşik İndeksler	328
14.1.2.1. Laspeyres Fiyat İndeksi	330
14.1.2.2. Paasche Fiyat İndeksi	331
14.1.2.3. Fisher’in İdeal İndeksi	332
14.2. Cari Fiyat Serisinin Deflate Edilmesi	333

KISALTMALAR

TÜİK : Türkiye İstatistik Kurumu
vb. : ve benzeri
vd. : ve diğerleri

YAZAR/YAZARLAR NOTU

Bu ders kitabı, İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi müfredatında “İstatistik” dersi olan tüm programlara uygun olacak şekilde tasarlanmıştır. Kitabın editörleri ve bölüm yazarları aşağıda listelenmiştir:

1. İSTATİSTİĞİN TANIMI KONUSU VE TEMEL KAVRAMLAR

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde istatistiğin tanımı ve kapsamına yönelik açıklamalar ve istatistiğe yönelik temel kavramlar ele alınacaktır.

Bu bağlamda, öncelikle istatistiğin tanımına yer verilecek, daha sonra veri, veriye yönelik temel kavramlar, veri toplama, toplanan verileri düzenleme konuları ayrıntılı olarak incelenecektir.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Sayıların dili olur mu?
- 2) Sayıların konuştuğu gizemli dil: İstatistik nedir?
- 3) Anketler, kamuoyu araştırmalarının temel dayanağı nedir?
- 4) Veri toplama aşamasında hata yapmak mümkün müdür?
- 5) İstatistik ile ilgili çalışmalarda bilinmesi gereken temel kavramlar nelerdir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistik nedir? İstatistiğin konusu nelerdir?	Gerek fen bilimleri gerekse sosyal bilimler alanında kullanılmakta olan istatistik ne olduğu ve ne işe yaradığı konusunda fikir sahibi olur.	Temel düzeyde istatistik kitapları ve yayınlanan istatistikler.
Tipik olay ve kolektif olay nedir?	İstatistiğin konusunu oluşturan kolektif olay kavramını bilir.	Temel düzeyde istatistik kitaplarında istatistiğin konusu ile ilgili bölümler.
Veri toplama işlemi nedir ve nasıl yapılır?	İstatistiğin temelini veri toplamak oluşturduğuna göre, doğru veri toplamak nasıl olacağı konusunda bilgi sahibidir.	Temel düzeyde istatistik kitaplarının veri toplama ile ilgili bölümleri.
Veri toplama aşamasında dikkat edilmesi gereken hususlar.	Veri toplama sonucunda elde edilen ham verilerin nasıl ve ne şekilde kullanılabilir hale getirilebileceğini bilir.	Verilerin düzenlenmesine yönelik, tasnif, gruplama, seriler oluşturma vb. yöntemlerin anlatıldığı istatistik kitapları.
Veri toplama aşamasında yapılabilecek hata türleri.	Tesadüfi ve sistematik hata konusunda bilgi sahibidir.	Temel düzeyde istatistik kitaplarının veri toplama ile ilgili bölümleri.

Anahtar Kavramlar

- İstatistik: Veri toplama, toplanan verileri düzenleme ve analiz ederek yorumlama ve tümevarım istatistiği kapsamında çıkarsama yapma amaçlı kullanılan yöntemler bilimi.
- Veri toplama: Araştırılan konuya yönelik doğru verilerin nasıl toplanacağına yönelik yöntemler topluluğu.
- Tesadüfi ve sistematik hatalar: Verilerin toplanması sırasında yapılabilecek temel hata türleri.

1.1. İstatistiğin Tanımı ve Konusu

İstatistik kelimesinin Modern Latince'deki *statisticum collegium* (devlet konseyi) ve İtalyancadaki *statista* (devlet adamı, politikacı) kelimelerinden türediği kabul edilmektedir. Kelime ilk olarak Almancada Gottfried Achenwall tarafından devlete ait verilerin sunulduğu *Statistik* (1749) adlı eserde devlet bilimi anlamında kullanılmıştır. (Çevrimiçi: <https://tr.wikipedia.org/wiki/%C4%B0statistik>)

Devlet bilimi denmesinin sebebi, devletlerin başta nüfus olmak üzere, nüfusun bileşimi ve özellikleri açısından bilgi sahibi olma amacıyla veri toplamalarıdır. Dolayısıyla, tarihsel anlamda istatistiğin ilk ortaya çıkışı veri toplama biçiminde olmuştur.

Veri toplama işlevi ile başlayan istatistik faaliyetleri; toplanan verileri düzenleme, özetleme, grafiklerle sunma gibi başlangıçta basit katkılarla, daha sonra geliştirilen teori ve yöntemlerle bugünkü kapsamına ulaşmıştır.

Bugün itibarıyla, istatistiğin ne olduğuna yönelik istatistikçiler arasında kabul görmüş tek bir tanımdan söz edemiyoruz. Bunun yerine, istatistiğin işlevsel tarafı ele alınarak başlıca iki temel işlevi yerine getirdiği kabul edilmektedir. Bunlar:

- Tarifsel / Tanımlayıcı (Betimsel) İstatistik
- Tümevarım (Çıkarsama) İstatistiği, Çıkarımsal İstatistik

şeklinde ifade edilmektedir.

1.1.1. Tarifsel (Tanımlayıcı / Betimsel) İstatistik

İstatistiğin ilk aşaması ilgi duyulan ya da araştırılan konuya yönelik olarak bilgi toplamaktır. Basitçe veri toplama olarak nitelenen bu aşama ile bir yandan toplanan veriler tablo ve grafiklerle düzenlenmekte ve sunulmakta, diğer yandan verinin özelliklerini tarif etmeye yönelik tanımlayıcı / özetleyici bazı bilgiler yaratılmaktadır.

Söz gelimi bir insan kaynakları yöneticisinin, personel alımı için başvuran adaylara bir test uyguladığını ve bu test sonuçlarına göre tercihte bulunacağını varsayalım.

Test, adaylara uygulandıktan ve değerlendirildikten sonra, elde edilen verilerle ilgili olarak yönetici şunları yapabilir:

İlk olarak test sonuçlarını en yüksek ya da en düşük puandan başlayarak sıralayabilir, aynı puanı alanları tasnif ederek bir araya toplayabilir, yakın puanları sınıflayabilir, ortalama puanı hesaplayabilir, puanların ne derece dağıldığını belirleyebilir, tablolar ve grafikler oluşturarak verileri düzenleyebilir.

Basitçe sıraladığımız verinin toplanmasından, toplanan verilerden hareketle elde edilen ölçülere, tablo ve grafik oluşturma sürecine **tarifsel (tanımlayıcı) istatistik ya da betimsel istatistik** adı verilmektedir.

1.1.2. Tümevarım (Çıkarsama) İstatistiği

Bir genel seçim öncesinde yapılan kamuoyu araştırmalarını ya da rating ölçümlerini hepimiz duymuşuzdur.

Halk arasında seçim araştırması olarak nitelenen bu araştırmalar ile ülke sınırları içinde oy verme hakkına sahip milyonlarca seçmenin seçim günü nasıl hareket edeceği ve hangi partinin ne oranda oy alabileceği, sözü edilen milyonlarca seçmen arasından seçilen çok az sayıda seçmen ile yapılan anket çalışmalarından hareketle tahmin edilebilmektedir.

Benzer şekilde, milyonlarca televizyon izleyicisi arasından, bu izleyicileri temsil etmek üzere seçilen az sayıda izleyici izlenerek, hangi TV kanalının ne ölçüde izlendiğine yönelik tahminler yapılmaktadır.

İşte, büyük bir kitleye yönelik oy oranı, izlenme oranı vb. tahminler **tümevarım istatistiği, çıkarsama istatistiği** ya da **çıkarımsal istatistik** olarak adlandırılmaktadır.

Bugünkü anlamıyla istatistik, betimsel istatistik olarak nitelenen veri toplama, toplanan verileri düzenleme, analiz etme ile ilgili olarak geliştirilmiş teori ve yöntemlerin yanı sıra tümevarım istatistiği alanıyla ilgili olarak geliştirilmiş teori ve yöntemleri içeren bir bilim dalı olarak karşımıza çıkmaktadır.

İstatistik; ekonomiye uygulanması ile ekonometri, psikolojiye uygulanması ile psikometri, sosyolojiye uygulanması ile sosyometri, sağlık alanında uygulanması ile biyoistatistik, eğitim bilimleri alanında da ölçme ve değerlendirme ile çok yaygın bir kullanım alanına sahip olmuştur.

İstatistiğin konusunu kollektif olaylar oluşturur. Kollektif olayın ne olduğunu anlamak için öncelikle tipik olaydan bahsetmek istiyoruz. Tipik olay, aynı koşullar altında her zaman aynı sonucu veren olaylardır. Örneğin; bir kalemin, belirli bir masadan yere düşüş süresi, kalemin ağırlığı ve masanın yüksekliği değişmediği sürece, değişmeyecektir. Böyle bir deneyi kaç kez tekrar edersek edelim hep aynı sonucu alırız. Benzer şekilde, belirli sıcaklıkta ve miktarda su, aynı şiddette ısıya maruz bırakıldığında hep aynı sürede kaynayacaktır. Kimya alanından örnek vermek istersek, 2 hidrojen ve 1 oksijen atomu daima 1 molekül su oluşturur.

Verdiğimiz örneklerden de anlaşılacağı gibi, fiziksel ve kimyasal hadiseler; koşullar değişmediği sürece aynı sonuçları doğurur. Burada herhangi bir farklılık ya da değişim oluşmaz. Dolayısıyla, deneyin sonucu da önceden kesin olarak bilinebilir. Öte yandan, bu tür hadiselerde, sonuç değişmediğinden bir tek gözlem yapmış olmak yeterlidir.

İşte, aynı koşullar altında her zaman aynı sonuçları veren hadiseler, **tipik olay** olarak adlandırılır.

Kollektif olay ise birbirine benzemeyen, bazı ortak özelliklere sahip olursa bile yine de farklılıklar gösteren hadiselerdir.

Bir mahallede yaşayan insanları düşünelim. İnsan olmak; bu mahallede yaşayanların ortak özelliğidir. O halde, bu insanlar birbirinin aynısı mıdır? Kadın ve erkek şeklinde iki cinse ayrılmaları yanında; yaşlarına, eğitim ve gelir seviyelerine, mesleklerine ve daha pek çok özelliğe göre farklılıklar göstermeleri doğaldır. O halde, bu mahalleden seçeceğimiz bir tek kişinin bütün mahalleyi temsil edeceğini söyleyemeyiz.

Ayrıca, mümkün olduğunca aynı koşullar sağlansa bile kolektif olaylar farklı sonuçlar verecektir. Örneğin, aynı eğitim ve kültür seviyesinde, aynı gelire sahip iki insan farklı harcama ya da tasarruf düzeylerine sahiptir. Benzer şekilde; aynı anne babadan dünyaya gelmiş, aynı eğitimi almış ve aynı çevrede yetişmiş ikiz kardeşlere bir test uygulansa farklı sonuçlar alabileceklerdir. Aynı tohum kullanılarak yetiştirilen iki bitkinin gelişimi de tıpatıp aynı olmayacaktır. Görüldüğü gibi, her ne kadar aynılık kriteri sağlanmaya çalışılsa da farklı neticeler ile karşılaşılabilir.

Özetle, kolektif olay olarak tanımlanan bu tür olaylarda, ortak özellikler bulunsa bile farklı sonuçlar alınmakta ve bu sebeple tek bir hadisenin gözlenmesi o konuya yönelik yeterli bilgi vermemektedir. Bu sebeple genel bir kanı oluşturacak düzeyde gözlem yapmak ya da veri toplamak suretiyle analizler gerçekleştirmek, benzerlik ya da farklılıkları ortaya çıkarmak amaçlanmaktadır.

İstatistik metodolojisi başlıca dört temel aşamadan oluşmaktadır. İlk olarak, gözlem yapmak ya da veri toplamak gerekir. İkinci olarak ham veri niteliği taşıyan, düzensiz durumda olan toplanan verilerin istatistik analizler için uygun hale getirilmesi, yani düzenlenmesi gerekecektir. Üçüncü aşamada, düzenlenen veriler, çeşitli tablo ve grafiklerle gösterilecek ya da sunulacak, ve son aşamada da analize hazır halde olan veriler uygun istatistik tekniklerle analiz edilecek ve değerlendirilecektir.

1.2. Verilerin Toplanması ve Rölöve (Derleme) Kavramı

İstatistik; kolektif olayların incelenmesine, gözlenmesine ve analiz edilmesine yönelik olarak kullanılan yöntemler topluluğudur. Dolayısıyla, istatistik çalışmalarda ilk aşama gözlem yapma ya da veri toplama aşamasıdır. Başka bir deyişle, incelenen olaya yönelik sayım ya da ölçüm yoluyla veri toplamak istatistik çalışmanın ilk aşamasını oluşturmaktadır. Bu aşamaya, ilk bilgilerin veya verilerin toplanması anlamına gelen **rölöve** veya **derleme** denir.

İstatistik çalışmalarda araştırmacılar başlıca iki tür veri toplama yöntemini kullanırlar. Bunlara “veri kaynakları” da denilmektedir.

İlk olarak, araştırmacılar bir kurum ya da kuruluş tarafından toplanmış veya düzenlenmiş verileri kullanabilirler. Söz gelimi, nüfus, dış ticaret, enflasyon, milli gelir vb. veriler Türkiye İstatistik Kurumu tarafından toplanmakta, düzenlenmekte ve yayınlanmaktadır. Burada veri bizzat araştırmacı tarafından toplanmadığından bu verilere **ikincil veri** denmektedir.

Veri toplama işlemi yukarıda olduğu gibi, resmi bir kurum tarafından yapılabileceği gibi, bir başka kurum ve hatta kişi tarafından da yapılabilir. Örneğin, sigorta sektöründe faaliyet

gösteren bir firma ile ilgili olarak bir takım analizler yapmayı amaçlıyorsak, ilgili sigorta şirketinden veri temin etme yoluna gideriz. Benzer şekilde, belirli bir hastalık ile ilgili bir çalışma yapıyorsak kamu ya da özel bir hastanenin bu konudaki verilerini kullanabiliriz.

Bazı durumlarda araştırmacılar, araştırma yaptıkları konu ile ilgili veriyi kendileri toplar. Bu tür veri toplama işlemine ise **birincil veri** toplama yöntemi denmektedir. Örneğin, bir üniversitede eğitim görmekte olan öğrencilerin yemekhane ve kantin memnuniyetini belirlemeyi amaçlıyorsak, veriyi bizzat toplamamız gerekir. Bu durumda, ilgili üniversitede eğitim görmekte olan tüm öğrencileri temsil etmek üzere yeterli sayıda öğrenci ile görüşerek veri toplanması yoluna gidilir.

Rölöve kavramı, araştırmacının veri toplama işini bizzat gerçekleştirdiği durumu ifade eden bir kavram olarak kullanılmaktadır. Bu açıdan bakıldığında; rölöve kavramından, veri toplama işini gerçekleştiren kişi ya da kurum farkı gözetmeksizin, bir veri yaratma süreci anlaşılmalıdır. Örneğin, Türkiye İstatistik Kurumu milli gelir ya da dış ticaret için veri toplar ve yayınlar. Burada, daha önce mevcut olmayan bir veri yaratma süreci ve bu sebeple birincil veri kaynağı söz konusudur. Araştırmacıların Türkiye İstatistik Kurumu tarafından toplanarak yayınlanan milli gelir ya da dış ticaret verilerini bu kurumdan temin etmesi ise ikincil veri kaynağı olarak nitelendirilmektedir. Dolayısıyla, ikincil veri niteliği taşıyan veriler de, verinin ilk yaratılma süreci açısından birincil veri niteliği taşımaktadır.

Birincil veri toplama yöntemlerini;

1. Deney tasarımı yoluyla veri toplama
2. Anket, soru formu yöntemiyle veri toplama
3. Gözlem yoluyla veri toplama

şeklinde sıralayabiliriz.

Deney tasarımı yoluyla veri toplama yöntemi sağlık alanında yapılan çalışmalarda çokça kullanılan bir yöntemdir. Örneğin yeni geliştirilmiş bir ilacın, belirli bir hastalığı tedavi etmede etkili olup olmadığı tespit edilmek istendiğinde, bu hastalığa sahip hastalar (denekler) kontrol ve deney grubu olarak ikiye ayrılarak, bir gruba bilinen ilaç diğerine ise yeni geliştirilen ilaç verilerek tedavi sonunda arada bir fark olup oluşmadığı istatistik testler kullanılarak test edilir.

Benzer şekilde, yeni bir cerrahi yöntemin operasyon sonrası iyileşme sürecini kısalttığı ya da karmaşa riskini azalttığı iddia ediliyorsa cerrahi müdahale geçirecek olan hastalar kontrol ve deney grubu olarak iki gruba ayrılarak bir gruba bilinen cerrahi yöntem, diğerine yeni geliştirilen cerrahi yöntem uygulanarak ameliyat sonrası iyileşme süreci izlenir ve istatistik testler kullanılarak yeni geliştirilen yöntemin daha etkili bir yöntem olup olmadığı tespit edilir.

Anket ya da soru formu yoluyla veri toplama yönteminin en bilinen örneği seçim araştırmalarıdır. Seçim öncesinde seçmenlerin siyasi eğilimleri, bütün seçmen kitlesini temsil etmek üzere seçilen bir örnek üzerinden yapılır. Bu örnek grup içinde yer alan seçmenlere seçim günü kime oy verecekleri sorulur. Böylelikle elde edilen cevaplar analiz edilerek partilerin oy oranları tahmin edilir.

Benzer şekilde, bir sađlık kurumundan sađlık hizmeti alan hastaların memnuniyet dereceleri ölçölmek istendiđinde, hastalara isteđe bađlı olarak, bir anket formu doldurtulur ve verdikleri cevaplar üzerinden hizmet kalitesi ya da hizmet memnuniyeti belirlenmeye çalışılır.

Anket yoluyla veri toplama işlemi yüz yüze görüşme şeklinde yapılabileceđi gibi, internet üzerinden veya telefon görüşmesi şeklinde de yapılabilir.

Gözlem yoluyla veri toplama yönteminde ise, araştırmacı incelediđi konuya yönelik olarak birimleri dođal ortamlarında gözler ve davranışlarını kaydeder. Örneđin, bir müzeyi ziyaret etmekte olan ziyaretçilerin, ziyaretleri süresince davranışları izlenerek, kaydedilir. Ziyaretçiler izlendiđini bilmemekte ve dođal bir ziyaret eylemi içinde müzeyi gezmektedir. Hangi objeler ile ilgilenildiđi, en çok ilgi çeken objeler, gezi güzergâhı, sorunlar vb. hepsi araştırmacı tarafından izlenerek kaydedilir. Daha sonra, çeşitli analizler vasıtasıyla genel bir deđerlendirme yapılmış olur.

Verilerin toplanması başlıđı altında son olarak veri toplama aşamasının önemini vurgulamak istiyoruz. Şöyle ki, veri toplama aşaması istatistik çalışmaların ilk aşamasını oluşturmakta ve bu aşamada toplanan veriler üzerinden tüm deđerlendirme ve analizler yapılmakta ve bir takım sonuçlara ulaşılmakta ya da kararlar verilmektedir. Dolayısıyla, alınan kararların veya elde edilen sonuçların dođruluđu toplanan verinin dođruluđuına bađlıdır. Ancak dođru veri ile dođru sonuçlara ulaşılabilir.

1.3. İstatistik Metodolojisine Yönelik Temel Kavramlar

1.3.1. Anakötle, Örnek Kötle, Parametre ve İstatistik

İstatistiđin konusunun kolektif olaylar olduđunu belirtmiştik. Şimdi bazı temel kavramlar üzerinde durmak istiyoruz.

Veri toplama işlemi incelemeye konu olan tüm birimlerin incelenip incelenmediđine göre ikiye ayrılır:

- Bütüne yönelik rölöve, tam sayım
- Kısmi rölöve, örnekleme

Araştırmaya konu olan kolektif olaya ilişkin tüm birimlerin oluşturduđu topluluđa **anakötle** denir. Anakötlede yer alan tüm birimlerin incelenmesi işlemine de **tam sayım** ya da **bütüne yönelik rölöve** adı verilir. Burada, kolektif olaya ilişkin hiçbir birim dışarıda tutulmaksızın bir araştırma gerçekleştirilmekte yani tüm birimler araştırmaya dâhil edilmektedir.

Türkiye'nin nüfusunu tespit etmeye yönelik çalışmalar anakötle düzeyinde gerçekleştirilir. Yani amaç, Türkiye nüfusunu oluşturan herkesin yaş, cinsiyet, eğitim, medeni durum vb. özellikleri açısından deđerlendirilmesidir.

Ancak, bazı hallerde tüm birimlerin gözlenmesi ya da incelenmesi çok maliyetli olmakta veya çok zaman almaktadır. Söz gelimi, bir siyasi partinin seçim öncesi seçmenlerin siyasi eğilimlerini tespit etmek ve partisinin oy oranını tahmin etmek istediğini düşünelim. Bu durumda, tüm seçmenlere ulaşarak hangi partiye oy vereceklerini sormak gerekecektir. Milyonlarca seçmen olduğu düşünüldüğünde böyle bir işlem çok sayıda anketör kullanımı gerektirecek ve oldukça maliyetli olacaktır. Öte yandan, milyonlarca seçmen ile görüşmek çok zaman alacaktır. Uzun zaman gerektirecek böyle bir zaman aralığı içinde, ülke gündemi değişebilecek ve buna bağlı olarak seçmenlerin tercihleri farklılaşacak, ölümler ve yeni seçmen niteliği kazanan insanlar olması sonucunda seçmen sayısı ve bileşimi değişecektir. Dolayısıyla, belirli bir an itibarıyla yapılması gereken seçim araştırmasının uzaması neticesinde A partisine oy vereceğini söyleyen bir seçmenin zaman içinde fikir değiştirerek B partisine kayması, C partisine oy verecek bir seçmenin hayatını kaybetmesi, Seçmen niteliği taşımayan birinin seçmen niteliği taşıyor hale gelmesine rağmen araştırma dışında kalması vb. değişimler olabilecektir.

Hattâ bazı durumlarda anakütlenin tamamını gözlemek mümkün dahi olamamaktadır. Sözgelimi bal arılarıyla ilgili bir çalışma yapmayı planlayalım. Tabiatla bulunan bütün arıları gözlemek, incelemek mümkün olamayacaktır. Zira, bütün arıları, hiçbirini dışlamayacak şekilde, bir araya toplamak ve incelemek mümkün değildir.

Zaman, maliyet ve fiili imkânsızlık durumlarında anakütlenin tamamını gözlemek yerine, anakütleyi temsil edecek nitelik ve yeterlilikte birimi gözlemek yoluna gidilmekte ve az sayıda birimden oluşan ve **örnek kütle** ya da **örneklem** adı verilen kütlede yer alan birimler gözlenmektedir.

Bu durumda, örneklem ya da örnek kütle üzerinden veri toplama işlemi gerçekleştirilir ki buna **kısmi rölöve** denmektedir. Kısmi rölöve ile toplanan veriler üzerinden yapılan analizler sonucunda anakütleye ilişkin çıkarımlarda bulunulur. Bu durumu bir örnek ile açıklamaya çalışalım. Bir ülkede 50 milyon seçmen olduğunu varsayalım. Seçimden birkaç hafta önce, seçim sonuçlarını tahmin etmek isteyelim. 50 milyon seçmenden oluşan anakütle içinden, bu anakütleyi temsil edebilecek bir örnek kütle seçmemiz gerekecektir. Burada önemli olan husus, seçilen örneğin anakütleyi yeterince temsil edebilecek bileşimde ve sayıda seçilmesidir. Bu seçimin nasıl yapılacağına yönelik çok sayıda örneklem yöntemi vardır. 50 milyon seçmeni temsil edecek 3000 kişi seçtiğimizi varsayalım. Örnek kütlelerimizde yer alan bu 3000 kişiye bugün seçim olsa hangi partiye oy vereceği sorularak, partilerin oy dağılımı belirlenir.

Hesaplanan oy dağılımı 3000 kişi üzerinden belirlenen oy dağılımı olmakla beraber, 50 milyon seçmen için ortaya çıkacak oy dağılımının tahmini olma özelliği taşır. Dolayısıyla, amaç 50 milyon seçmen için oy dağılımını tahmin etmektir ve bu tahmin 3000 kişi üzerinden yapılmaktadır.

Konunun daha iyi anlaşılması için örneklem çalışmalarına yönelik basit bir örnek vermek istiyoruz. Bir tencerede dört kişilik bir çorba yaptığımızı varsayalım. Çorbanın içinde çeşitli malzemeler, tuz, baharat vb. bulunacaktır. Çorbayı hazırlayıp pişirdikten sonra, tadı, tuzu ve kıvamının yerinde olup olmadığını anlamak üzere hepsini içmemiz gerekmez. Tencerenin

içinden aldığımız tek bir kaşık bize bilgi vermesi açısından yeterlidir. Dolayısıyla, burada bir örnekleme çalışması söz konusudur.

Örneğimizi biraz geliştirelim. Şimdi de, dört kişilik değil, bir öğrenci yemekhanesini için 1000 kişilik bir çorba yaptığımızı varsayalım. Yaptığımız çorbanın miktarı artmıştır. Bu durumda, çorbanın tadı, tuzu ve kıvamını anlamak üzere bir kaşık değil bir kâse mi çorba içmek gerekecektir. Tabii ki hayır.. Yine bir kaşık çorba ile bu tahmini yapabilir ve çorbanın kıvamı, tadı, tuzu hakkında bilgi sahibi olabiliriz. Bir kaşık çorba, tenceredeki çorba miktarı artmış dahi olsa, yine yeterli bir tahmin aracı olacaktır. Hangi koşulda? Büyük bir tencerenin içindeki çorbanın çok iyi karıştırılmış olması durumunda.

Tencerenin içindeki çorbanın çok iyi karıştırılmış olması sağlandığında, tencereden alınan tek bir kaşık, tencerenin içindeki çorbanın tamamını temsil edebilecek bileşimdedir. Karıştırma işlemiyle, çorbadan aldığımız o bir kaşığın içine, çorbada yer alan bütün malzemelerden biraz girmiş olmasını sağlamış oluruz. Dolayısıyla, başarılı bir örnekleme çalışması, çok fazla sayıda örnek çekmekle değil, anakütleyi oluşturan farklı nitelikteki birimlere örnekte yer vermek suretiyle gerçekleştirilebilir.

Örnek seçme işleminin nasıl yapılacağına yönelik, anakütlenin yapısını, birim sayısını ve araştırma yapılan konuyu temel alan çok sayıda örnekleme yöntemi mevcuttur. Bu tekniklere, örnekleme başlığı altında değineceğiz.

Bu başlık altında son olarak parametre ve istatistik kavramlarına değinmek istiyoruz. Anakütleyi karakterize eden değerlere **parametre**, örnek kütle ya da örnekleme karakterize eden değerlere ise **istatistik** ya da **statistik** denir.

Seçim araştırması örneğimizi hatırlayalım. Ülkede 50 milyon seçmen olduğunu ve 3000 seçmen ile görüşerek hangi partiye oy vereceğini sorduğumuzu varsaymıştık. 3000 kişiye sormak suretiyle elde ettiğimiz oy oranları, örnek kütle için hesaplanmıştır ve istatistik adını almaktadır. 50 milyon seçmen için hesaplanan oy oranları ise parametredir.

Dikkat edilecek olursa, örnek istatistiği seçilen örneğe göre farklılık, değişkenlik gösterebilirken parametre o anakütle için tek ve kesin değerdir. Şöyle ki, seçtiğimiz 3000 kişi değişirse, örnek istatistiğimiz de değişir. Farklı örnek kütleler, farklı oy dağılımları ve dolayısıyla farklı örnek istatistikleri yaratır. Oysa, anakütle için hesaplanacak oy dağılımı yani parametre değerleri tek ve kesin değer olmaktadır.

Daha önce istatistiğin tanımı ve konusu başlığı altında tümevarım ya da çıkarsama istatistiği olarak adlandırdığımız çalışma alanı yukarıda bahsettiğimiz örnek istatistiğinden hareketle anakütle parametresini tahmin etme alanıdır.

Bu başlık altında son olarak veri toplama aşamasında yapılan hatalara yani rölöve hatalarına değinmek istiyoruz.

1.3.2. Rölöve Hataları

Çok çeşitli sebeplerle ortaya çıkabilecek rölöve hatalarını çalışmanın bütününe yönelik etkileri açısından tesadüfi hatalar ve sistematik hatalar olmak üzere iki başlıkta inceliyoruz:

Tesadüfi hatalar, farklı yönlerde ve genellikle zıt yönlerde ortaya çıkan hatalar olup, genel olarak değerlendirildiğinde çalışmanın bütününe yönelik etkisi düşük düzeyde olan hatalardır. Örneğin, sayım görevlilerinin bir kadını erkek ya da bir erkeği kadın olarak kodlaması. Bu tür bir hata hep aynı yönde gerçekleşmez. Yani, devamlı olarak erkekler kadın ya da kadınlar erkek olarak kodlanmaz. Dolayısıyla, her iki yönde de yapılabilecek bu tip bir hatanın çalışma üzerinde etkisi, zıt yönde gerçekleşen hatalar birbirini yok edeceği için, çok düşük olacaktır.

Sistematik hatalar ise hep aynı yönde tezahür eden, gerçekleşen hatalardır. Bu tür hatalara örnek olarak, genç görünmek arzusuyla yaşın olduğundan küçük söylenmesi ya da vergi korkusu ile gelirin olduğundan düşük beyan edilmesi verilebilir.

Sistematik hataları tesadüfi hatalardan ayıran en temel özellik, hataların hep aynı yönde gerçekleşmesi ve birim sayısını arttırmak suretiyle bu tür hataların azaltılamamasıdır. Oysa, birim sayısı arttıkça tesadüfi hatalar hem oransal olarak azalmakta hem de artı ve eksi yönde gerçekleştikleri için birbirlerinin etkisini nötr hale getirmektedirler.

Sonuç olarak, tesadüfi hatalar çalışmanın sonuçları üzerinde etkili bir hata olmamakta ancak sistematik hatalar ise dikkate alınması gereken ve çalışmanın sonuçlarını doğrudan etkileyen hatalar olarak karşımıza çıkmaktadır.

Bu Bölümde Neler Öğrendik?

Bu bölümde, istatistiğin tanımı ve konusu üzerinde durduk. Bu bağlamda önce, veri toplama, toplanan verileri düzenleme, özetleme ve sunmayı amaçlayan istatistik alanı olan betimsel istatistikten söz ettik. Daha sonra, tümevarım istatistiği alanına değinerek bu alanla ilgili temel kavramları öğrendik.

İstatistik disiplininde kullanılan temel kavramlardan olan, anakütle, örnek kütle, parametre, istatistik kavramlarını ayrıntılı olarak ele alarak, rölöve hataları üzerinde durduk.

Bölüm Soruları

- 1- İstatistikte veri toplama işleminedenir.
 - a- Rölöve
 - b- Veri
 - c- Vasıf
 - d- Şık
 - e- Birim
- 2- Araştırmacı tarafından doğrudan yapılan veri toplama işleminedenir.
 - a- Veri toplama
 - b- Rölöve
 - c- Birincil veri
 - d- İkincil veri
 - e- Soru formu
- 3- Aynı koşullar altında her zaman aynı sonuçları veren olaylaradenir.
 - a- Atipik olay
 - b- Tipik olay
 - c- Kollektif olay
 - d- Gözlem
 - e- Veri
- 4- İnceleme ya da gözleme konu olan tüm birimlerin oluşturduğu kütle.....denir.
 - a- Örnek
 - b- Örnek kütle
 - c- Anakütle
 - d- Tam sayım
 - e- Gözlem

- 5- Anakütleyi karakterize eden, niteleyen değerlere.....denir.
- a- İstatistik
 - b- Tam sayım
 - c- Örnekleme
 - d- Parametre
 - e- Rölöve
- 6- Örnek kütleyi karakterize eden, niteleyen değerleredenir.
- a- İstatistik (Statistik)
 - b- Tipik olay
 - c- Vasıf
 - d- Şık
 - e- Parametre
- 7- Araştırmacının bir kurum ya da kuruluş tarafından toplanmış verileri kullanması durumunda..... söz konusu olmaktadır.
- a- Birincil veri
 - b- Veri
 - c- İkincil veri
 - d- Anket
 - e- Gözlem
- 8- Bir araştırmada konu ile ilgili tüm birimlerin gözlenmesi durumuna denir.
- a- Tam sayım
 - b- Örnekleme
 - c- Anakütle
 - d- Parametre
 - e- İstatistik

- 9- Zaman, maliyet vb. sebeplerle, incelemeye konu olan tüm birimlerin değil, bu birimlerin içinden seçilen az sayıda birimi gözlemek suretiyle anakütleye yönelik bilgi sahibi olma yöntemine..... denir.
- a- Anakütle
 - b- Rölöve hatası
 - c- Veri toplama
 - d- Örneklem
 - e- Parametre
- 10- Veri toplama işlemi sırasında ortaya çıkan ve hep aynı yönde hataya sebep olan hata türü aşağıdakilerden hangisidir?
- a- Rölöve hatası
 - b- Hata
 - c- Tesadüfi hata
 - d- Önemli hata
 - e- Sistematik hata

Cevaplar

- 1) a
- 2) c
- 3) b
- 4) c
- 5) d
- 6) a
- 7) c
- 8) a
- 9) d
- 10) e

Yararlanılan Kaynaklar

ALOBİ, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), "On the Theory of Scales of Measurement" , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

2. BİRİM, VASİF, ŞİK KAVRAMLARI VE VERİNİN ÖLÇÜM DÜZEYİ

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜK AKMAN

DR.ÖĞR.ÜYESİ HAKAN BEKTAŞ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde istatistiğin konusunu teşkil eden “birim” kavramı üzerinde durulacak ve birimin özellikleri ve çeşitleri incelenecektir.

Birimin özellikleri ile ilgili olarak vasıf ve şık kavramları ele alınacak ve örneklerle açıklanacaktır.

Veri toplama aşamasında birimin ölçümüne yönelik ölçüm düzeyleri konusu incelenecek ve karşılaştırmalı olarak değerlendirilecektir.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) İstatistiğin konusu nedir?
- 2) İstatistikte “Birim” nedir?
- 3) Vasıf ve şık kavramları ne demektir?
- 4) Verinin ölçüm düzeyi ne demektir?
- 5) Ölçüm düzeyi açısından verilerin sınıflanması nasıldır?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistik nedir? İstatistiğin konusu nelerdir?	Gerek fen bilimleri gerekse sosyal bilimler alanında kullanılan istatistik nedir ve ne işe yarar konularında fikir sahibi olur.	Temel düzeyde istatistik kitapları ve yayınlanan istatistikler.
Birim kavramı	İstatistikte birim kavramını ve çeşitlerini bilir.	Temel düzeyde istatistik kitaplarında istatistiğin konusu ve temel kavramlarla ilgili bölümleri.
Vasıf ve şık kavramı	İstatistikte birimin özellikleri ile vasıf ve şık kavramlarını bilir.	Temel düzeyde istatistik kitaplarında istatistiğin konusu ve temel kavramlarla ilgili bölümleri.
Ölçüm düzeyi kavramı	Veri toplama ile ilgili olarak verinin ölçüm düzeyi ve önemi konularında bilgi sahibi olur.	Temel düzeyde istatistik kitaplarının veri, ölçek, ölçüm düzeyi ile ilgili bölümleri.
Verilerin ölçüm düzeylerine göre sınıflanması	Ölçüm düzeylerinin çeşitleri ve istatistik çalışmalardaki önemini bilir.	Temel düzeyde istatistik kitaplarının veri, ölçek, ölçüm düzeyi ile ilgili bölümleri.

Anahtar Kavramlar

- Birim: İnceleme ya da gözleme konu olan kollektif olaylardan her biri.
- Vasıf: Birimin sahip olduğu özellikler.
- Şık: Vastın birimde ortaya çıkan görüntüsü.
- Ölçüm düzeyi: Birimin ilgilenilen vasıf açısından nasıl ölçüldüğü.

2.1. İstatistikte Birim Kavramı ve Çeşitleri

Bir istatistik çalışmada, inceleme ya da gözleme konu olan kollektif olaylardan her birine **birim** denir. Birimin mahiyeti incelenen olaya göre değişir. Aşağıda, yapılan istatistik çalışmaya göre birimin ne olduğuna yönelik örnekler verilmiştir:

Tablo 2.1: İstatistikte Çalışmanın Niteliğine Göre Birimler

İstatistik çalışma	Birim
Nüfus sayımı	İnsan
Hane halkı tüketim araştırması	Aile
Ziraat sayımı	Zirai işletme
Trafik kazası	Her bir trafik kazası
Boşanma istatistikleri	Her bir boşanma olayı
Doğum istatistikleri	Her bir doğum

İstatistikte birim, maddi bir varlığa sahip olabileceği gibi maddi varlığa sahip olmayabilir de. Örneğin nüfus sayımında birim insandır ve fiziksel ya da maddi bir varlık söz konusudur. Benzer şekilde bina sayımında birim binadır ve yine gözle görülür maddi bir varlığa sahiptir. Bu tür fiziksel varlığa sahip olan birimlere **maddi birim** denir. Öte yandan, birimin mutlaka fiziki bir varlığa sahip olması gerekmez. Sözgelimi doğum istatistiklerinde soyut anlamda doğum vakası bir birimdir. Benzer şekilde boşanma istatistiklerinde de her bir boşanma vakası bir birim olarak değerlendirilmektedir. Fiziksel varlığı olmayan birimlere ise **maddi olmayan birim** adı verilir.

Birim tek bir varlık olabileceği gibi bir gruptan da oluşabilir. Örneğin nüfus istatistikleri açısından yaş farkı gözetilmeksizin her insan bir birimken, hane halkı tüketim araştırmalarında aile birim özelliği taşımaktadır.

Birimin sayılabilir ya da ölçülebilir olması gerekir. Nüfus, evlenme, boşanma istatistiklerinde sayma esastır. Tarım istatistikleri kapsamında hesaplanan buğday üretimi ise ölçüm ile gerçekleştirilmektedir.

Birim, uzun ömürlü olabileceği gibi, kısa sürede ortadan kaybolan kısa ömürlü de olabilir. Nüfus sayımlarında birim özelliği taşıyan insan uzun ömürlü bir birimdir. Bugün var olmaktadır ve yarın da var olacaktır. Benzer şekilde, motorlu taşıt istatistiklerinde her bir motorlu taşıt bir birimdir ve uzun ömürlüdür. Bugün de gözlenebilir yarın da gözlenebilir. Bu tür göreceli olarak uzun ömürlü olarak niteleyebileceğimiz birimlere **devamlı birim** diyoruz. Bu tür birimler belirli bir zaman dilimi içinde gözlenebilme imkânına sahiptir.

Oysa bazı birimler, kısa ömürlüdür ve ortaya çıktıkları anda gözlenmeleri ya da kaydedilmeleri gerekir. Örneğin, trafik kazaları istatistiklerinde her bir trafik kazası bir birimdir ve ortaya çıktıkları anda gözlenmeleri gerekir. Aksi takdirde, ortadan kalkmakta ve gözlenebilir olma özelliklerini yitirmekte ve dolayısıyla kayıt edilmeleri mümkün olmamaktadır. Boşanma istatistikleri de böyledir. Her bir boşanma olayının birim olarak değerlendirildiği bu istatistiklerde boşanma olayı gerçekleştiği anda kayıt edilmelidir. Kayıt edilmemesi halinde daha sonra tespit edilerek gözlenmesi çok güçtür.

İşte, gerçekleştiği anda kayıt altına alınması gereken , kısa ömürlü bu birimlere **ani birim** adı verilmektedir.

Birimin homojen olması ve belirli bir tarife ya da tanıma uygunluk göstermesi gerekir. Örneğin, İstanbul'a göç etmiş aileler üzerinde bir sosyolojik araştırma yaptığımızı varsayalım. Burada birim, İstanbul'a göç etmiş ailelerdir. Çalışmamız Türkiye'nin çeşitli yerlerinden göç etmiş aileler ile ilgiliyse, yurt dışından göç etmiş aileler çalışmanın birimi niteliği taşımaz. Ya da benzer şekilde üniversite eğitimi için İstanbul'da bulunan öğrenciler de bu çalışmanın birimi niteliği taşımaz.

Benzer şekilde bina sayımı yaptığımızı varsayalım. Sayımı gerçekleştirecek kişilerin herhangi bir eğitime tabi tutulmaksızın sayıma gönderilmesi bazı sakıncaları beraberinde getirecektir. Örneğin; görevlilerden biri, boş ve içinde kimsenin yaşamadığı bir binayı sayarken diğeri saymayabilecektir. Benzer şekilde, bir binanın yanında boş bir arsaya yapılmış, kulübe benzeri bir bina bazı sayım görevlilerince sayılacak bazıları tarafından ise sayılmayacaktır. Yine depo ya da araç park yeri olarak kullanılan temelsiz prefabrik görüntüde bir yapı da görevlilerin inisiyatif kullanmasına göre sayılacak ya da sayılmayacaktır.

Görüldüğü gibi, basit bir bina sayımında bile görevlilerin bir yapıyı hangi koşullarda bina olarak kabul edeceğinin, hangi durumlarda ise bina olarak kabul etmemesi gerektiğinin çok net ve anlaşılır şekilde açıklanması gerekmektedir.

2.2. Vasıf ve Şık Kavramı

Birimlerin sahip oldukları özelliklere **vasıf** denilmektedir. Örneğin, nüfus sayımlarında birimin insan olduğunu daha önce ifade etmiştik. İnsanı niteleyen özellikler olan; cinsiyet, yaş, medeni durum, eğitim, meslek vb. özellikler birer vasıftır.

Benzer şekilde, bina sayımı yapıyorsa, birim binalardır ve binanın yaşı, kat sayısı, şekli, kat mülkiyetli olup olmadığı gibi özellikleri ise vasıflarıdır.

Vasıfların ortaya çıkış şekillerine ya da diğer deyişle bireylerdeki görünümüne ise **şık** denilmektedir.

Cinsiyetin insan açısından bir vasıf olduğunu söylemiştik. Cinsiyet, kadın ve erkek şeklinde iki şıklı bir vasıftır. Medeni hâl ise; bekâr, evli ve dul-boşanmış şeklinde üç şıklı bir vasıftır.

Bu başlık altında vasıf türlerine de kısaca değinmek istiyoruz:

Vasıflar, sahip oldukları şık sayılarına göre, **az şıklı vasıflar** ve **çok şıklı vasıflar** olmak üzere ikiye ayrılır. Az şıklı vasıflar cinsiyet, medeni durum gibi şık sayısının az olduğu vasıflardır. Meslek, doğum yeri, yaş, gibi vasıflar ise çok sayıda şıka sahip oldukları için çok şıklı vasıf grubunda yer alırlar.

Vasıflar, nicel ve nitel vasıf olarak da ikiye ayrılırlar, Şıkları sayısal nitelik taşıyan vasıflara **nicel vasıf** denmektedir. Yaş, boy uzunluğu, ağırlık gibi vasıflar nicel vasıflara örnek verilebilir. Şıkları sayısal değer almayan vasıflara ise **nitel vasıf** denmektedir. Cinsiyet, medeni durum, meslek, göz rengi gibi vasıflar ise sayısal olmayan özellik gösterdikleri için nitel vasıflara örnektir.

Vasıflara yönelik bir diğer ayırım, sürekli ve süreksiz (kesikli) vasıf ayırımıdır. Şıkları sadece belli değerleri alabilen vasıflara süreksiz (kesikli) vasıf denmektedir. Bireylerin çocuk sayısı, yaşı, binaların kat sayısı süreksiz vasfa örnek olarak verilebilir. Çocuk sayısı ya birdir ya da ikidir, bir buçuk çocuk sahibi olmak mümkün değildir. Benzer şekilde yaş da ya 21 ya da 22'dir, 21,3 olması söz konusu değildir. Süreksiz (kesikli) vasıflarda birimlerin ara değer alması mümkün değildir, kesikli vasıf denmesinin sebebi de esasen budur.

Şıkları birbirine sonsuz derecede yaklaşan, başka bir deyişle iki değer arasında sonsuz değer alabilen vasıflara ise **sürekli vasıf** denmektedir. Ağırlık ve boy vasıfları sürekli vasfa örnek olarak verilebilir. Hassas ölçüm aletleri ile ölçüm yapıldığı takdirde bir bireyin boyu 171 ile 172 cm arasında örneğin 171,57 cm olabileceği gibi, 171,576 da olabilir. Bu tür vasıflar, görüldüğü gibi, küsuratlı olarak ifade edilebilmektedir.

Vasıflar şıklarının belirli bir sıra izleyip izlememelerine göre sıralı ve sırasız vasıf şeklinde sınıflanmaktadır. Örneğin, yaş, eğitim gibi vasıfların şıkları belirli bir sıra izlemekte olup bu tür vasıflara **sıralı vasıf** denmektedir. Sıralı vasıflar, sıra özelliği gösterdiğinden bir birimin şık değerine bakarak o vasıf açısından pozisyonunu değerlendirebiliyoruz. Sözgelimi, 25 yaşında bir bireyin 18 yaşında bir bireyden büyük, 45 yaşında bir bireyden ise küçük olduğunu anlayabiliyoruz. Benzer şekilde üniversite mezunu birinin lise mezunu birinden daha eğitilmiş olduğunu söyleyebiliyoruz.

Şıkları belirli bir sıra izlemeyen vasıflara ise **sırasız vasıf** denmektedir. Sırasız vasıflara örnek olarak, meslek, doğum yeri, cinsiyet, göz rengi vb. vasıflar verilebilir. Burada bir öncelik-sonralık sıralaması bulunmamaktadır. Dolayısıyla, bu tür vasıflarda şıkları sıralamak mümkün değildir.

Vasıflar şık sayılarına göre de sınıflanmaktadır. Şık sayısı az olan vasıflara **az şıklı vasıf**, şık sayısı çok olan vasıflara ise **çok şıklı vasıf** denmektedir. Sadece iki şıkka sahip olan cinsiyet, üç şıklı olan medeni durum gibi vasıflar az şıklı vasıflara örnek olarak verilebilir. Benzer şekilde, yaş, doğum yeri gibi çok sayıda şıklı olan vasıflar ise çok şıklı vasıflara örnektir.

Vasıf türleri ile ilgili son olarak zaman ve mekân vasıflarına değinmek istiyoruz. Şıkları zaman özelliği gösteren vasıflara **zaman vasfı** adı verilmektedir. Doğum yılı, mezuniyet yılı gibi vasıflar zaman vasfına örnek verilebilir. Öte yandan, mekân özelliği gösteren vasıflara ise **mekân vasfı** denmektedir. Doğum yeri, ikamet edilen il, yaşanan semt vb. vasıflar ise mekân vasfına örnek teşkil etmektedir.

2.3. Verinin Ölçüm Düzeyi

İstatistiksel çalışmalarda araştırmacının incelediği konuya ait veriyi toplayabilmesi, ilgilendiği vasıf açısından birimin ölçümünü gerektirmektedir. Ölçüm, genel olarak birimleri kategorilere atamak şeklinde tanımlanmaktadır. Ancak vasfın şıklarının belirli bir kural dahilinde kurgulanmış olması gerekmektedir. Burada bahsedilen kural, ilgilenilen özellik açısından incelenen birimler için değerlendirme sistematığını belirtmektedir. Bu kuralın tanımına göre, dört ölçüm düzeyi bulunmaktadır. Bunlar; **sınıflama (nominal)**, **sıralama (ordinal)**, **aralık (interval)** ve **oran (ratio)** ölçüm düzeyleridir (Stevens, 1946:678). Veri toplama sonucunda elde edilen verinin ölçüm düzeyi hakkına bilgi sahibi olunması; verinin düzenlenmesi ve uygun istatistiksel yöntemlerin seçilmesi hususunda önem arz etmektedir. Dolayısıyla incelenen birimlerin ölçülmesi ile hedeflenen özellikler ve elde edilecek ölçümlerin ne amaçla kullanılacağı dikkate alınarak uygun ölçüm düzeyinin belirlenmesi gerekmektedir. Bir araştırmacının İstanbul Üniversitesi lisans programlarına kayıtlı öğrencilerin ağırlıklarını düşük, orta ve yüksek kilolu olmak üzere üç kategoride ölçtüğünü varsayalım. Bu araştırmacı, veri toplama aşamasının sonunda İstanbul Üniversitesi lisans programlarına kayıtlı öğrencilerin ağırlıklarının ortalamasını sayısal bir değer (71 kilogram gibi) yerine kategorik (orta kilolu gibi) olarak belirleyecektir. Eğer bu kategorik bilgi, araştırmacının çalışmasının amacına uygun ise vasfın şıklarını (düşük, orta, yüksek) değiştirme gereği duymayacaktır. Ancak araştırmacı incelediği öğrencilerin ağırlıklarının ortalamasını sayısal bir değer olarak elde etmek istiyorsa, bu durumda vasfın şıkları (0, 1, 2, ...) için kural tanımını değiştirmesi gerekecektir.

Ölçme işlemi ile incelenen birimler belli bir vasıf açısından kural dahilinde şıklara atanmaktadır. Bu şıklar; kadın/erkek veya bekâr/evli gibi isimlendirilerek veya 78 kilogram veya 161 santimetre gibi sayısal değerlerle gösterilmektedir. Şıklar kümesi ve aralarındaki ilişkiler, ölçüm düzeyini belirlemektedir. Şimdi bahsedilen ölçüm düzeylerini tanımlayalım.

Sınıflama (nominal) ölçüm düzeyi, incelenen birimleri belirli bir özellik açısından kategorilere atamaktadır. Diğer bir ifadeyle, incelenen birimleri belirli bir vasıf açısından şıklara atamaktadır. Bu atama işleminde, şıklar nitel yani isimseldir. Bu ölçüm düzeyine örnek olarak cinsiyet (kadın/erkek), medeni hâl (bekâr/evli), sağlık durumu (hasta, hasta değil), kan grubu (A, B, AB, 0) vasıfları verilebilir. Söz konusu ölçüm düzeyi ile amaçlanan, bir özelliğin varlığı ya da yokluğuna göre birimleri ayırt edici kategorilere atamaktır. Yani, tanımlama yapılmaktadır. Dolayısıyla şıklar, özelliğin miktarı hakkında bilgi vermemektedir. Bir otelin ikinci katında aynı donanım ve büyüklüğe sahip odaların kapı numaraları 200, 201, 202, ... , 208 şeklindedir. Yani, 208 numaralı oda 200 numaralı odaya göre 8 metrekare büyük değil ve konaklama ücreti aynıdır. Burada amaçlanan, her bir odayı diğerinden ayırt edilebilir kılmaktır. Buna ek olarak, sınıflama ölçüm düzeyine insan isimleri, futbolcuların sırt numaraları, apartmanların numaraları da örnek gösterilebilir. Başka bir örnek olarak ise, bir topluluğa ilişkin

toplanan veride medeni durum özelliği açısından insanların kategorize edilmesi verilebilir. Bilindiği üzere; bekâr bireylerin evli bireylere göre bir üstünlüğü ya da dezavantajı bulunmamaktadır. Dolayısıyla bireyler, medeni durum özellikleri açısından ayırt edilebilir kılınmak amacıyla kategorilere atanmıştır.

Sıralama (ordinal) ölçüm düzeyi, incelenen birimleri belirli bir özelliğe göre sıra dahilinde kategorileştirmektedir. Sınıflama ölçüm düzeyinin özelliğine ilave olarak, bu ölçüm düzeyinde kategoriler arasında önem düzeyi ya da büyüklük-küçüklük ilişkisi bulunmaktadır. Yani bu ölçüm düzeyi; ayırt ediciliğin yanı sıra kategoriler arasındaki farklılığın miktarı/önemi hakkında da bilgi sunmaktadır. Bu ölçüm düzeyine örnek olarak başarı durumu (kötü, orta, iyi, pekiyi) , eğitim durumu (ilkokul, ortaokul, lise, ön lisans, lisans), gelir durumu (düşük, orta, yüksek) vasıfları verilebilir. Bu vasıfların sınıfları incelendiğinde; sınıflar arasındaki mesafenin/ aralığın eşit olmadığı görülmektedir. Dolayısıyla sınıflar arasında bir sıra ilişkisi olmasına rağmen, bu sınıflar arasındaki mesafeler eşit olmadığından aradaki fark ölçülememektedir. Bu ölçümle toplanmış veriye örnek olarak, üç farklı gazete markasına ilişkin tüketici tercihi (en çok tercih edilen, ikinci tercih edilen, üçüncü tercih edilen) verilebilir. Burada, en çok ve ikinci tercih yanıtları arasındaki fark ile ikinci ve üçüncü tercih edilen arasındaki fark eşit değildir. Dolayısıyla sınıflar (yanıtlar) üzerinden matematiksel işlemler gerçekleştirilememektedir.

Aralık (interval) ölçüm düzeyi ile incelenen birimlerin belirli bir özellik açısından sahip olduğu göreceli miktarları ölçülmektedir. Sıralama ölçüm düzeyinin özelliklerine ek olarak, bu ölçüm düzeyinde kategoriler (sınıflar) arasındaki mesafeler (farklar) eşittir. Dolayısıyla birimler arasındaki farklar hesaplanabilmektedir. Yani, bu ölçüm düzeyi ile toplanan veri üzerinden matematiksel işlemler gerçekleştirilebilmektedir. Fakat aralık ölçüm düzeyinde sıfır noktası, keyfi olarak belirlendiğinden mutlak yokluğu yansıtmamaktadır. Bu ölçüm düzeyine yaygın olarak sıcaklık örneği verilmektedir. Çünkü Fahrenheit ve Santigrat birimleri için sıfır noktası, mutlak yokluğu göstermemektedir. Bu durumu anlatabilmek için hava sıcaklığının Adana’da 20 santigrat derece ve Kars’ta 5 santigrat derece olduğunu varsayalım. Bu örnek kapsamında, iki şehir arasındaki sıcaklık farkının 15 santigrat derece olduğu söylenebilmektedir. Fakat Adana’nın, Kars’tan dört kat sıcak olduğu söylenememektedir. Bu örneği daha da pekiştirmek için Moskova’nın -5 santigrat derece olduğunu varsayıp; Adana şehrinin sıcaklık değeri ile kıyaslamanız yeterli olacaktır.

Oran (ratio) ölçüm düzeyi, aralık ölçüm düzeyinin özelliklerine ek olarak mutlak yokluğu gösteren sıfır noktasına sahiptir. Oran ölçüm düzeyine sahip veri, matematiksel ve istatistiksel işlemlere elverişlidir. Bireylerin boy uzunluğu, ağırlık, yaş gibi özellikleri (sayısal sınıflara sahip olduğunda) mutlak yokluğu gösteren sıfır noktasına sahip olduğundan oran ölçüm düzeyi ile toplanmış verilere örnektir. Bu topluluktan Ali’nin ağırlığı 100 kilogram, Ayşe’ninki 50 kilogramdır. Bu bilgiden hareketle; Ali’nin Ayşe’den 50 kilogram daha ağır olduğu ve Ali’nin Ayşe’den iki kat ağır olduğu sonuçlarına ulaşılmaktadır. Hatırlanacağı üzere; aralık ölçüm düzeyinde oranlama işlemi yapılamamasına rağmen, mutlak yokluğu gösteren sıfır noktasına sahip olan oran ölçüm düzeyinde yapılabilmektedir. Mutlak yokluğu gösteren sıfır noktasının zihninizde pekişmesi için “Çocuk sayınız?” sorusuna vereceğiniz yanıtı düşünün. Bu soruya sıfır yanıtı vererseniz, hiç çocuğunuz olmadığı göstermektedir. Bu sorunun yanıtı,

-2 gibi bir deęer olamamaktadır. Hâlbuki santigrat cinsinden hava sıcaklığı negatif deęerler alabilmektedir.

Bahsedilen ölçüm düzeylerinin sahip oldukları özellikler Tablo 2.2’de özetlenmiştir.

Tablo 2.2: Ölçüm Düzeyleri Arasındaki İlişki

Ölçüm Düzeyi	Şıklar birbirinden farklı mı?	Şıkların sıralamasının bir önemi var mı?	Şıklar arasındaki farklar eşit mi?	Mutlak yokluğu gösteren sıfır noktası var mı?
Sınıflama	Evet	Hayır	Hayır	Hayır
Sıralama	Evet	Evet	Hayır	Hayır
Aralık	Evet	Evet	Evet	Hayır
Oran	Evet	Evet	Evet	Evet

Bu kısmı sonlandırırken verinin ölçüm düzeyi ile vasıf türü arasındaki ilişkiyi de belirtmek gerektiğini düşünüyoruz. Bilindiğı üzere; vasıf, nitel ve nicel olmak üzere temel iki türe ayrılmaktadır. Nitel vasıf, sınıflama ve sıralama ölçüm düzeyi ile ölçülürken; nicel vasıf ise aralık ve oran ölçüm düzeyi ile ölçülmektedir. Ayrıca sınıflama ve sıralama ölçüm düzeyi ile toplanmış veri **nitel**, aralık ve oran ölçüm düzeyi ile toplanmış veri ise **nicel** olarak ifade edilmektedir. Dolayısıyla veriyi de nitel ve nicel olmak üzere iki başlıkta gruplayabiliriz.

Bu Bölümde Neler Öğrendik?

Bu bölümde, istatistiğin konusunu teşkil eden kollektif olaylardan her biri olarak tanımlanan birim kavramı üzerinde durduk. Birimin çeşitlerini ele alarak, birimin türüne bağlı olarak hangi tür rölöve ile veri toplama işleminin yapılacağını inceledik.

Birimin özelliklerini gösteren vasıf ve şık kavramlarını ve çeşitlerini öğrendik.

Birimin ölçümüne yönelik olarak ölçüm düzeyi kavramını ve ölçüm düzeyi sınıflamasını öğrendik.

Bölüm Soruları

- 1- İstatistiğin konusunu oluşturan kollektif olaylardan her birine denir.
 - a- Birim
 - b- Vasıf
 - c- Şık
 - d- Veri
 - e- İnsan
- 2- Fiziksel bir varlığa sahip olan birimlere denir.
 - a- Varlık birim
 - b- Fiziksel birim
 - c- Maddi birim
 - d- Somut birim
 - e- Soyut birim
- 3- Birim, uzun süre gözlenebiliyorsa ya da uzun ömürlü olarak tanımlanıyorsa.....
birim söz konusu olmaktadır.
 - a- Maddi
 - b- Devamlı
 - c- Ani
 - d- Soyut
 - e- Somut
- 4- Birimlerin sahip oldukları özelliklere denir.
 - a- Şık
 - b- Özellik
 - c- Vasıf
 - d- Ölçüm düzeyi
 - e- Ölçek

5- Vasıfların ortaya çıkış şekillerine denir.

- a- Özellik
- b- Ölçüm düzeyi
- c- Ölçek
- d- Şık
- e- Vasıf

6- Aşağıdakilerden hangisi maddi birimdir?

- a- İnsan
- b- Doğum
- c- Ölüm
- d- Evlenme
- e- Boşanma

7- Aşağıdakilerden hangisi nicel vasfa örnektir?

- a- Doğum yeri
- b- Meslek
- c- Yaş
- d- Eğitim durumu
- e- Medeni durum

8- Aşağıdakilerden hangisi nitel vasıftır?

- a- Medeni durum
- b- Yaş
- c- Boy
- d- Kilo
- e- Bel çevresi

- 9- Mutlak sıfır noktasına sahip olan ölçüm düzeyidir.
- a- Sınıflayıcı ölçüm düzeyi
 - b- Sıralayıcı ölçüm düzeyi
 - c- Aralık ölçüm düzeyi
 - d- Oran ölçüm düzeyi
 - e- Nominal ölçüm düzeyi
- 10- Birimleri, sıralama işlemi olmaksızın sadece kategorilere ayırmaya yarayan ölçüm düzeyine.....denir.
- a- Sıralayıcı ölçüm düzeyi
 - b- Aralık ölçüm düzeyi
 - c- Interval ölçüm düzeyi
 - d- Oran ölçüm düzeyi
 - e- Sınıflayıcı ölçüm düzeyi

Cevaplar

- 1) a
- 2) c
- 3) b
- 4) c
- 5) d
- 6) a
- 7) c
- 8) a
- 9) d
- 10) e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

3. VERİLERİN DÜZENLENMESİNDE TASNİF VE GRUPLAMA

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde önce veri toplama işlemi ile elde ettiğimiz ham verilerin nasıl işleneceğini ve düzenleneceğini öğreneceğiz. Bu bağlamda verileri düzenlemek amacıyla kullanılabilecek yöntemlerden olan tasnif ve gruplama işlemlerinin nasıl yapıldığı üzerinde duracağız. Özellikle gruplama işlemi ile ilgili olarak dikkat edilmesi gereken hususları ele alacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Toplanan veriler, herhangi bir işleme tabi tutulmazsa anlaşılabilir mi?
- 2) Verileri düzenlemek neden gereklidir?
- 3) Verileri düzenlerken hangi yöntemler kullanılabilir?
- 4) Verileri düzenleme, verinin kullanımına yönelik kısıtlara sebep olur mu?
- 5) Verileri tasnif ve gruplama yöntemlerini kullanarak düzenlerken dikkat edilmesi gereken noktalar nelerdir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Toplanan verilerin düzenlenmesi.	Veri toplama süreciyle elde edilen verilerin nasıl düzenleneceğini ve bu amaçla kullanılabilecek yöntemlerin neler olduğunu bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Verilerin tasnif yöntemiyle düzenlenmesi	Tasnif yöntemiyle verinin nasıl düzenleneceğini ve bu işlem sırasında nelere dikkat edilmesi gerektiğini bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Tasnif yönteminin yetersiz kaldığı durumlar	Tasnif yönteminin verileri düzenlemede yetersiz kaldığı durumlar konusunda fikir sahibidir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Verilerin gruplama yöntemiyle düzenlenmesi	Toplanan verilerin gruplama yöntemiyle nasıl düzenleneceğini bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar, veriler ve çeşitli kurumlar tarafından oluşturulmuş istatistik tablolar.
Gruplama yönteminde dikkat edilmesi gereken hususlar	Toplanan verilerin gruplama yöntemiyle düzenlenmesi konusunda dikkat edilmesi gereken hususların neler olduğunu bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar.

Anahtar Kavramlar

- Verilerin düzenlenmesi: Veri toplama süreciyle elde edilen ham verilerin kullanılabilir hale getirilmesi.
- Tasnif: Toplanan verilerin düzenlenmesi amacıyla sıklıkla kullanılan veri düzenleme yöntemi.
- Gruplama: Büyük veri kümelerini düzenlemede kullanılan veri düzenleme yöntemi.

Giriş

Rölöve yani veri toplama süreci ile elde edilen veriler ancak bazı işlemlerden geçirildikten sonra anlamlı hale gelir.

İstatistik dersini almış 100 öğrencinin dönem sonu başarı notları ile ilgili bir araştırma yaptığımızı düşünelim. İlgili notlara sınıf sınav sonuç listesi üzerinden kolaylıkla ulaşmamız mümkündür ve görüleceği gibi bu oldukça kolay bir veri toplama işlemidir. Dolayısıyla, elimizde 100 öğrencinin numara sırasına göre oluşturulmuş başarı notları mevcuttur. Liste düzeni öğrencilerin numara sırasına göre yapılmış olduğundan 0-100 aralığında öğrencilerin aldığı başarı notları karışık bir şekilde liste üzerinde görülmektedir. Bu yapı içinde örneğin kaç öğrencinin 100, kaç öğrencinin 0 aldığını tespit etmek için tüm listeyi gözden geçirerek ilgili notları alan öğrencileri saymamız gerekecektir. Benzer şekilde kaç öğrenci 50'nin üzerinde not almıştır sorusuna cevap vermek için de yine tüm listeyi gözden geçirerek 50 puanın üzerinde notu olan öğrencileri saymamız gerekecektir. Bunun gibi soruları arttırabiliriz.

Görüldüğü gibi, veriler elimizde bir liste düzeni içinde bulunsada dahi, bu veri topluluğundan bilgi üretebilmek için bazı işlemler gerekmektedir. Verilerin düzenlenmesi işlemi, ham veri niteliği taşıyan toplanmış bir veri setinin anlamlı şekilde sıralanması, anlaşılır şekilde düzenlenmesi işlevini yerine getiren bir süreçtir.

3.1. Tasnif

Verilerin düzenlenmesi başlığı altında yapılabilecek en basit düzenleme olan tasnif, verileri incelenen vasfın şıklarına göre ayırmaktır. Başka bir deyişle, aynı şıkta sahip birimler kümelerine ayrılarak her bir şıkta kaç birimin yer aldığı tespit edilerek yapılan düzenleme işlemine **tasnif** adı verilmektedir.

Tasnif işlemi sonucunda, hangi şıkkın kaç birimde gözleendiği ya da kaç kez ortaya çıktığı kolayca görülebilmektedir. Örneğin, kaç öğrencinin 100 aldığı, kaç kişinin 0 aldığı merak edildiğinde bu bilgiye kolayca ulaşılabilir.

Burada bir kavram üzerinde durmak istiyoruz: Frekans.

Bir veri kümesinde hangi şıkkın kaç kez gözleendiğini ifade etmek üzere **frekans** kavramı kullanılmaktadır. Örneğin, yukarıdaki örnekte 100 alan öğrenci sayısı 5 ise, 100'ün frekansı 5 olmaktadır. Yani 100 değeri veri kümesi içinde 5 kez görülmüş ya da 5 kez tekrar etmiş demektir.

Tasnif işlemi bir örnekle açıklamaya çalışalım:

100 öğrencinin İstatistik dersi dönem sonu başarı notlarının tasnif edilmesi suretiyle oluşturulmuş tablo aşağıda verilmiştir:

Tablo 3.1: 100 Öğrencinin İstatistik Dersi Başarı Puanlarının Dağılımı

Notlar	Frekans
0	1
10	3
20	8
30	10
40	14
50	20
60	16
70	12
80	8
90	5
100	3
Toplam	100

Yukarıdaki tabloda hangi puanı kaç öğrencinin aldığını kolaylıkla görebildiğimiz gibi, çok basit işlemlerle belirli bir puanın altında ya da üstünde puan alan öğrenci sayısını da yine kolaylıkla bulabiliyoruz. Söz gelimi, kaç öğrencinin başarı puanının 100 olduğunu merak ediyorsak, yukarıdaki tablonun son satırından 3 öğrencinin puanının 100 olduğunu anlıyoruz. Benzer şekilde, sıfır alan öğrenci sayısının 1, on alan öğrenci sayısının 3 olduğunu yine kolayca görebiliyoruz.

Ellinin üzerinde başarı puanı olan öğrenci sayısını merak ediyorsak, yine aynı şekilde elli ve üzerinde puanı olan öğrencilerin frekanslarını toplayarak $(20+16+12+8+5+3)$ 64 kişinin bu aralıkta puan aldığını hesaplayabiliyoruz. Benzer şekilde elli puanın altında kalan öğrenci sayısını da ya bu puanlara karşılık gelen frekansları toplayarak $(1+3+8+10+14)$ ya da toplam 100 öğrenci bulunduğundan, bu sayıdan elli ve ellinin üzerinde puan alan öğrenci sayısı olan 64'ü çıkararak 36 olarak kolayca hesaplayabiliyoruz.

Tasnif işlemi, yukarıda verdiğimiz örnekte olduğu gibi nicel yani sayısal özellik gösteren vasıflarda kullanılabileceği gibi, nitel vasıflarda da veriyi düzenleme amaçlı kullanılabilir.

Nitel vasıf söz konusu olduğunda tasnifin nasıl yapılabileceğini bir örnekle açıklamaya çalışalım.

Bir fakültenin birinci sınıfına kayıt yaptıran 150 öğrencilerin mezun oldukları liselere yönelik bir verimiz olduğunu düşünelim. 150 öğrenci değişik lise türlerinden gelmiştir: Düz lise, Anadolu lisesi, fen lisesi, yabancı okullar, teknik lise, imam-hatip lisesi vb.

Birinci sınıfa kayıt yaptırmış 150 öğrenciyi mezun oldukları lise türlerine göre tasnif edebiliriz. Yani, aynı lise türünden mezun olan öğrencileri bir araya toplayarak, bu tür liseden mezun olmuş kaç öğrencimiz olduğunu görebiliriz.

Tablo 3.2 150 öğrencinin mezun oldukları lise türüne göre dağılımını göstermektedir.

Tablo 3.2: 150 Öğrencinin Mezun Oldukları Lise Türüne Göre Dağılımı

Lise Türü	Frekans
Düz Lise	35
Anadolu Lisesi	43
Fen Lisesi	25
Teknik Meslek Lisesi	10
İmam-Hatip Lisesi	10
Yabancı Lise	25
Güzel Sanatlar Lisesi	2
Toplam	150

Tasnif işlemi sadece aynı şıkka sahip birimleri bir araya toplamayı amaçlayan en basit veri düzenleme yöntemidir. Aynı şıkka sahip birimlerin bir araya toplanması suretiyle tasnif işlemi uygulanmış olur. Şık sayısının çok fazla olması halinde tasnif işlemi ile yetinilmeyip grupta yapma yoluna gidilir.

3.2. Gruplama

Bir vasfın tasnif edilecek şık sayısı çok fazla ise yalnızca tasnif veriyi düzenlemek açısından yeterli olmayabilir.

Örneğin, bir şehirde yaşamakta olan insanların yaş dağılımları ile ilgili bir çalışma yaptığımızı varsayalım. En yaşlı bireyin 100 yaşında olabileceğini varsayarsak yaş vasfının şıkları 0-100 aralığında olacaktır. Bu durumda sadece tasnif yoluna gidersek, hangi yaşta kaç kişi olduğunu tespit etmemiz gerekecek ve verimiz 100 satırlık bir tablo ile düzenlenebilecektir. Böylelikle, bu şehirde yaşamakta olan insanların kaç tanesinin 1, kaç tanesinin 2, kaç tanesinin 3,, kaç tanesinin 99 ve kaç tanesinin 100 yaşında olduğunu tasnif işlemiyle görebiliriz. Ancak 100 satırdan oluşan bir tablodan bilgi almak oldukça güçtür.

Benzer şekilde, ülkemizde faaliyet gösteren işletmeleri çalıştırdıkları işçi sayılarına göre tasnif ettiğimizi varsayalım. Hiç işçi çalıştırmayan iş yerleri ile binlerce işçi çalıştıran işyerlerine kadar çok şıklı bir vasıf ile karşı karşıyayız. Bu durumda, tasnif işlemi ile veriyi düzenleme yoluna gidersek binlerce satırlık bir tablo oluşturmamız gerekecektir. Binlerce satırdan oluşan böyle bir tabloyu takip etmek ve oradan sonuçlar çıkarmak neredeyse mümkün değildir.

Yukarıda örneklerini verdiğimiz, çok sayıda şıkkı olan nicel vasıflar, benzer şıklar bir araya toplanmak suretiyle grupta yoluna gidilerek de düzenlenebilir. Burada birbirine yakın şıkları bir araya toplamak esastır.

Örneğin, yaş dağılımı ile ilgili çalışmada bireylerin yaşları mümkün olduğunca homojen gruplara toplanabilir. Böylelikle, çok fazla şıkkı olan veriler daha anlaşılır bir şekilde düzenlenmiş olur. Sözelimi, 0, 1, 2,3,...., 8,9 yaşlarındaki bireyler 0-10 yaş arası grup, 10, 11, 12,....,17, 18, 19 yaşlarındaki bireyler 10-20 yaş arası grup şeklinde düzenlenebilir.

Nicel vasıfla ilgili olarak gruplamanın neden gerekli olduğunu ve nasıl yapılabileceğini bir örnekle açıklamaya çalışalım.

Bir köyde yaşayan 225 kişinin yaş dağılımı aşağıda yer alan Tablo 3.3. ile düzenlenmiştir:

Tablo 3.3: Bir Köyde Yaşayanların Yaşlara Göre Dağılımı

Yaş	Frekans	Yaş	Frekans	Yaş	Frekans	Yaş	Frekans	Yaş	Frekans
1	1	21	-	41	4	61	2	81	2
2	1	22	2	42	6	62	3	82	1
3	2	23	2	43	3	63	4	83	1
4	1	24	3	44	4	64	3	84	1
5	2	25	1	45	6	65	5	85	1
6	2	26	2	46	6	66	3	86	-
7	1	27	4	47	4	67	2	87	-
8	2	28	4	48	4	68	2	88	1
9	2	29	3	49	3	69	1	89	-
10	3	30	3	50	4	70	1	90	1
11	2	31	4	51	4	71	1	91	1
12	2	32	3	52	5	72	2	92	-
13	3	33	4	53	5	73	3	93	-
14	1	34	-	54	4	74	3	94	-
15	1	35	4	55	3	75	4	95	-
16	2	36	4	56	1	76	2	96	-
17	2	37	2	57	4	77	2	97	-
18	1	38	3	58	4	78	2	98	-
19	2	39	2	59	3	79	1	99	-
20	2	40	4	60	3	80	1	100	-

Tablo 3.3.'te bir köyde yaşayan 225 kişinin yaşlara göre dağılımını görmekteyiz. Yaşlar küçükten büyüğe doğru sıralanmış ve her yaş için o yaşta kaç kişinin bulunduğu tasnif yoluyla belirlenmiş ve veri buna göre düzenlenmiştir.

Daha önce ifade ettiğimiz gibi yaş, nicel bir vasıf olup çok şıklıdır. Dolayısıyla, her şık tabloda bir satıra karşılık gelmekte ve oldukça büyük bir tablo kullanmayı gerektirmektedir. Çok fazla satırı olan tablolardan bilgi almak, onları yorumlamak oldukça güçtür. Sözgelimi, yukarıdaki tablodan hangi yaşta kaç kişinin bulunduğunu ilk bakışta tespit edebilirken köyün genç bir nüfusa mı yoksa yaşlı bir nüfusa mı sahip olduğunu ise kolayca söyleyemeyiz.

Örneğimizde olduğu gibi, çok şıklı nicel vasıflarda birbirine yakın şıklar bir araya toplanıp gruplanarak daha anlaşılır bir düzenleme yoluna gitmek tercih edilir. Bir insanın 32 ya da 33 yaşında olması yaş özelliği düşünüldüğünde iki insan arasında büyük bir farklılık olarak düşünülemez. Benzer şekilde; 51, 52, 54 yaşlarında olmak da insanlar arasında yaş açısından ayırt edici derecede önemli bir farklılık olarak değerlendirilmez.

İşte böyle durumlarda çok şıklı vasıflarda birbirine yakın olan ve çok farklı olarak değerlendirilemeyecek şıklar bir araya getirilerek, yani gruplanarak veriyi daha anlaşılır bir düzenleme yoluna gitmek tercih edilmektedir.

Örneğimizde köyde yaşayan 225 kişinin yaşları, birbirine yakın yaşlar bir araya getirilerek gruplanmış ve Tablo 3.4. ile düzenlenmiştir. Gruplamanın nasıl yapıldığını ve hangi yaşların hangi grupların içinde yer aldığını tablonun ilk sütunundan kolaylıkla izleyebiliriz. Örneğin; 1, 2, 3, 4, 5, 6, 7, 8, 9 yaşlarında olanlar 0-10 yaş grubunda toplanmıştır ve köyde bu yaş grubunda 14 kişi bulunmaktadır. Benzer şekilde, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 yaşlarında olanlar da 10-20 yaş grubunda toplanmış olup, bu gruptaki insan sayısı ise 19 kişi olarak tespit edilmiştir.

Tablo 3.4: Bir Köyde Yaşayanların Yaşlara Göre Dağılımı

Yaşlar	Yaş Grupları	Frekanslar
1,2,3,4,5,6,7,8,9	0-10	14
10,11,12,13,14,15,16,17,18,19	10-20	19
20,21,22,23,24,25,26,27,28,29	20-30	23
30,31,32,33,34,35,36,37,38,39	30-40	29
40,41,42,43,44,45,46,47,48,49	40-50	44
50,51,52,53,54,55,56,57,58,59	50-60	37
60,61,62,63,64,65,66,67,68,69	60-70	28
70,71,72,73,74,75,76,77,78,79	70-80	21
80,81,82,83,84,85,86,87,88,89	80-90	8
90,91,92,93,94,96,97,98,99,100	90-100	2
Toplam		225

Bu noktada bir hususa dikkat çekmek istiyoruz. Tablo 2.4.'den görüleceği gibi yaş grupları 0-10, 10-20, 20-30 şeklinde düzenlenmiştir. Dolayısıyla 10 değeri hem 0-10 hem 10-20 sınıfında yer alıyor görünmektedir. Tablo 2.2.'den 10 yaşında olan 3 kişi bulunduğu anlaşılıyor. Bu 3 kişinin hem 0-10 hem 10-20 yaş grupları içinde yer alması tabii ki

düşünülemez. Eğer her iki yaş grubunda yer alırlarsa 10 yaşında olan aynı 3 kişi her iki grubun içinde yer alarak 2 kez sayılmış, değerlendirilmiş olur.

O halde 10 yaşında olan 3 kişi hangi sınıfa dahil edilecektir?

Tablo 3.4’de kullanılan yaş gruplarını, her bir yaşın sadece bir gruba dâhil edilmesi başka bir deyişle hiçbir yaşın iki gruba aynı anda girmemesi için değerlendirmemizi şu şekilde yapmak doğru olacaktır.

Bu hususu daha iyi kavrayabilmek için gruplama ile ilgili bazı tanımlamalar yapmak istiyoruz.

Yaş gruplarını 0-9, 10-19, 20-29.... şeklinde oluştursaydık yukarıdaki sorunu yaşamayacaktık. Tablo 3.5. bu şekilde yapılmış gruplama ile elde edilen yaş dağılımını göstermektedir.

Tablo 3.5:Bir Köyde Yaşayanların Yaşlara Göre Dağılımı

Yaş Grupları	Frekanslar
0-9	14
10-19	19
20-29	23
30-39	29
40-49	44
50-59	37
60-69	28
70-79	21
80-89	8
90-99	2
Toplam	225

Tablo 3.5’ de yer alan gruplama biçiminde her bir yaş sadece bir gruba dâhil olmaktadır ve dolayısıyla bir önceki tabloda görülen durum söz konusu olmamaktadır.

Ancak uygulamada 0-10, 10-20, 20-30..... şeklinde yapılan düzenlemenin daha çok tercih edildiğini görüyoruz. Bunun öncelikli sebebi görsel olarak gruplarda devamlılık sağlanmasıdır. Burada yukarıda sözünü ettiğimiz aynı değer iki grupta da yer alması sakıncasından basit bir varsayımla uzaklaşıyoruz: **-den az varsayımı**.

Tablo 3.4 ‘de yer alan grupları, -den az varsayımı ile birlikte değerlendiriyoruz. Aşağıda Tablo 3.6 ile düzenlediğimiz yeni tabloyu görebilirsiniz:

Tablo 3.6: Bir Köyde Yaşayanların Yaşlara Göre Dağılımı

Yaş Grupları	Frekanslar
0-10 den az	14
10-20 den az	19
20-30 den az	23
30-40 den az	29
40-50 den az	44
50-60 den az	37
60-70 den az	28
70-80 den az	21
80-90 den az	8
90-100 den az	2
Toplam	225

Tablo 3.6'yı yorumlarken her bir yaş grubunun üst sınırını, sınır değerinden az şeklinde düşünüyoruz. Örneğin, 0-10'dan az yaş grubu 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 yaşlarını kapsamaktadır, zira 10'dan az şeklindedir yaş gruplaması.

Benzer şekilde, 10-20'den az yaş grubu da 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 yaşlarını kapsamaktadır ve yaş grubu 20'den az şeklinde nitelendiği için 20 bu yaş grubuna dâhil olmamaktadır.

Yukarıda aynı veri için düzenlenmiş üç farklı tablo görmektesiniz. Tablo 3.4 ile Tablo 3.5 ve Tablo 3.6'yı karşılaştırdığımızda yaş gruplarına ilişkin sınır değerlerinin farklı olduğunu ancak frekansların değişmediğini görmekteyiz. Her bir tablo, ilgili veriyi düzenleme açısından doğrudur ve herhangi bir hata söz konusu değildir. O halde hangi gösteriş şekli daha uygun ya da daha kullanışlıdır?

Bu durumu izah etmek üzere kesikli ve sürekli vasıf özelliğini hatırlatmak ve konuya bu açıklamalar üzerinden yaklaşmak istiyoruz.

Daha önce de ifade ettiğimiz gibi yaş kesikli bir vasıftır ve sadece tam sayı ile ifade edilir. Dolayısıyla, bir insanın yaşı ya 9'dur ya da 10'dur. 9 ile 10 arasında bir değer alması mümkün değildir.

Fakat boy ya da ağırlık gibi vasıflar ise sürekli vasıf özelliği taşımakta ve gözlenen birimler iki değer arasında sonsuz değer alabilmektedir. Söz gelimi bir objenin uzunluğu 9 cm ya da 10 cm olabileceği gibi, 9,5 ya da 9,25 cm'de olabilir.

Gruplama işlemi sırasında, grup ya da sınıfların sınır değerlerini belirlerken –den az varsayımını kullanarak alt sınır değerinin dâhil olduğu ve üst sınır değerine en yakın değere kadar olan değerlerin tamamının kapsandığı ancak üst sınır değerinin dışarıda tutulduğu bir yaklaşım benimsenmektedir. Bu durumu bir örnek yardımıyla açıklamaya çalışalım. Sözelimi 0-10'dan az yaş grubu 0 yaş ile başlayıp, 1, 2, 3, 4, 5, 6, 7, 8, 9 yaşlarını kapsamakta, 10'a en

yakın ancak 10'a eşit olmayan en son değer, örneğin 9,99999 değerine dek, bu yaş grubunda değerlendirilmektedir. Yaş, kesikli bir vasıf olduğundan ve küsuratlı değer alması mümkün olmadığından, yaşla ilgili dağılımda 0-10'dan az yaş grubu için, yaş sınırları içinde kalan yaşlar 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 olacaktır. 10-20'den az yaş grubunda ise 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 yaşları yer alacaktır.

Benzer sınırlar söz gelimi bir uzunluk vasfı için söz konusu olduğunda ise, 0-10'dan az sınıfı 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,99999 değerine kadar tüm değerleri kapsayacaktır. Benzer şekilde 10-20'den az sınıfı da 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,9999 değerine kadar tüm değerleri bu sınıfta toplayacaktır.

Sınıflama konusunda, aynı veriye ilişkin olarak düzenlediğimiz yukarıda verilen 3 tablodan hangisinin kullanılacağına yönelik olarak son bir açıklama ile bu konuyu kapatmak istiyoruz.

Görsel olarak yaş gruplarının 0-10, 10-20, 20-30 şeklinde benimsendiği ve –den az şeklinde bir açıklamaya tabloda yer verilmeksizin bu kriterin geçerli olduğunun kabul edildiğini görmekteyiz.

Gruplama İle İlgili Temel Kavramlar Ve Sınıf Aralığının (Genişliğinin) Önemi

Bu noktada gruplama ya da sınıflamaya yönelik olarak basit birkaç kavram üzerinde durmak istiyoruz. 0-10 sınıfını ele alalım. Burada sıfır değerine alt sınır değeri, 10 değerine ise üst sınır değeri denmektedir. Üst ve alt sınır değerleri arasındaki farka ise **sınıf aralığı (genişliği)** adı verilmektedir.

Benzer şekilde 10-20 sınıfı için de, alt sınır değeri 10, üst sınır değeri 20 ve sınıf genişliği de 10 olmaktadır.

Bu açıklamaları da dikkate aldığımızda, sınıfların içinde yer alacak birimlerin tespit edilmesi sırasında, alt sınır değerinin ilgili sınıfa ait olduğu, dolayısıyla alt sınır değerine sahip birimlerin bu sınıfta değerlendirileceği, üst sınır değerinin ise ilgili sınıfa ait olmadığı ve dolayısıyla üst sınır değerine sahip birimlerin bu sınıfta yer almayacağı ve bu birimlerin izleyen sınıfta sayılacağı anlaşılmalıdır.

Gruplama ya da sınıflama ile ilgili olarak dikkat edilmesi gereken birkaç nokta üzerinde durmak istiyoruz.

Gruplama birbirine yakın şıkka sahip birimleri bir araya toplama esasına dayanır. Dolayısıyla, aynı grup ya da sınıf içinde yer alan birimlerin homojen olması, birbirinden çok farklı özellik göstermemesi gerekecektir.

Bu konuyu basit bir örnekle açıklayalım. Bir sınavda 0-100 aralığındaki sınav sonuçlarının, sınıf genişliği 20 olacak şekilde düzenlendiğini ve geçme notunun 50 olduğunu varsayalım. Sınıflar,

0-20
20-40
40-60
60-80
80-100

şeklinde düzenlenecek ve bu not sınıflarına düşen notlar sayılarak her bir sınıfın frekansı belirlenecektir.

Yukarıdaki gibi bir sınıf düzenlemesinde 40-60 sınıfı hem 50'nin altında not alıp başarısız olanları, hem 50 veya üzerinde not alıp başarılı olanları bir arada değerlendirecektir. Dolayısıyla, böyle bir sınıflama yapılması halinde kaç kişinin bu sınavda 50 ve 50'nin üzerinde not alarak başarılı olduğu tespit edilemeyecektir.

Gruplama birbirine benzer sınıfları bir araya toplama esasına dayandığından, gruplama yapılması halinde veriye ilişkin detaylar kısmen kaybolmaktadır. Belli bir sınıfta not alan öğrenci sayısını bilmekte ancak, o sınıf içinde yer alan hangi notu kaç öğrencinin aldığı bilinmemektedir.

Gruplama veri ile ilgili detayların kaybolmasına sebep olduğundan, gruplama sırasında kullanılacak sınıf genişliğinin belirlenmesi oldukça önemlidir.

Çok büyük sınıflar, sınıf içinde toplanan verinin homojenliğini zedelemenin yanında, yüksek miktarda bilgi kaybına da yol açmaktadır.

Not örneğimizden devam edelim ve sınıflarımızı 0-50 ve 50-100 şeklinde belirlediğimizi varsayalım. Sınıf genişliğimiz bir önceki sınıflamaya göre çok arttığından sınıf içi homojenlik azalmıştır. Öte yandan, 50-100 sınıfında not alan öğrencilerin başarı düzeyleri arasındaki farklar da göz ardı edilmiş olmaktadır. Örneğin, 90 ve üzerinde not alan en başarılı öğrenciler ile 50 alıp dersi geçen başarısı en düşük öğrencilerle aynı grupta değerlendirilmiş olmaktadır.

Bu başlıkta son olarak sınıf genişliğinin ne olması gerektiğini belirlemek üzere geliştirilmiş bir kuraldan söz etmek istiyoruz.

Sınıf genişliğinin doğru olarak tespit edilmesi, veriyi doğru şekilde düzenlemek ve söz konusu veri kullanılarak gerçekleştirilecek analizlerden doğru sonuçlar elde etmek açısından son derece önemli bir konudur.

Sınıf genişliğinin ne olması gerektiği Sturge Kuralı adı verilen bir kural ya da formül vasıtasıyla kolaylıkla belirlenebilmektedir.

Sturge Kuralına ve formülüne kitabın ilerleyen sayfalarında yer verilecektir.

Nitel vasıf için gruplama örneği

Yukarıda verdiğimiz örnekler şıkları nicel özellik gösteren vasıflara yönelik olmakla beraber, şık sayısı çok fazla olan nitel vasıflar için de benzer durum söz konusudur. Sözelimi, meslek dağılımını ele alalım. Bakkal, manav, kuaför,, doktor, avukat, sanayici vb. çok sayıda meslek bulunmaktadır ve dolayısıyla veri düzenlemede sadece tasnif yoluna gidilmesi halinde yine çok fazla sayıda satırı olan bir tablo ile karşı karşıya kalınmaktadır.

Nitel vasıf durumunda da birbirine yakın şıklar bir arada toplanmak suretiyle gruplama yoluna gitmek ve veri setini bu şekilde düzenlemek mümkündür.

Meslek örneğimizi ele alalım. Çok sayıda değişik meslek mensuplarını mümkün olduğunca birbirine yakın gruplarda toplamaya çalışalım. Bakkal, manav, kasap, kuaför vb. meslekleri küçük esnaf başlığı altında gruplayabiliriz. Benzer şekilde, özel sektörde beyaz yakalı tabir edilen ofis çalışanlarını, özel sektörde memur; özel sektörde mavi yakalı olarak tanımlanan ofis temizliği ve çay- kahve servisi yapanları, yemekhane çalışanlarını, işçileri ise özel sektörde işçi grubunda değerlendirebiliriz.

Tablo 3.7: Bir Mahallede Yaşayanların Meslek Gruplarına Göre Dağılımı

Meslek Grupları	Frekanslar
Kamuda işçi	140
Özel sektörde işçi	170
Kamuda memur	85
Özel sektörde memur	125
Küçük esnaf	48
Serbest Çalışan (doktor, avukat, mali müşavir vb.)	32
Küçük ölçekli işletme sahibi (1-5 işçi çalıştıran)	15
Orta büyüklükte işletme sahibi (6-25 işçi çalıştıran)	7
Büyük ölçekli işletme sahibi (25 + işçi çalıştıran)	3
Toplam	625

Nitel vasıfla ilgili olarak yapılabilecek gruplama işlemine Trafik Sicil Amirliğinde kayıtlı taşıt araçları da örnek verilebilir. Bilindiği gibi, çok sayıda marka ve modelde çeşitli taşıt araçları trafikte kayıtlıdır. Bu araçlarda, binek otomobil, 4x4 arazi taşıtları, minibüs, midibüs, otobüs, kamyon, kamyonet, çekici, motosiklet ve iş makineleri şeklinde gruplanabilir. Böylelikle, marka ve modeli ne olursa olsun tüm binek otomobiller bir grupta toplanmakta ve bu sınıfta kayıtlı kaç araç olduğu kolaylıkla görülmektedir.

Özetlersek, şıkları çok fazla olan vasıflara yönelik verileri düzenlemede sadece tasnif işleminin yeterli olmadığını ve benzer ya da birbirine yakın şıkları bir araya toplamak suretiyle gerçekleştirilen gruplama / sınıflama işleminden faydalanılması gerektiğini söyleyebiliriz.

Bu Bölümde Ne Öğrendik?

Bu bölümde; toplanan verilerin ham veri olarak nitelendiğini ve gerek istatistik analizler yapabilmek gerekse verilerin ilgililerce anlaşılır halde sunulabilmesi için kullanılabilecek veri düzenleme yöntemlerini öğrendik.

Bu amaçla öncelikle, verilerin aynı sınıfların bir araya toplanması suretiyle nasıl tasnif edilebileceğini ve tablolarla gösterilebileceğini öğrendik.

Daha sonra, göreceli olarak daha çok gözlem ihtiva eden verilerin, birbirine yakın değer alan gözlem değerlerinin gruplanması suretiyle nasıl düzenlendiğini ve gruplama işleminin nasıl yapılması gerektiğini ve grupların oluşturulması sırasında sınır değerlerin belirlenmesi aşamasında nelere dikkat edilmesi gerektiğini öğrendik.

Bölüm Soruları

- 1- Verilerin aynı şıkların bir araya toplanması şeklinde düzenlenmesine..... denir.
 - a- Gruplama
 - b- Tasnif
 - c- Sınıflama
 - d- Sıralama
 - e- Düzenleme
- 2- Verilerin, birbirine yakın şıklarını mümkün olduğunca homojen sınıflara toplamak suretiyle yapılan veri düzenleme işleminedenir.
 - a- Gruplama
 - b- Tasnif
 - c- Sıralama
 - d- Düzenleme
 - e- Ayırma
- 3- Verileri birbirine yakın homojen gruplara toplamak suretiyle oluşturulan alt ve üst sınırı belli aralıkların her birinedenir.
 - a- Frekans
 - b- Tasnif
 - c- Birim
 - d- Sınıf
 - e- Sınır
- 4- Bir gözlem değerinin seri içinde kaç kez tekrar ettiğini gösteren değeredenir.
 - a- Frekans
 - b- Tasnif
 - c- Birim
 - d- Sınıf
 - e- Grup

- 5- Bir sınıfın üst sınırı ile alt sınırı arasındaki farkadenir.
- a- Sınıf üst sınırı
 - b- Sınıf alt sınırı
 - c- Sınıf genişliği
 - d- Sınıf orta noktası
 - e- Sınıf sayısı
- 6- Gruplama işlemi yapılırken, oluşturulacak sınıflarınolmasına dikkat edilmelidir.
- a- Homojen
 - b- Heterojen
 - c- Sayısal
 - d- Geniş
 - e- Dar
- 7- Gruplama işlemi sırasında, sınıfların tutulması sınıfların homojenliğine zarar vererek detayların kaybolmasına yol açar.
- a- Dar
 - b- Geniş
 - c- Aralıklı
 - d- Sayısal
 - e- Sayısal olmayan
- 8- Bir gruplama işlemi ile oluşturulan sınıflarda, o sınıfta bulunabilecek en küçük değeredenir.
- a- Sınıf genişliği
 - b- Sınıf aralığı
 - c- Sınıf ortası
 - d- Sınıf alt sınırı
 - e- Sınıf üst sınırı

9- Bir sınıfın içinde kaç tane gözlem yer aldığını gösteren değeredenir.

- a- Tasnif
- b- Sınıf
- c- Frekans
- d- Gözlem
- e- Grup

10- Gözlem sayısının çok fazla olması durumunda, verileri homojen gruplara toplamak suretiyle yapılan düzenleme işlemine.....denir.

- a- Tasnif
- b- Sınıf
- c- Frekans
- d- Gözlem
- e- Gruplama

Cevaplar

- 1- b
- 2- a
- 3- d
- 4- a
- 5- c
- 6- a
- 7- b
- 8- d
- 9- c
- 10- e

Yararlanılan Kaynaklar

ALOBİ, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), "On the Theory of Scales of Measurement" , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

4. VERİLERİN DÜZENLENMESİNDE SERİLER VE GRAFİKLER

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

DR.ÖĞR. ÜYESİ HAKAN BEKTAŞ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde önce veri toplama işlemi ile elde ettiğimiz ham verilerin nasıl işleneceğini ve düzenleneceğini öğreneceğiz. Bu bağlamda verileri düzenlemek amacıyla kullanılacak yöntemlerden olan seri türlerini öğrenecek ve bu işlemlerinin nasıl yapıldığı üzerinde duracağız. Ayrıca, grafik türlerini öğrenecek ve verilerin grafiklerle nasıl gösterildiği üzerinde duracağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) İstatistikte seriler kavramı neyi ifade eder?
- 2) Basit, Tasnif edilmiş ve sınıflanmış serilerle verilerin düzenlenmesi nasıl yapılır?
- 3) Zaman serisi ve mekân serisi kavramları nedir?
- 4) Panel veri nedir?
- 5) Grafikler ile verilerin sunulması nasıl olur?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Toplanan verilerin düzenlenmesi.	Veri toplama ve düzenleme süreciyle ilgili olarak seriler kavramını ve türlerini bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Verilerin basit, tasnif edilmiş ve sınıflanmış seri biçiminde düzenlenmesi	Veri toplama süreciyle elde edilen verilerin seriler ile nasıl düzenleneceğini bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Verilerin zaman ve mekân serisi olarak toplanması ve düzenlenmesi.	Zaman ve mekân serilerini bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar ve veriler.
Bileşik seriler ve panel veriler.	Bileşik serilerin ve panel verilerin nasıl düzenlendiğini ve ne olduğunu bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar, veriler ve çeşitli kurumlar tarafından oluşturulmuş istatistik tablolar.
Grafikler	Verilerin grafiklerle nasıl sunulacağını ve hangi seri türünde hangi grafiğin kullanılacağını bilir.	Temel düzeyde veri toplama, düzenleme ve sunumunu konu alan istatistik kitaplar.

Anahtar Kavramlar

- Verilerin düzenlenmesinde seriler: Veri toplama süreciyle elde edilen ham verilerin basit, tasnif edilmiş ve sınıflanmış seri şeklinde düzenlenmesi.
- Zaman ve mekân serileri: Veri toplama işleminin zaman ve mekân vasfına göre yapılarak verilerin düzenlenmesi.
- Grafik Türleri: Verinin türüne ve düzenlenme şekline göre kullanılabilecek grafiklerin neler olduğu.

4.1. Seriler

Verilerin düzenlenmesiyle ilgili olarak tasnif ve gruplama işlemlerinden bahsettikten sonra bu bölümde de “dağılımlar” ya da başka deyişle “seriler” konusunda açıklamalar yapmak istiyoruz.

Seriler de toplanan verilerin belirli bir esasa göre düzenlenmesidir ve bu şekilde verilerin anlaşılır ve kullanılabilir hale gelmesi amaçlanmaktadır.

Seriler,

- 1- Zaman serileri
- 2- Mekân serileri
- 3- Panel veriler
- 4- Bileşik seriler
- 5- Bölünme (Frekans) serileri

şeklinde sıralanabilir. Şimdi sırasıyla yukarıda saydığımız dağılımlarla ilgili bilgi verelim.

4.1.1. Zaman Serileri

Verilerin zaman vasfının şıklarına göre düzenlenmesiyle elde edilen serilere **zaman serileri** denir.

Türkiye’nin yıllara göre enflasyon oranı, bir işletmenin aylar itibarıyla satışları, Bodrum’da tatil yapan turist sayısının yıllara göre dağılımı ve Türk Hava Yollarının yıllar itibarıyla taşıdığı yolcu sayısı, günlük altın fiyatları ya da döviz kurları zaman serilerine örnek olarak verilebilir.

Örneklerden de anlaşılacağı gibi, zaman serileri yıllık ve aylık olarak düzenlenebileceği gibi günlük hattâ saatlik olarak da düzenlenebilir. Ayrıca nüfus gibi yıldan yıla çok büyük değişimler göstermeyen veriler de 5 ya da 10 yıllık olarak düzenlenebilir.

Aşağıda zaman serilerine yönelik örnekler verilmiştir:

Tablo 4.1: Gayrı Safi Yurt İçi Hâsıla (2009 yılı fiyatlarıyla)

Yıllar	Gayrı Safi Yurt İçi Hâsıla
2010	1.160.013.978
2011	1.394.477.166
2012	1.569.672.115
2013	1.809.713.087
2014	2.044.465.876
2015	2.337.529.940
Kaynak: TÜİK	

Tablo 4.1 ‘de Türkiye’nin 2009 fiyatlarıyla Gayrı Safi Yurt İçi Hasılası’na ilişkin olarak düzenlenmiş veriler, başka bir deyişle Türkiye’nin Gayrı Safi Milli Hasılasının 2010-2015 yılları dağılımı görülmektedir. Tablodan da kolaylıkla anlaşılacağı gibi verimiz yıllık olarak düzenlenmiştir.

Tablo 4.2: Tüketici Fiyat İndeksi (2016 yılı)

Aylar	Aylık Değişim Oranları
Ocak	1,82
Şubat	-0,02
Mart	-0,04
Nisan	0,78
Mayıs	0,58
Haziran	0,47
Temmuz	1,16
Ağustos	-0,29
Eylül	0,18
Ekim	1,44
Kasım	0,52
Aralık	1,64
Kaynak: TÜİK	

Tablo 4.2’de Türkiye’nin 2016 yılında aylar itibarıyla tüketici fiyat indeksi değerlerine ilişkin aylık değişim oranlarını görmektedir. Bu değerler 2016 yılı aylık enflasyon değerlerini göstermektedir. Tablodan da kolayca anlaşılabilceği gibi aylık esasta düzenlenmiş bir veri söz konusudur.

Tablo 4.3: 24 Ayar Külçe Altın Gram Fiyatı

Günler	Altın Satış Fiyatı
12 Haziran 2017	144,47
13 Haziran 2017	143,29
14 Haziran 2017	143,54
15 Haziran 2017	141,68
16 Haziran 2017	141,92
Kaynak: http://www.altinpiyasa.com/arsiv Çevrimiçi: 30.06.2017	

Tablo 4.3’de 26-29 Haziran 2017 tarihleri arasında 24 ayar külçe altın gram fiyatına ilişkin altın satış değerlerini görmektedir. Burada ise günlük esasta düzenlenmiş bir veri ile karşı karşıyayız.

4.1.2. Mekân Serileri

Verilerin mekân vasfının şıklarına göre düzenlendiği verilere mekân serileri denmektedir. Ülkelere göre nüfus, illere göre gayri safi milli hâsıla, ilçelere göre nüfus yoğunluğu, İstanbul'da mahallelere göre okul sayısı vb. veriler mekân serilerine örnek olarak verilebilir. Burada zaman sabit kalmakta, verinin toplandığı mekân yani yer değişmektedir.

Tablo 4.4: Türkiye'nin En Çok İhracat Yaptığı 5 Ülke (2016 yılı)

Ülkeler	İhracat değeri (Bin ABD doları)
Almanya	13.999.121
İngiltere	11.686.213
Irak	7.637.154
İtalya	7.580.999
ABD	6.623.368

Tablo 4.4'de 2016 yılında Türkiye'nin en çok ihracat yaptığı 5 ülke ve bin ABD doları cinsinden ihracat değerleri verilmektedir. Tablodan da görüleceği gibi, verinin düzenlenmesinde yıl yani zaman aynı kalmakta, yerler yani ülkeler değişmektedir.

Tablo 4.5: Kişi Başına Milli Gelir Sıralamasında İlk On Ülke (2016 yılı)

Ülkeler	Kişi Başına Milli Gelir (ABD Doları)
Katar	129.726
Lüksemburg	101.936
Singapur	87.082
Kuveyt	71.263
İrlanda	69.374
Norveç	69.296
Birleşik Arap Emirlikleri	67.696
İsviçre	59.375
ABD	57.293
Suudi Arabistan	54.078
Kaynak: http://aa.com.tr/tr/dunya/kisi-basina-dusen-milli-geliri-en-yuksek-ulke-belli-oldu/735249 Çevrimiçi: 29.06.2017.	

Tablo 4.5'te 2016 yılında kişi başına milli gelir sıralamasındaki ilk 10 ülkeyi ve bu ülkelerin kişi başına milli gelir değerlerini görmekteyiz. Bu örnekte de yine zamanın yani yılın sabit olduğunu ve kişi başına milli gelir değeri açısından farklı ülkelere ilişkin verilerin gösterilmekte olduğunu görüyoruz.

Her ne kadar örneklerimiz ülkelere göre düzenlenmiş dağılımları gösterse de; kıta, ülke, şehir, ilçe, köy gibi coğrafi birimler için olduğu gibi; belirli bir mahallede okullara göre öğrenci sayısı, büyük ölçekli şirketlere göre personel sayısı gibi seriler de belirli bir yılı esas alan mekân serilerine örnek olarak verilebilir.

4.1.3. Panel Veriler

Panel veriler hem zaman hem mekân serisi özelliğini aynı anda taşıyan, başka bir deyişle hem zaman hem mekân serisi olma özelliği gösteren serilere denir.

Panel verilere örnek olarak Türkiye'nin 2010-2016 yılları arasında illere göre milli geliri, 2015 ve 2016 yılları itibarıyla Avrupa ülkelerinin kişi başına milli gelir değerleri, son 3 yıl itibarıyla ülkemize gelen yabancı turistlerin geldikleri ülkelere göre dağılımı verilebilir. Burada hem zaman hem mekân özelliği birlikte değişmekte ve veri bu koşullar altında toplanmakta ve düzenlenmektedir.

Aşağıda yer alan Tablo 4.6'de 2016 yılı verilerine göre kişi başına milli gelir sıralamasında ilk 10 ülkenin kişi başına milli gelirleri ve bunun yanında aynı ülkelerin 2015 yılına ilişkin kişi başına milli gelir değerleri verilmektedir.

Söz konusu tabloda iki yıllık bir zaman ve 10 ülke için kişi başına milli gelir değerleri yer almakta olup, hem zaman hem ülkeler (mekân) değişimi bulunmaktadır.

Panel veriler mekân birimi ve zaman aralığı arttığı ölçüde büyük veri matrisleri yaratırlar. Örneğin son 10 yıllık periyotta ülkemizdeki 81 ilin gayri safi milli hâsıla değerlerini düzenlediğimizi varsayalım. 10x81 boyutlarında bir veri matrisi ve dolayısıyla 810 hücreden oluşan bir dağılım ortaya çıkacaktır.

Tablo 4.6: Kişi Başına Milli Gelir Sıralamasında İlk On Ülke (2015 ve 2016 yılları)

Ülkeler	Kişi Başına Milli Gelir (ABD Doları)	
	2015	2016
Katar	132.870	129.726
Lüksemburg	99.505	101.936
Singapur	85.382	87.082
Kuveyt	70.542	71.263
İrlanda	65.806	69.374
Norveç	68.592	69.296
Birleşik Arap Emirlikleri	67.217	67.696
İsviçre	58.647	59.375
ABD	56.084	57.293
Suudi Arabistan	53.802	54.078
Kaynak: http://aa.com.tr/tr/dunya/kisi-basina-dusen-milli-geliri-en-yuksek-ulke-belli-oldu/735249 Çevrimiçi: 29.06.2017.		

4.1.4. Bileşik Seriler

Bileşik seriler ya da bileşik dağılımlar tek bir değişken ya da veri için değil, iki ya da daha çok veri için düzenlenen dağılımlardır. Daha önce zaman serilerine örnek olarak verdiğimiz Türkiye'nin en çok ihracat yaptığı 5 ülke için ithalat değerlerini de tabloya dâhil eder ve düzenlersek bileşik bir dağılım yaratmış oluruz.

Benzer şekilde ülkelerin milli gelirleri ile kişi başına milli gelirleri, ABD doları, Euro ve İngiliz Sterlini döviz kurları zaman sabit kalmak kaydıyla bileşik dağılımlara örnek verilebilir.

Tablo 4.7: Tüketici ve Üretici Fiyat İndeksleri Aylık Değişim Oranları (2016 yılı)

Aylar	Üretici Fiyat İndeksi	Tüketici Fiyat İndeksi
Ocak	0,55	1,82
Şubat	-0,20	-0,02
Mart	0,40	-0,04
Nisan	0,52	0,78
Mayıs	1,48	0,58
Haziran	0,41	0,47
Temmuz	0,21	1,16
Ağustos	0,08	-0,29
Eylül	0,29	0,18
Ekim	0,84	1,44
Kasım	2,00	0,52
Aralık	2,98	1,64
Kaynak: TÜİK		

Tablo 4.7, Türkiye'nin 2016 yılında üretici ve tüketici fiyat indekslerinde gözlenen aylık değişim oranlarını göstermektedir. Tabloda hem üretici hem tüketici fiyat indekslerine ilişkin iki dağılım birlikte düzenlenmiştir ve dolayısıyla bileşik bir dağılım söz konusudur.

Tablo 4.8: 24 Ayar Külçe Altın Gram Fiyatı ve ABD Doları ve Euro Döviz Kurları

Günler	Altın Satış Fiyatı	ABD Doları	Euro
12 Haziran 2017	144,47	3,5254	3,9549
13 Haziran 2017	143,29	3,5232	3,9495
14 Haziran 2017	143,54	3,5199	3,9152
15 Haziran 2017	141,68	3,5093	3,9232
16 Haziran 2017	141,92	3,5187	3,5294
Kaynak: http://www.altinpiyasa.com/arsiv Çevrimiçi: 30.06.2017			
http://www.tcmb.gov.tr/kurlar/201706/Jun_tr.html Çevrimiçi: 30.06.2017			

Tablo 4.8, 12-16 Haziran 2017 tarihleri arasında bir haftalık süreçte, 24 ayar külçe altın gram satış fiyatı ile ABD doları ve Euro döviz kurlarının satış fiyatlarını göstermektedir. Tabloda 3 ayrı değişkenin verileri düzenlenmiştir dolayısıyla yine bir bileşik dağılım söz konusudur.

4.1.5. Bölünme (Frekans) Serileri

Önceki başlıklarda zaman, mekân ve bileşik seriler ile panel verilerden söz ettik. Bu seriler zaman özellikli, mekân özellikli ya da her iki unsuru birlikte içerecek şekilde düzenlenen serilerdir. Bölünme ya da frekans serileri ise zaman ve mekân vasıflarından bağımsız ve başlı başına bir dağılım yaratan serilerdir ve istatistikte oldukça geniş bir kullanım alanına sahiptir.

İstatistiğe giriş dersini alan öğrencilerin vize sonuçlarının dağılımı, İstanbul Üniversitesi İktisat Fakültesine 2017 yılında kayıt yaptıran öğrencilerin mezun oldukları lise türüne göre dağılımı, bir futbol takımının 2016-2017 sezonunda oynadığı maçlardan elde ettiği puanların dağılımı ya da bir bebeğin 0-12 aylık gelişimi boyunca boy ve/veya ağırlık dağılımı bölünme serilerine örnek verilebilir.

Bölünme serileri nicel bir veriye yönelik olabileceği gibi nitel veriye yönelik olarak da düzenlenebilir. Sözelimi mezun olunan lise türüne yönelik bir dağılım nitel bir veriye dayanmaktadır. Benzer şekilde İstatistik dersini alan öğrencilerin cinsiyet dağılımı da yine bir nitel dağılım örneğidir.

Öğrencilerin not dağılımı, zekâ testi skorlarının dağılımı, boy, ağırlık dağılımları ise sayısal bir vasfın şıklarına yönelik bir dağılım gösterdiğinden nicel dağılımlara örnektirler.

Bölünme ya da frekans dağılımları da kendi içinde 3 ana başlıkta değerlendirilmektedir. Şimdi sırasıyla bunları inceleyelim.

4.1.5.1. Basit Seri

Nicel vasıflara yönelik olarak toplanan verilerin küçükten büyüğe doğru sıralanması ile elde edilen basit dağılımlardır. En küçük değerle başlayarak verilerin sıralanması esasına dayanır. Sıralama işleminin genel olarak küçükten büyüğe doğru yapılması söz konusu olsa da büyüktan küçüğe doğru azalan bir dağılım yaratılması da mümkündür.

Amaç, karışık bir veri görüntüsünden küçükten büyüğe ya da büyükten küçüğe doğru sıralanmış bir veri yaratmak ve bir dağılım yaratmaktır.

Basit seriye yönelik örnek olmak üzere 20 öğrencinin katıldığı bir sertifika programı sonrasında uygulanan bir değerlendirme testinden katılımcıların 100 üzerinden aldıkları puanları ele alalım.

Aşağıda verilen Tablo 4.9’da ilk sütun sertifika programına katılan katılımcıların 100 üzerinden değerlendirme puanlarını ikinci sütun ise puanların küçükten büyüğe doğru sıralanması suretiyle oluşturulan basit seriyi göstermektedir.

Tablo 4.9: Bir Sertifika Programına Katılan 20 Katılımcının Değerlendirme Sonuçları

Puanlar	Basit Seri
90	10
80	10
10	20
20	30
100	40
90	50
40	50
10	50
30	60
50	60
80	70
70	70
60	70
50	70
70	80
50	80
70	90
60	90
90	90
70	100

Tablo 4.9’a birlikte bakalım. İlk sütun katılımcıların değerlendirme sonuçlarını sınav kâğıtlarını veriş sırasına göre ya da isim listesine göre göstermektedir. Dolayısıyla burada sınav sonuçları açısından bir düzenleme söz konusu değildir. İstatistik çalışmalarda buna ham veri denilmektedir. Veri geliş sırasına göre ya da bir liste düzeni içinde kayıt sırasına göre hazırlanmıştır.

Bir vize ya da final sınavı sonrasında öğrencilerin sınıf listesine göre ilan edilen sınav sonuçları ham bir veridir. Yukarıdaki örnekte 20 sınav sonucu olduğu için ham veriye bakarak sınavın başarı durumuna yönelik bir bilgi edebiliriz belki. Ancak 20 değil 100, 150 öğrenci olursa ne olur? Bir an düşünelim. 150 öğrencilik bir sınıfın sınav sonuç listesinde, öğrencilerinin isimlerinin karşısında aldıkları puanların yazıldığı sınav sonuç listesinden sınıfın başarısı hakkında ne kadar bilgi sahibi olabiliriz? 150 öğrencinin notları 150 satırlık bir tablo

gibi düşünöldüğünde bu yapıda, notların karışık şekilde yer aldığı bir tablodan ilk bakışta bilgi sahibi olmak mümkün değildir.

Basit seri yaratmak, yani gözlem değeriğini sıralamak en basit veri düzenleme biçimidir. 150 öğrencinin sınav sonuçlarının yer aldığı listedeki notları küçükten büyüğe doğru sıraladığımızda basit seri oluşturmuş oluruz. Tablomuz yine 150 satırdan oluşacak ancak notlar sıralandığı için en düşük notları, en çok alınan notları daha kolay tespit etme imkânı bulacağız.

Tablo 4.9'a dönersek, tablonun ikinci sütununda sertifika programına katılan 20 katılımcının değerlendirme sonuçlarının sıralanmış şeklini yani basit serisini görmekteyiz. Tablodan kolaylıkla, sıfır alan hiç kimsenin olmadığını, iki kişinin 10, bir kişinin 20 aldığını, sadece beş katılımcının 50 puanın altında sonuç aldığını, 15 katılımcının 50'nin üzerinde sınav sonucuna sahip olduğunu, iki kişinin 80, üç kişinin 90 ve bir kişinin 100 aldığını görmekteyiz.

Veriyi basit seri haline getirmek yani sıralamak uygulamada kullanılan en basit veri düzenleme yöntemidir ve gözlem sayımız çok fazla olduğu durumlarda yetersiz kalmaktadır.

Yukarıda verdiğimiz örnekleri düşünelim. 20 öğrenci söz konusu olduğunda sınav sonuçlarını sıralamak yani basit seri oluşturmak oldukça faydalı bir düzenleme iken 150 ya da 200 ya da daha da fazla gözlem içeren serilerde çok da anlamlı bir düzenleme özelliği taşımamaktadır.

Böyle durumlarda ise şimdi açıklayacağımız başkaca veri düzenleme yöntemlerinden yararlanılmaktadır.

4.1.5.2. Tasnif Edilmiş Seri

Verilerin düzenlenmesinde kullanılan tasnif yönteminden daha önce bahsetmiştik. Tasnif işlemi veride yer alan aynı gözlem değeriğinin bir araya toplanması ve sayılması esasına dayanmaktadır. Böylelikle, veri seti içinde kaç değişik gözlem değeri olduğu ve her bir gözlem değeriğinden kaç tane bulunduğı tespit edilmiş olmaktadır.

20 katılımcının bir sertifika programı sonrasında tabi tutulduğu değerlendirme sınavına ilişkin sonuçları basit seri haline getirmiştik. Şimdi de söz konusu sınav sonuçlarını ayrıştırarak kaç değişik sınav sonucu olduğunu ve hangi sonucun kaç kişi tarafından alındığını tespit ederek veriyi tasnif edelim.

Tablo 4.9'dan kolaylıkla izlenebileceğı gibi katılımcılar arasında,

10 alan 2 kişi
20 alan 1 kişi
30 alan 1 kişi
40 alan 1 kişi
50 alan 3 kişi
60 alan 2 kişi
70 alan 4 kişi

80 alan 2 kiři
90 alan 3 kiři ve
100 alan 1 kiři

bulunmaktadır.

Sınav sonuçlarını bu şekilde ayrıştırma ve her sınav sonucundan kaç tane bulunduğunu belirlemekle tasnif işlemini gerçekleştirmiş olduk. Şimdi bu işlemle elde ettiğimiz verimizi bir tablo ile düzenleyelim.

Tablo 4.10: Bir Sertifika Programına Katılan 20 Katılımcının Değerlendirme Sonuçlarından Oluşturulan Tasnif Edilmiş Seri

Puanlar	Frekanslar
10	2
20	1
30	1
40	1
50	3
60	2
70	4
80	2
90	3
100	1
Toplam	20

Tablo 4.10'dan da görüleceği gibi katılımcıların aldıkları puanlar ve hangi puanın kaç katılımcı tarafından alındığı kolaylıkla tespit edilebilmektedir.

Tasnif edilmiş seri de, basit seri gibi veride herhangi bir bilgi kaybına yol açmamaktadır. Dolayısıyla, gerek basit seri gerekse tasnif edilmiş seri, ham veri setini sadece düzenlemeye tabi tutmakta ve ham veri setinde mevcut veriyi ya da bilgiyi tamamıyla korumaktadır.

Tasnif edilmiş seri özellikle verinin çok fazla sayıda gözlemden oluşması durumunda daha faydalı ve tercih edilen bir yöntemdir.

Verinin daha da fazla gözlemden oluşması halinde veriyi düzenleyebilmek için tasnif işleminden yararlanmak da yetersiz kalabilmektedir.

Böyle durumlarda veri setinde yer alan birbirine yakın gözlem değerlerinin gruplanması ya da sınıflanması yoluna gidilmektedir. Sınıflanmış frekans dağılımı yaratma süreci olarak niteleyebileceğimiz veri düzenleme yöntemini açıklayalım şimdi de.

4.1.5.3. Sınıflanmış Frekans Serisi

Sınıflanmış frekans serisi göreceli olarak fazla gözlem değerine sahi verilerin, birbirine yakın gözlem değerlerini homojen sınıflarda toplamak suretiyle gerçekleştirilen bir veri düzenleme yöntemidir.

250 öğrencinin girdiği ve 100 sorudan oluşan bir test sınavı düşünelim. 100 sınav sorusu olduğuna göre her bir soru 1 puan değerinde olacak ve hiçbir soruya doğru cevap veremeyen öğrenci 0, bir tane doğru cevabı olan öğrenci 1, iki tane doğru cevabı olan öğrenci 2, yetmiş beş doğru cevabı olan öğrenci 75, yetmiş altı doğru cevabı olan öğrenci 76, doksan sekiz doğru cevabı olan öğrenci 98, doksan dokuz doğru cevabı olan 99 ve tüm sorulara doğru cevap veren öğrenci 100 almış olacaktır.

Bir an için 250 öğrencinin sınav notlarını küçükten büyüğe doğru sıraladığımızı ve basit seri oluşturduğumuzu düşünelim. Elde ettiğimiz veri 250 satırdan oluşan sıralanmış bir veri olacaktır. Dolayısıyla, sıralanmış olması dışında veriyi değerlendirme ya da veriyle ilgili bilgi sahibi olma anlamında bir katkı söz konusu olamayacaktır. Böyle bir durumda basit seri oluşturmak suretiyle veriyi düzenlemek tercih edilecek bir veri düzenleme yöntemi değildir.

Veriyi, her nottan kaç öğrencinin aldığını tespit etmek suretiyle tasnif yoluna giderek tasnif edilmiş seri yarattığımızı düşünelim. 0-100 aralığında 101 adet değişik notu alt alta sıraladığımızı ve her bir notun karşısına da o notu alan öğrenci sayısını yani frekansı yazdığımızı düşünelim.

101 satırlık bir veri ile karşı karşıyayız demektir. Böyle bir durumda tasnif edilmiş seri oluşturmak da veriyi düzenlemek anlamında tercih edilebilecek bir yöntem olamamaktadır.

Burada basit bir varsayımda bulunmak yoluna giderek, veriyi sınıflama yöntemi benimsenmektedir.

Verileri sınıflamada kullanılan temel varsayım, birbirine yakın gözlem değerlerini aynı sınıf içinde değerlendirmektir. Örneğin, not sınıflarının 0-10, 10-20, 20-30,80-90, 90-100 şeklinde belirlendiğini düşünelim. Bu durumda, 0-10 not aralığında not alan öğrencilerin başarı düzeyinin benzer olduğu kabul edilmektedir. Yine aynı şekilde 20-30 sınıfında ve diğer tüm sınıflarda yer alanlar için de aynı ya da benzer başarı durumunun geçerli olduğu kabul edilmektedir.

Sınıflamanın temelini teşkil eden bu varsayım ile 100 satırdan oluşan bir tasnif edilmiş frekans serisini sadece 10 sınıftan oluşan bir sınıflanmış frekans serisine dönüştürebiliriz.

Böylelikle, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 notları alan öğrenciler ayrı ayrı gösterilmek yerine 0-10 sınıfında not alan öğrenciler olarak gösterilmektedir. Gerçekte de bir öğrencinin 100

üzerinden 3 ya da 4 veya 6 ya da 7 almış olması önemli bir başarı farkı yaratmamakta ve 3, 4, 5, 6, 7 notlarını detaylı bir şekilde ayrı ayrı göstermek yerine 0-10 not sınıfı içinde toplayarak göstermek yaklaşımı benimsenmektedir.

Şüphesiz, sınıflanmış frekans serisi oluşturduğumuzda her bir nottan kaç öğrencinin almış olduğunu görmek artık mümkün olmamaktadır. Dolayısıyla, basit seri ya da tasnif edilmiş frekans serisinden farklı olarak sınıflanmış frekans serisi veride bilgi kaybına yol açmaktadır. 0-10 aralığında kaç öğrencinin not aldığını görmek mümkünken bu aralıkta yer alan notlar için ayrıntılı bir bilgi alınamamaktadır. Sınıflanmış frekans serisi oluşturmada dikkat edilmesi gereken en önemli husus da budur. Yani, veride kabul edilebilir düzeyde bir bilgi kaybına razı olmak suretiyle sınıfları belirlemek ya da sınıflama yoluna gitmek. Zira, sınıflama özünde bir özetleme niteliği taşır ve aşırı detaylardan seriyi arındırmak isterken tüm bilgiyi kaybetmek de söz konusu olabilir.

Örneğin, öğrencilerin notlarını 0-50 ve 50-100 şeklinde iki sınıfta toplamayı benimsemek, veride çok ciddi bilgi kaybına yol açacak bir yaklaşımdır. Zira böyle bir sınıflama 50 alan öğrenci ile 70, 85 ve hatta 99 ve 100 alan öğrencileri aynı sınıfta değerlendirme hatasına düşecektir. Bu noktada sınıfları belirlemede sınıfların homojen verileri bir araya toplayacak şekilde belirlenmesi gerektiği unutulmamalıdır.

Daha önce “Gruplama” başlığı altında bu konuya değinmiştik. Geldiğimiz noktada gruplama başlığı altında verdiğimiz bilgileri bir kez daha gözden geçirmek okuyucu açısından faydalı olacaktır. Tekrara düşmemek adına sınıfların homojen olması konusunu gruplama başlığı altından takip edilmesini öneriyoruz.

Tablo 4.11: Bir Sertifika Programına Katılan 20 Katılımcının Değerlendirme Sonuçlarından Oluşturulan Sınıflanmış Frekans Serisi

Sınıflar	Frekanslar
0-20	2
20-40	2
40-60	4
60-80	6
80-100	6
Toplam	20

Tablo 4.11’de daha önce basit seri ve tasnif edilmiş seri düzenlemesi yaptığımız bir sertifika programına katılan 20 katılımcının değerlendirme sonuçlarından elde edilen sınıflanmış frekans serisi görülmektedir.

Daha önceki basit seri ve tasnif edilmiş frekans serileri ile kıyaslandığında çok daha yalın bir veri düzenlemesi yapılmış olduğu görülmekle beraber veriye yönelik detaylar kaybedilmiş olmaktadır.

Veriyi sınıflanmış frekans serisi haline dönüştürmek sadece veriyi düzenleme amacıyla yapılan bir işlem gibi değerlendirilmemelidir. Bu düzenleme işlemi, veri üzerinde yapılacak çeşitli analizleri kolay yoldan gerçekleştirme amacına da hizmet etmektedir.

250 öğrencinin girdiği test sınavının sonuçları örneğini hatırlayalım. 101 satırlı bir tasnif edilmiş seri ile karşı karşıya olduğumuzu söylemiştik. Sınav sonuçlarının basit şekilde ortalamasını almak istesek bile bunun için uzun uzun işlemler yapmak zorunda kalırız. Oysa, sınıflanmış frekans serisi üzerinden ortalama hesaplamak oldukça basit ve kısa zamanda gerçekleştirilebilir bir çözüm olmaktadır.

Bölünme ya da frekans serilerinde ortalama ve dağılma ölçülerinin nasıl hesaplandığını izleyen bölümde ele alacağız.

Bu noktada sınıflamayla ilgili olarak bazı küçük hatırlatmalar yapmak istiyoruz.

Tablo 4.11’de yer alan sınıflara bakalım.

0-20 sınıfında 2, 20-40 sınıfında 2, 40-60 sınıfında 4, 60-80 sınıfında 6 ve 80-100 sınıfında yine 6 kişinin puan aldığı görülmektedir. Diğer deyişle sertifika programına katılan 20 kişinin söz konusu sınıflara nasıl dağıldığı görülmektedir. Nitekim Frekans dağılımı denmesinin sebebi de serinin 20 kişinin sınıflara hangi frekanslarla dağıldığını gösteriyor olmasındandır.

Daha önce grupta başlığı altında değindiğimiz gibi, sınıfların alt sınır değeri ilgili sınıfa dâhil, üst sınır değeri ise dâhil değildir.

0-20 sınıfını ele alalım. 0 değeri bu sınıfın alt sınır değeridir, 20 değeri de yine bu sınıfın üst sınır değeridir. Dolayısıyla 0-20 sınıfına 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 değerleri dâhildir ancak üst sınır değeri olan 20 kural gereği bu sınıfta değildir ve bir sonraki sınıfta yer almaktadır.

Aynı kural diğer sınıflar için de geçerlidir ve üst sınır değerleri, alt sınır değeri oldukları sınıflarda değerlendirilmekte ve bu sınıflara dâhil edilmektedirler.

80-100 sınıfı son sınıf olması sebebiyle bu konuyla ilgili olarak bir istisnaya sahiptir. Normal koşullar altında 100 değeri 80-100 sınıfı içinde yer alamayacağından, 100 değeri için 100-120 sınıfının açılması gerektiği düşünülebilir. Oysa notlar 0-100 aralığındadır ve 100’den fazla not alınması mümkün değildir.

Dolayısıyla, 100-120 sınıfı açılması gerektiğini düşünürsek gerçekte 100 alan 6 öğrenciyi 100-120 arasında not almış şekilde değerlendirmiş oluruz. En yüksek notun 100

olması mümkünken 100-120 sınıfı açmak, hem tuhaf bir görüntü yaratacak hem de gerçekte 100 alan 6 öğrencinin yanlış şekilde değerlendirilmesine yol açacaktır.

Sözünü ettiğimiz üst sınır değerinin ilgili sınıfa dâhil olmadığı kuralı son sınıf için ihlâl edilmekte ve üst sınır değeri de o sınıf içinde değerlendirilmektedir. Daha açık bir ifadeyle, 80-100 sınıfı 100 değerini de kapsayacak şekilde düşünülmektedir.

Bir sınıfın üst sınır değeri ile alt sınır değeri arasındaki farka sınıf genişliği dendiğini daha önce ifade etmiştik.

Burada bir kavramdan daha söz etmek istiyoruz: Sınıf orta noktası.

Sınıf orta noktası, o sınıf içinde yer alan tüm birimlerin eşit olduğu varsayılan değeridir. Oldukça basit bir formülle hesaplanmakta olan sınıf orta noktası m harfi ile gösterilmektedir.

$$\text{Sınıf Orta Noktası} = m = (\text{Üst sınır değeri} + \text{Alt sınır değeri}) / 2$$

$$0-20 \text{ sınıfının sınıf orta noktası } (0 + 20) / 2 = 10$$

$$20-40 \text{ sınıfının sınıf orta noktası } (20 + 40) / 2 = 30$$

şeklinde hesaplanmaktadır.

Tablo 4.12’de sınıf orta noktalarının da gösterildiği sınıflanmış frekans dağılımı gösterilmektedir.

Sınıf orta noktası kavramı özellikle, sınıflanmış frekans serilerinde ortalama ve dağılma ölçüleri hesaplayabilmek açısından önemli bir kavramdır ve ilerleyen bölümde bu konu tekrar ele alınacaktır.

Tablo 4.12: Bir Sertifika Programına Katılan 20 Katılımcının Değerlendirme Sonuçlarından Oluşturulan Sınıflanmış Frekans Serisi

Sınıflar	Sınıf Orta Noktaları (m)	Frekanslar
0-20	10	2
20-40	30	2
40-60	50	4
60-80	70	6
80-100	90	6
Toplam		20

4.2. Sınıf Genişliğini Tayin Etmede Sturge Kuralı

Sınıflanmış frekans dağılımı oluşturmak istediğimizde dikkat edilmesi gerek en önemli husus sınıfların belirlenmesidir. Bu noktada verinin büyüklüğü, yani gözlem sayısı ve verinin niteliği önem taşımakta ve mümkün olduğunca homojen sınıflar oluşturmak amaçlanmaktadır.

Bu noktada verinin yapısını dikkate alarak sınıf genişliğinin ne olması gerektiğine yönelik fikir vermek üzere Sturge Kuralı geliştirilmiştir.

Sturge Kuralı;

$$S = (X_{\text{maks}} - X_{\text{min}}) / (1 + 3,322 \log n)$$

formülü ile hesaplanmaktadır.

Formülde yer alan,

X_{maks} değeri, serideki en büyük gözlem değerini,

X_{min} değeri, serideki en küçük gözlem değerini,

n ise gözlem sayısını

ifade etmektedir.

Sturge Kuralı bize sınıf genişliğinin ne olması gerektiği konusunda bilgi vermektedir. Araştırmacının bu öneriyi dikkate alarak buna uygun bir sınıf genişliği tayin etmesi uygun olacaktır.

Sturge Kuralının nasıl uygulandığını bir örnek yardımıyla görelim.

Örnek:

100 işyerinde kaç personel istihdam edildiği tespit edilmiş ve gözlem sonuçları aşağıdaki tablo ile düzenlenmiş olsun.

Tablo 4.13: 100 İşyerinde Çalışan Personel Sayılarının Dağılımı

Personel Sayısı	Frekans
20	3
25	2
30	8
35	16
40	13
45	18
50	12
55	11
65	10
70	7
Toplam	100

Sturge Kuralını uygulayarak bu dağılım için sınıf genişliğinin ne olması gerektiğini belirleyelim:

$$S = (X_{\text{maks}} - X_{\text{min}}) / (1 + 3,322 \log n)$$

$$S = (70 - 20) / (1 + 3,322 \log 100)$$

$$S = 6,54 \cong 7$$

Sturge Kuralını uyguladığımızda örnek verimiz için sınıf genişliğinin 7 alınması gerektiği ortaya çıkıyor.

Şimdi sınıf genişliklerini 7 olacak şekilde düzenleyerek sınıflanmış frekans serimizi oluşturalım.

Tablo 4.14: 100 İşyerinde Çalışan Personel Sayısının Sınıflanmış Frekans Dağılımı

Sınıflar	Frekans
20-27	5
27-34	8
34-41	29
41-48	18
48-55	12
55-62	11
62-69	10
69-76	7
Toplam	100

Tablo 4.14’de Sturge Kuralı uygulayarak elde ettiğimiz 7 birimlik sınıf genişliklerine göre düzenlediğimiz sınıflanmış frekans tablosu görülmektedir.

Sturge Kuralını sınıf genişliğini tespit etmede bir yol gösterici olarak kabul etmek ve bu öneriyi dikkate alarak sınıf genişliğini tayin etme kararını araştırmacı vermelidir.

Örneğimizde sınıf genişliğinin Sturge Kuralına göre 7 olarak belirlenmesi neticesinde, sınıfların sınır değerlerinin ne olması gerektiğinin tespiti bu kurala tamamen uyularak yapılmıştır.

Ancak uygulamaya baktığımızda, genel olarak sınıf sınır değerlerinin sonu beş ya da onun katları şeklinde alınmasının tercih edildiğini görüyoruz. Bunun en önemli sebebi beş veya onun katlarının daha anlaşılır ve akılda kalıcı değerler olmalarıdır.

Dolayısıyla, sınıf genişliği olarak Sturge Kuralı 7 birimlik bir genişlik önerse de araştırmacı bu öneriyi dikkate almak ve önerilen değerden çok da uzaklaşmamak kaydıyla bir sınıf genişliği belirleyebilir.

Örneğin yukarıdaki örneğimizde sınıf genişliğini 10 olarak belirlemek suretiyle hem Sturge Kuralının önerisi olan 7 birimlik genişlik önerisinden çok uzaklaşılmamış ve hem de onun katları şeklinde ortaya çıkacak olan sınıf sınırları ile daha anlaşılır, akılda kalıcı alt ve üst sınır değerleri belirlenmiş olur.

4.3. Verinin Grafik Yardımıyla Sunulması

Veri toplama süreci ile ulaşılan bilginin açık ve anlaşılır bir biçimde sunulması için düzenlenmesi gerekmektedir. Bu süreçte, toplanan verinin tablo ile özetlenmesinin yanı sıra

grafikler de kullanılmaktadır. Grafik kullanımı ile amaçlanan, toplanan verinin dağılımının görsel olarak ortaya konulmasıdır. Ancak bu aşamada, vasıf ve şık özellikleri dikkate alınarak uygun grafik türünün tercih edilmesi önem arz etmektedir. Ayrıca bir grafikte yatay ve dikey eksenlerin tanımlanması ve grafiğin başlığının olması gerekmektedir. Bu kısımda nitel ve nicel verilerin grafikleri ele alınacaktır.

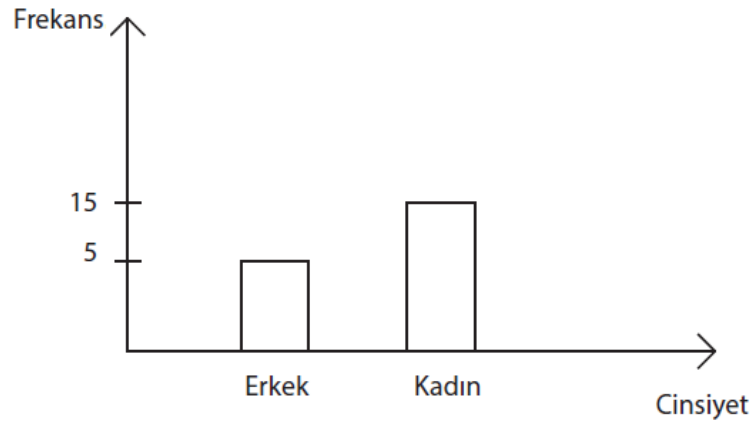
4.3.1. Nitel Veriler için Grafikler

Bu başlık altında sınıflama (nominal) ve sıralama (ordinal) ölçüm düzeyi ile ölçülmüş nitel vasa ait verinin grafiğinin çizilmesi üzerinde durulacak olup; sırasıyla **çubuk grafik**, **daire dilimleri grafiği** ve **bindirmeli grafik** tanıtılacaktır.

Çubuk Grafik

Nitel vafın şıklarına ilişkin sıklık (frekans) ya da yüzde sıklık (% frekans) değerlerinin ayrı ayrı çubuklarla gösterildiği grafik türüdür. Bu grafik türünü örneklendirmek için nitel vasıflardan biri olan cinsiyet özelliğini ele alalım. Bilindiği üzere; bu vafın şıkları kadın ve erkektir. Bu grafiği çizerken; yatay ve dikey eksen isimlerini belirlememiz gerekmektedir. Şıklara ait frekansların dikey ekseninde, vafın şıklarının ise yatay ekseninde olduğunu varsayalım.

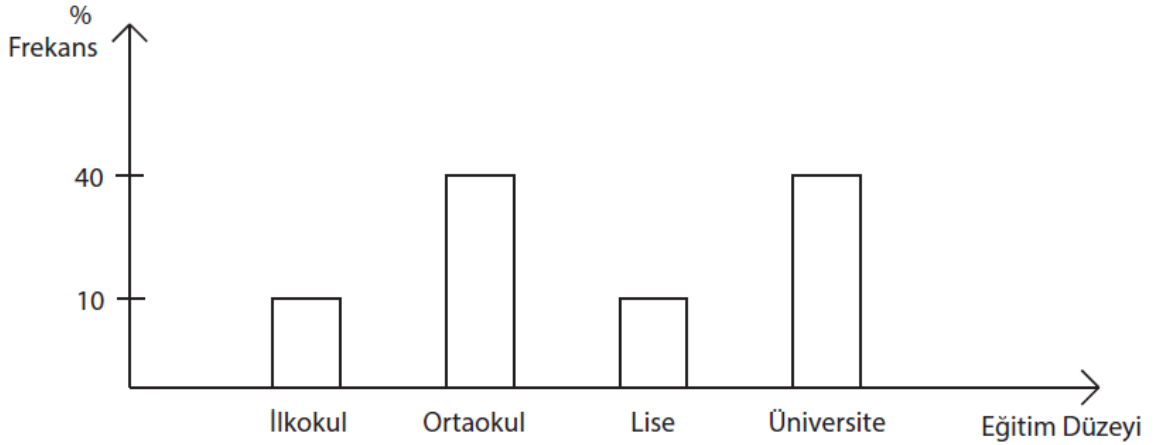
Grafik 4.1: Öğrencilerin Cinsiyet Özelliğine Göre Dağılımı



Grafik 4.1. incelendiğinde, topluluktaki kişilerin 15'ini kadınların oluşturduğu gözlenmiştir. Cinsiyet vafı, sınıflama (nominal) ölçüm düzeyi ile ölçüldüğünden şıkların sırası önem arz etmemektedir. Ayrıca dikey ekseninde frekans ya da yüzde frekansın kullanılması grafiğin taşıdığı bilgi miktarını ve şeklini değiştirmeyecektir.

Sıralama (ordinal) ölçüm düzeyi ile toplanan veri için örnek vermek gerekirse; bir hastanenin dâhiliye kliniğine gelen hastaların eğitim düzeylerini (ilkokul, ortaokul, lise, üniversite) ele alalım. 12 Temmuz 2017 tarihinde bu polikliniği ziyaret eden 50 kişi olduğunu varsayalım. Grafik çizimi aşamasında geçtiğimizde, öncelikle grafiğin dikey ve yatay eksenlerinin tanımlanması gerekmektedir. Dikey eksen şıkların yüzde frekansını, yatay eksen ise şıkları göstermektedir.

Grafik 4.2: Dahiliye Polikliniğini Gelen Hastaların Eğitim Düzeyine Göre Dağılımı



Grafik 4.2.'de görüldüğü üzere; grafiğin yatay ekseninde yer alan şıklar, bir sıra dâhilinde yazılmıştır. Bu grafik için şıkların sıralaması önem arz etmektedir. Çünkü veri, sıralama (ordinal) ölçüm düzeyi ile toplanmıştır.

Daire Dilimleri Grafiği

Bu grafik türünde, nitel vasa ilişkin şıklar bir dairenin dilimleri şeklinde sunulmaktadır. 2017-2018 eğitim-öğretim yılında İstanbul Üniversitesi İktisat Fakültesi'ne kayıt olan öğrencilerin mezun oldukları lise türüne göre dağılımı, bir hastanenin cildiye polikliniğine başvuran hastaların medeni durum özelliğine göre dağılımı, İstanbul Üniversitesi İletişim Fakültesi'ne kayıtlı olan öğrencilerin bölümlere göre dağılımı bu grafik türü ile özetlenebilmektedir.

İletişim Fakültesi'nde *Gazetecilik, Halkla İlişkiler ve Tanıtım, Radyo, Televizyon ve Sinema* olmak üzere üç bölüm bulunmaktadır. Bu bölümler göre öğrencilerin dağılımı Tablo 4.15.'deki gibi olsun:

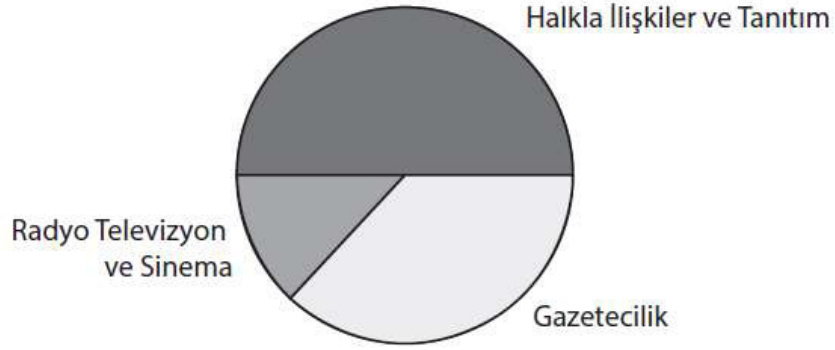
Tablo 4.15: İletişim Fakültesi Öğrencilerinin Bölümlere Göre Dağılımı

Bölüm	Öğrenci Sayısı	% Öğrenci Sayısı
Gazetecilik	150	30
Halkla İlişkiler ve Tanıtım	300	60
Radyo, Televizyon ve Sinema	50	10

Tablo 4.15'ten hareketle; daire dilimleri grafiğini çizebilmek için her bir bölümdeki öğrenci sayısına veya oranına karşılık gelen açılar hesaplanmıştır. Bu aşamada, %100 değeri 360 dereceye karşılık gelirse, %30'a karşılık gelen açının derecesi basit orantıyla bulunmuştur. Yani $(30/100) \times 360 = 108$ derece, $(60/100) \times 360 = 216$ derece, $(10/100) \times 360 = 36$ derece olarak

hesaplanmıştır. Bu değerlerden hareketle daire dilim grafiği çizilmiş ve Grafik 4.3'te sunulmuştur.

Grafik 4.3: İletişim Fakültesi Öğrencilerinin Bölümlere Göre Dağılımı



Grafik 4.3 incelendiğinde; bölümlere göre öğrencilerin dağılımının yapıldığı görülmüştür.

Bindirmeli Grafik

Nitel vasfa ait şıkların dağılımını göstermek amacıyla kullanılan bir diğer grafik türü ise bindirmeli grafikdir. Dolayısıyla bahsedilen örnekler, bu grafik türü ile de çizilebilmektedir. Yine İletişim Fakültesi öğrencilerinin bölümlere göre dağılımını ele alırsak; dikey eksen yüzde frekansı, yatay eksen ise bölümlerin olduğu Grafik 4.4'te sunulmuştur.

Grafik 4.4: Bölümlere Göre Öğrencilerin Dağılımı



Grafik 4.4.'te görüldüğü üzere; dikey eksen yüzdesel olarak 0'dan 100'e kadar bölünmüştür. Her bir şıkkın toplam içindeki payı üst üste gelecek şekilde tanımlanmıştır. Dolayısıyla her bir şık, toplam içindeki yüzdesel payı kadar alan kaplamıştır.

4.3.2. Nicel Veriler için Grafikler

Bu başlık altında aralık ve oran ölçüm düzeyi ile ölçülmüş nicel vasfa ait verinin grafiğinin çizilmesi üzerinde durulacak olup; sırasıyla **kartezyen grafik**, **histogram** ve **dağılım poligonu**, **çizgi grafiği** ve **serpilme diyagramı** tanıtılacaktır.

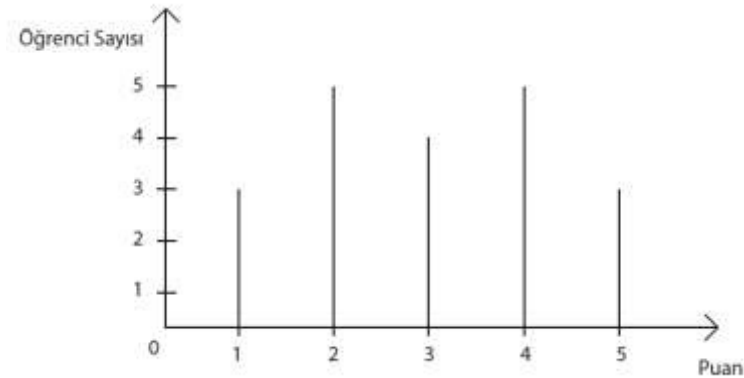
Kartezyen Grafik

Tasnif edilmiş seriler, bu grafik türü ile özetlenebilmektedir. Bu grafik türünü bir örnek yardımıyla açıklayalım. İstatistik dersini seçen 20 öğrencinin olduğunu ve öğrencilerin 1 ile 5 puan arasında değerlendirildiğini varsayalım. Bu öğrencilerin sınav puanına göre dağılımı Tablo 4.16’da sunulmuştur.

Tablo 4.16: İstatistik Dersini Seçen Öğrencilerin Puan Dağılımı	
Puan	Öğrenci Sayısı
1	3
2	5
3	4
4	5
5	3

Tablo 4.16.’da yer alan bilgidен hareketle; dikey eksenin öğrenci sayısını, yatay eksenin ise sınav puanını gösterdiği bir grafik çizilmiştir. Bu grafikte, puan dağılımına göre öğrenci sayısının görülmesi amaçlanmış olup; Grafik 4.5. çizilmiştir.

Grafik 4.5: Puanlara Göre Öğrenci Dağılımı



Grafik 4.5. incelendiğinde; puanlara göre öğrenci dağılımı *kartezyen grafik* ile gösterilmiştir.

Histogram

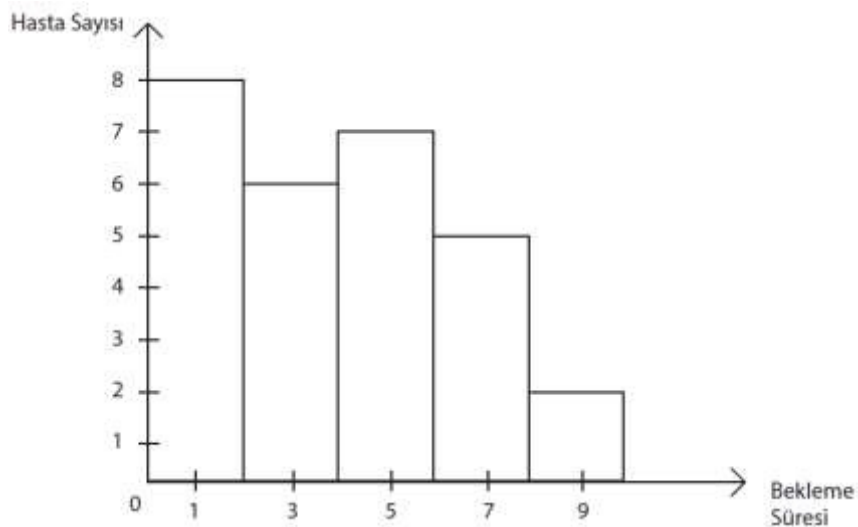
Sınıflanmış seriler de tasnif edilmiş seriler gibi kartezyen grafik ile gösterilmektedir. Ancak sınıflanmış serilerde sütunlar kullanılmaktadır. Histogram olarak adlandırılan bu grafikte ise sütunların alanı frekansları göstermektedir. Sınıf aralıklarının eşit olması durumunda, sütunların taban uzunlukları aynı olmaktadır. Böylelikle, her bir sınıfa karşılık gelen sütun alanı o sınıfa karşılık gelen sıklık ya da yüzde sıklığı göstermektedir. Bu grafik türünü, bir örnek yardımıyla açıklayalım. Bir hastanenin acil servisine 20 başvurunun olduğu ve başvurulara kaç dakika sonra müdahale edildiğine dair veri, Tablo 4.17’de verilmiştir.

Tablo 4.17: Acil Serviste Hastaların Bekleme Süresi

Bekleme Süresi (dakika)	Hasta Sayısı
0-2	8
2-4	6
4-6	7
6-8	5
8-10	2

Tablo 4.17.’den hareketle, dikey ekseninde hasta sayısının yatay ekseninde ise bekleme süresinin yer aldığı histogram çizilmiş olup; Grafik 4.6’da gösterilmiştir.

Grafik 4.6: Hastaların Bekleme Süresine Göre Dağılımı



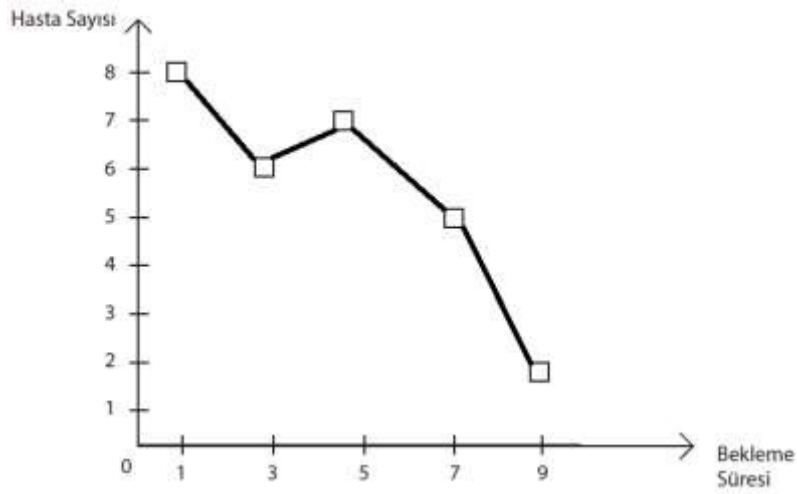
Grafik 4.6. incelendiğinde; sütunların taban uzunluklarının eşit olduğu görülmektedir. Çünkü sınıf aralıkları eşittir. Bu grafikte dikey eksenin frekans ya da yüzde frekans olarak

düzenlenmiş olması grafiğin şeklini değiştirmemektedir. Ayrıca yatay eksende sütunların alt ve üst sınır değerleri yazılabileceği gibi bu sınıfları temsil eden sınıf orta noktası (m) de yazılabilir. Grafik 4.6.’da yatay eksende sınıf orta noktalarının yazıldığı görülmektedir.

Dağılım Poligonu

Histogram çizildikten sonra sütunların üstlerindeki sınıf orta noktaları birleştirilerek dağılım poligonu çizilmektedir. Bu çizgiler, histogram üzerinde gösterilebileceği gibi, sütunlar silinerek de gösterilebilmektedir. Bu grafik için örnek çizim yapılırken, histogram başlığındaki örnekten yararlanılmıştır. Yine dikey eksende hasta sayısı, yatay eksende ise bekleme süresi bulunmakta olup; Grafik 4.7’de gösterilmiştir.

Grafik 4.7: Hastaların Bekleme Süresine Göre Dağılımı



Grafik 4.7. incelendiğinde; Grafik 4.6.’daki histogramdaki sütun çizgilerinin silinmiş olduğu görülmektedir. Nicel vasfın şık sayısı çok olduğunda, sınıf orta noktalarının birleştirilmesiyle elde edilen dağılım poligonunun kullanılması daha açık ve kolay anlaşılırdır.

Çizgi Grafiği

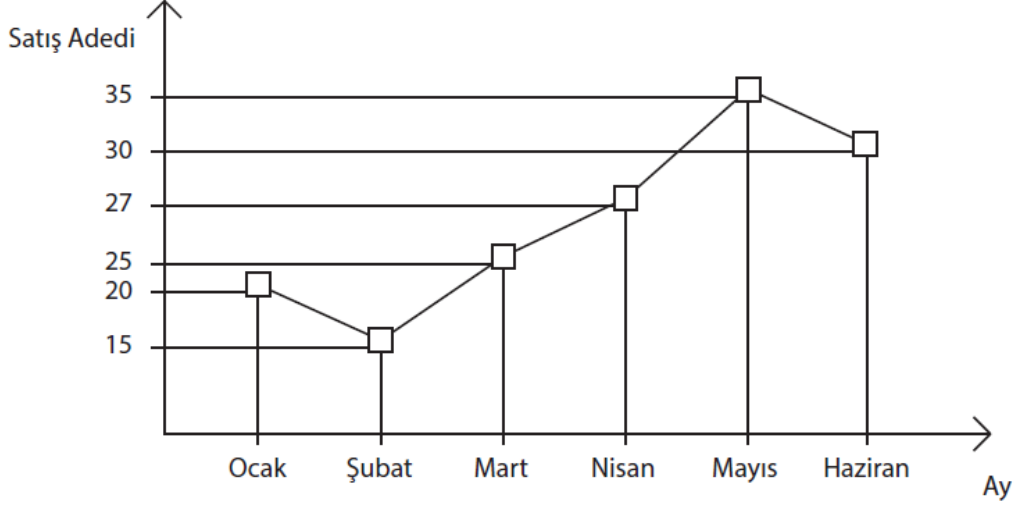
İncelenen vasfın zaman içinde gösterdiği değişimi göstermek için çizgi grafiği kullanılmaktadır. Bu grafik türünde, yatay eksen zaman dilimini (gün, hafta, ay, yıl gibi) ve dikey eksen ise vasfın şıklarını göstermektedir. Her iki eksenin kesiştiği nokta, başlangıç noktası olan sıfırdır. Bu grafik türüne örnek olarak, bir cep telefonu mağazasının aylar itibarıyla sattığı cep telefonu adedi verilebilmektedir. Tablo 4.18’de bu mağazanın ilk altı ay için satış adetleri verilmiştir.

Tablo 4.18: Aylara Göre Cep Telefonu Satışları

Ay	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran
Satış Adedi	20	15	25	27	35	30

Tablo 4.18.'de yer alan bilgilerden hareketle, çizgi grafiği çizilmiştir. Çizgi grafiği, ilgili aylardaki satış adetlerinin çizgi ile birleştirilmesi ile elde edilmiş olup Grafik 4.8.'de gösterilmiştir.

Grafik 4.8: Aylara Göre Cep Telefonu Satışları



Serpilme Diyagramı

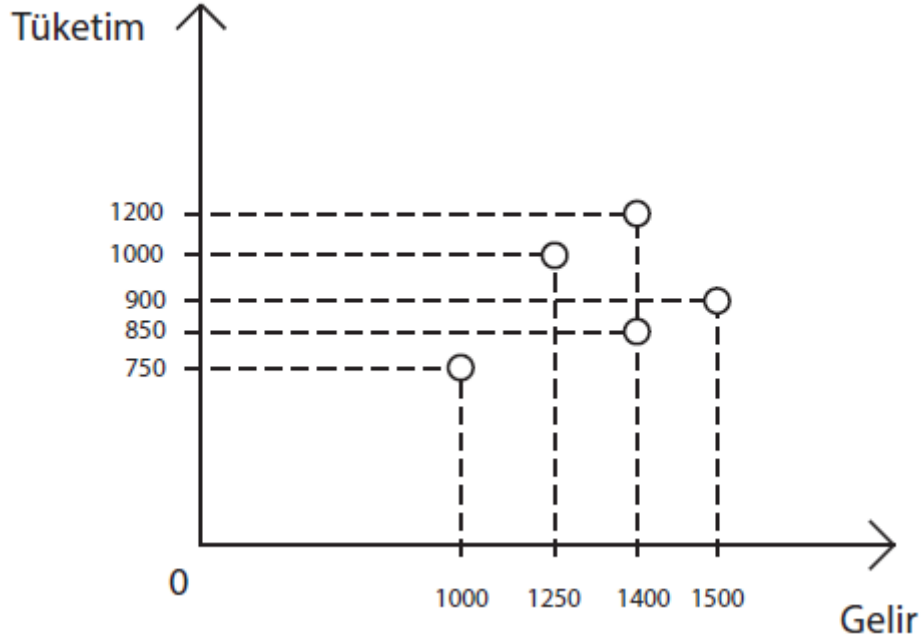
Serpilme diyagramı kısmına kadar anlatılan grafikler, vasıf sayısının bir olması durumunda kullanılmaktadır. Birimlerin iki özelliği gösterilmek istendiğinde, serpilme diyagramında yararlanılmaktadır. Bu grafik türünün çizilmesinde, ilgilenilen özelliklerden (vasıflardan) biri yatay eksen, diğer ise dikey eksen yer almaktadır. Örneğin, bireylerin gelir ve tüketim vasıfları incelenmiş ve Tablo 4.19'da sunulmuştur.

Tablo 4.19: Gelir ve Tüketim Vasıfları Açısından Bireylerin Değerleri

	Gelir (lira)	Tüketim (lira)
Ali	1000	750
Melek	1400	1200
Doğan	1250	1000
Mehmet	1400	850
Dilek	1500	900

Tablo 4.19. incelendiğinde, her bir birey için iki değer verildiği gözlenmiştir. Dolayısıyla dikey eksene tüketim, yatay eksene gelir vasfı konularak grafik çizilip; her bir bireye ait değer bu grafikte nokta olarak gösterildiğinde serpilme diyagramı tamamlanmış olacaktır. Grafik 4.9'da serpilme diyagramı gösterilmiştir.

Grafik 4.9: Gelir ve Tüketim Değişkenlerine Ait Serpilme Diyagramı



Grafik 4.9.'da görüldüğü üzere; bireylerin gelir ve tüketim değerlerinin kesiştiği noktalar işaretlenmiştir. İstenirse, bu noktalar çizgi kullanılarak birleştirilebilir.

Bu Bölümde Neler Öğrendik?

Bu bölümde; toplanan verilerin ham veri olarak nitelendiğini ve gerek istatistik analizler yapabilmek gerekse verilerin ilgililerce anlaşılır halde sunulabilmesi için kullanılabilecek veri düzenleme yöntemlerini öğrendik.

Verilerin düzenlenmesiyle ilgili olarak, zaman serisi, mekân serisi, bileşik seriler, panel veriler ve bölünme serilerini örneklerle inceleyerek, basit seri, tasnif edilmiş seri ve sınıflanmış seriler üzerinde ayrıntılı bilgiler verdik.

Son olarak, toplanan verilerin görsel olarak sunulması amacıyla kullanılan grafiklerden söz ederek çeşitli grafik yöntemlerinden örnekler verdik.

Bölüm Soruları

- 1- Sınıflanmış serilerin grafik gösterimine..... denir.
 - a- Çubuk grafik
 - b- Histogram
 - c- Serpilme diyagramı
 - d- Daire dilimleri
 - e- Bindirmeli grafik
- 2- Verilerin, birbirine yakın şıklarını mümkün olduğunca homojen sınıflara toplamak suretiyle oluşturulan seriyedenir.
 - a- Sınıflanmış seri
 - b- Tasnif edilmiş seri
 - c- Sıralama
 - d- Düzenleme
 - e- Ayırma
- 3- Zaman vasfının şıklarına göre düzenlenen serilere denir.
 - a- Mekân serileri
 - b- Bölünme serileri
 - c- Panel veriler
 - d- Zaman serileri
 - e- Bileşik seriler
- 4- Mekân vasfının şıklarına göre düzenlenen serileredenir.
 - a- Mekân serileri
 - b- Bölünme serileri
 - c- Panel veriler
 - d- Zaman serileri
 - e- Bileşik seriler

- 5- Hem zaman ve hem mekân vasfına göre düzenlene serileredenir.
- a- Mekân serileri
 - b- Bölünme serileri
 - c- Panel veriler
 - d- Zaman serileri
 - e- Bileşik seriler
- 6- Bir gözlem değerinin seri içinde kaç kez tekrar ettiğini gösteren değeredenir.
- a- Frekans
 - b- Tasnif
 - c- Birim
 - d- Sınıf
 - e- Grup
- 7- Gözlem değerlerini küçükten büyüğe doğru sıralayarak yaratılan seriye.....denir.
- a- Tasnif edilmiş seri
 - b- Basit seri
 - c- Sınıflanmış seri
 - d- Gruplanmış seri
 - e- Bileşik seri
- 8- Bir sınıfın alt sınırı ile üst sınırının toplanarak ikiye bölünmesi ile elde edilen ve sınıfı temsil eden değere.....denir.
- a- Sınıf üst sınırı
 - b- Sınıf alt sınırı
 - c- Sınıf genişliği
 - d- Sınıf orta noktası
 - e- Sınıf aralığı

- 9- Bir sınıfın üst sınırı ile alt sınırı arasındaki farkadenir.
- a- Sınıf üst sınırı
 - b- Sınıf alt sınırı
 - c- Sınıf genişliği
 - d- Sınıf orta noktası
 - e- Sınıf sayısı
- 10- İki ya da daha fazla veri setini bir arada düzenleyen serilere.....denir.
- a- Mekân serileri
 - b- Bölünme serileri
 - c- Panel veriler
 - d- Zaman serileri
 - e- Bileşik seriler

Cevaplar

- 1- b
- 2- a
- 3- d
- 4- a
- 5- c
- 6- a
- 7- b
- 8- d
- 9- c
- 10- e

Yararlanılan Kaynaklar

ALOBİ, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), "On the Theory of Scales of Measurement" , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

5. ANALİTİK ORTALAMALAR – ARİTMETİK ORTALAMA

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Daha önceki bölümlerde verilerin toplanmasını ve düzenlenerek sunulmasını öğrendik.

Bu bölümde de toplanmış ve düzenlenmiş verilerin, verilerden elde edilen bilgilerin özetlenmesini öğreneceğiz.

Bu kapsamda, çok sayıda veriyi özetlemek amacıyla kullanılabilecek ortalamaları analitik ve analitik olmayan başlığı altında öğrenecek ve analitik ortalamalar içinde en yaygın kullanım alanına sahip olan aritmetik ortalamanın nasıl hesaplandığını öğreneceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Çok sayıda birimden oluşan bir veri setini nasıl akılda kalıcı tek bir değerle ifade edebiliriz?
- 2) Çok sayıda birimden oluşan veri setlerini birbiriyle karşılaştırmak nasıl mümkün olabilir?
- 3) Analitik ortalama nedir ve hangi ortalamalar bu grupta değerlendirilir?
- 4) Analitik olmayan ortalama nedir ve hangi ortalamalar analitik olmayan ortalama grubunda yer alır?
- 5) Bir veri setini özetlemek için sadece ortalama yeterli midir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Verileri özetlemek neden önemli?	Toplanan verileri düzenledikten sonra, verilerden sonuç çıkarmanın ilk aşaması verileri özetlemektir. Verilerin nasıl özetleneceğinin ve bu işleme neden gerek duyulduğunu bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından yararlanılabilir.
Verileri özetlemek amacıyla kullanılan analitik ve analitik olmayan ortalamalar	Analitik ve analitik olmayan ortalamaların genel özellikleri ve türlerini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından yararlanılabilir.
Verileri özetlemede kullanılan analitik ortalamaların ayrıntılı olarak ele alınması ve en önemli analitik ortalama olarak aritmetik ortalama.	Uygulamada en çok kullanılan analitik ortalamaları ve özel olarak aritmetik ortalamayı ayrıntılı olarak öğrenerek, aritmetik ortalamanın dikkatli kullanılması gerektiği durumlar konusunda da bilgi sahibidir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Ağırlıklı aritmetik ortalama	Ağırlıklı aritmetik ortalamaya neden ihtiyaç duyulduğunu ve nasıl hesaplandığını bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Aritmetik ortalamanın özellikleri	En önemli ortalama olan aritmetik ortalamanın özelliklerini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.

Anahtar Kavramlar

- Ortalama: Verileri özetleme, tek bir değer olarak nitelendirme ve akılda kalıcı bir değer yaratma amacıyla kullanılan temel istatistik ölçüler.
- Analitik ortalamalar: Ortalama hesabında tüm gözlem değerlerinin kullanılmasıyla elde edilen ve duyarlı ortalamalar olarak adlandırılan ortalamaların genel adı.
- Analitik olmayan ortalamalar: ortalama hesabında tüm gözlem değerlerinin değil, bazı gözlem değerlerinin kullanıldığı ve duyarsız ortalamalar olarak adlandırılan ortalamaların genel adı.
- Aritmetik ortalama: En yaygın kullanılan analitik ortalama.

Giriş

Daha önceki bölümlerde veri toplama aşamasından bahsettik. Bu kapsamda temel kavramlardan söz ettikten sonra veri toplama yöntemlerini, veri toplama aşamasında yapılabilecek hataları ve dikkat edilmesi gereken hususları inceledik. Daha sonra, ham veri olarak nitelenen toplanmış verilerin nasıl düzenlendiğini ve istatistik analizler yapılmaya nasıl uygun hale getirildiğini öğrendik.

Bu bölümde çok sayıda gözlem sonucundan oluşan bir verinin nasıl özetleneceğini öğreneceğiz.

Basit bir örnek yardımıyla veri özetleme işleminin ne amaçla ve nasıl yapıldığını açıklamaya çalışalım:

100'er kişiden oluşan iki sınıfa aynı içerikte İstatistik dersi anlattığımızı ve dönem sonunda öğrencileri, final sınavına alarak, başarılarını sınav sonucu ile ölçtüğümüzü düşünelim. Sınıfların dönem sonu başarılarını kıyaslamak istersek yüzer öğrenciden oluşan sınıfların notlarını tek tek karşılaştırmamız düşünülemez.

Bu durumda hangi sınıfta kaç kişinin tam not yani 100 aldığını tespit ederek karşılaştırmak akla gelebilir. Hangi sınıfta 100 alan öğrenci sayısı fazlaysa o sınıfı daha başarılı kabul etmeyi düşünebiliriz belki. Bir an için bu yaklaşımı benimsediğimizi farz edelim.

Şimdi bir sınıfta dört tane 100 alan öğrenci olduğunu ve kalan 96 öğrencinin 50'nin altında not aldığını, buna karşılık diğer sınıfta iki öğrencinin 100 aldığını ve kalan 98 öğrencinin 50'nin üzerinde notlar aldığını varsayalım.

Verdiğimiz örnek oldukça uç bir örnek olmakla beraber durumu açık bir şekilde ortaya koymaktadır. 100 alan öğrenci sayısını başarı kriteri olarak ele aldığımızda, dört öğrencinin 100 aldığı ve geriye kalan 96 öğrencinin 50'nin altında not alarak başarısız olduğu sınıfı daha başarılı olarak değerlendirmiş oluyoruz. Oysa diğer sınıfta iki öğrenci 100 almış olmasına rağmen tüm öğrenciler başarılı olmuştur.

Dolayısıyla, hangi notu kaç kişinin aldığı tespit edilerek yapılacak kıyaslamalar bizi yanıltabilir.

Gerek iki veri setine yönelik herhangi bir vasfı kıyaslamak gerekse bir veri setini özetleyecek ve/veya o veriyle ilgili hatırdaki kalacak ve o veriyi niteleyecek tek bir değer belirlemek istendiğinde ortalamalardan yararlanılır.

Ortalamalar, merkezi eğilim ölçüleri olarak ele alınır. Başka bir deyişle bir veri seti içinde yer alan verilerin hangi değer etrafında toplandığını ifade eder.

Ortalamalar aynı zamanda bir veri seti içindeki küçük ve büyük değerleri yumuşatarak serinin genel eğilimini ortaya koymayı amaçlar. Bir sınıfta çok başarılı ve çok başarısız öğrenciler olabilir ancak bir sınıfın başarısı ortalama başarı olarak değerlendirilir.

Ortalamaların önemini ve neden ortalama ihtiyacı duyulduğunu sağlık alanından bir örnekle açıklamaya çalışalım. Bildiğimiz gibi, tansiyon bir bireyin genel sağlık durumu ile ilgili bilgi veren önemli ölçülerden biridir. Bir hekim, hastasının tansiyonuna yönelik fikir sahibi olmak için hastayı kontrolü sırasında tansiyonunu ölçmekle yetinmez. Bazen hastanın vücuduna bağlanan bir ölçüm ve kayıt cihazıyla tansiyon gün içinde takip edilir, bazen de gün içinde birkaç kez hastanın tansiyonunu ölçerek ya da ölçtürerek kaydetmesini ister. Böylece tek bir ölçüm değeri değil, gün içinde gözlenmiş çok sayıda ölçüm değeri ortaya çıkar.

Bu tansiyon ölçüm değerleri içinde farklı farklı sonuçlar olacaktır. Örneğin hastanın dinlenme zamanlarında ölçüm değerleri düşük, hareket halinde olduğu ya da endişe, korku ya da heyecan gibi duygu durumlarında ise ölçüm değerleri yüksek çıkacaktır.

Hastanın uyuduğu ya da sakin şekilde dinlendiği durumlar ile hastanın hareket halinde ya da duygu durumu açısından bir uyarana maruz kaldığı durumlar açısından tansiyon ölçüm değerleri farklılık gösterecektir. O halde hekim, karar vermek açısından hangi ölçüm değerini dikkate alacaktır?

Bu durumda hekim, hastanın çok sayıdaki tansiyon ölçüm değerlerindeki aşırı düşük ya da aşırı yüksek olan uyku ve hareket halindeki ölçüm değerlerinin etkisini yumuşatarak ortalama tansiyon değerini elde etmeye çalışacaktır. Başka bir deyişle, hastanın normal, olağan, genel tansiyon değerine ulaşmayı amaçlayacaktır.

Bu işlem ortalama hesaplanarak yapılmakta ve serideki düşük ya da yüksek değerler birbirini dengeleyerek ya da etkisini gidererek serinin genel eğilimini yani ortalamasını ortaya çıkarmak amaçlanmaktadır.

Ortalama, adı üzerinde, ortalama durumu ya da genel durumu ifade eden bir ölçüdür. Dolayısıyla, ortalama tansiyon değeri de bir insanın gün içindeki genel tansiyon değerini ifade edecektir.

Ortalamanın neden gerekli olduğuna yönelik olarak gelir örneğini de verebiliriz. Bir toplumda yüksek gelirliler olduğu kadar düşük gelirliler de vardır. Hâl böyle iken bir toplumun gelir seviyesini nasıl ifade etmek doğru olacaktır?

Bir toplum ya da ülkenin gelir durumu da yine ortalama olarak değerlendirilir. Yani, düşük ya da yüksek gelirliler değil, ortalama gelir seviyesi belirlenir ve bir gelir ölçüsü olarak kullanılır.

Genel olarak baktığımızda ortalama hesaplamanın başlıca 3 temel işlevi olduğunu söylemek mümkündür:

Ortalama hesaplayarak,

- Bir serinin hesaplanan ortalama vasıtasıyla kolayca hatırlanabilir olmasını sağlama,
- Birden çok seriyi ortalamaları üzerinden kolaylıkla karşılaştırabilme,

- Serinin normal deęerini ortalama yoluyla bulma ve bylelikle seri hakkında genel bir fikir sahibi olma ve bir takım yorumlar yapabilme.

İstatistik sınavına giren yzer kiřilik iki sınıfın ortalamalarını karřılařtırdığımız rneęi hatırlayalım. A sınıfının ortalamasını 60, B sınıfının ortalamasını 75 olarak hesapladığımızı varsayalım. Bu sonulara gre A ve B sınıflarının genel olarak bařarısını ortalamaları ile deęerlendirebilir, hesapladığımız ortalama deęerler ile kolayca hatırlayabilir ve iki sınıfın bařarı dzeylerini karřılařtırabiliriz.

Benzer řekilde bir ęrencinin 100 zerinden 90 ortalaması olan bir ęrencinin bařarı dzeyinin yksek olduęunu anlarız. Ya da 100 zerinden 55 ortalamaya sahip bir ęrencinin de bařarısının yksek olmadığı anlařılır.

ęrencinin not ortalaması rneęinden de grleceęi gibi, ortalama seriye ynelik olarak da bize bilgi vermektedir. Not ortalaması bir ęrencinin bařarı durumunu zetleyen ve ęrencinin genel bařarı durumunu deęerlendirmede kullanılan evrensel aralardan biridir.

Ortalama, adı zerinde bir serinin birimleri sıralandıęında ortalarda yer alması beklenen bir deęerdir. Bu noktada karřımıza ortalamanın temsili olup olmadığı yani ait olduęu serinin birimlerini iyi temsil edip edemedięi sorunu ıkmaktadır.

Uygulamada ama řphesiz temsili bir ortalama hesaplamaktır. Bir ortalamanın temsili olması konusuna daha sonra ayrıntılı olarak deęineceęiz.

Bu bařlık altında son olarak basit bir kurala deęinmek istiyoruz.

Bir serideki,

en kk deęer, $X_{\min}$,

en byk deęer, $X_{\max}$ řeklinde gsterilirse,

hesapladığımız ortalama deęer,

$$X_{\min} < \text{Ortalama} < X_{\max}$$

sınırları iinde olacak, yani serideki en kk deęer ile en byk deęerin arasında bir deęer alacaktır.

Bu sınırlara zellikle dikkat ekmek istiyoruz. Bir ortalama, hibir zaman serideki en kk deęerden kk ya da en byk deęerden byk olamaz. Nitekim bu basit kural ortalama kavramının anlamı gereęi de byledir.

Yukarıdaki eřitsizlik seride yer alan birimlerin birbirine eřit olmaması halinde geerlidir. İstisnai bir durum olmakla beraber seride yer alan btn birimler birbirine eřitse bu durumda yukarıdaki eřitsizlik, eřitlik halini alır.

Bir öğrencinin bir dönem boyunca 7 ders aldığını ve her ders için sınav notunun 85 olduğunu varsayalım. Bu durumda öğrencinin bütün notları 85 olduğu için not ortalaması da 85 olacaktır. Söz konusu örnekte, serideki en küçük değer, en büyük değer ve ortalama aynı değer olup 85’dir.

Ortalamalar ile ilgili olarak yaptığımız bu genel açıklamalardan sonra şimdi de ortalamaların, türlerinin neler olduğu ve başlıca hangi ortalamaların hangi durumlarda tercih edilmesi gerektiği üzerinde duralım.

5.1. Ortalamalar

Başlıktan da anlaşılacağı gibi, tek bir tane değil; çok sayıda ortalama söz konusudur. Bu ortalamalar, hesaplanmış şekillerine göre iki ana grupta değerlendirilir:

- Analitik ortalamalar
- Analitik olmayan ortalamalar

Şimdi bu iki ortalama grubunu ayrıntılarıyla ele alalım ve bu ortalamaların neler olduğunu görelim.

5.1.1. Analitik Ortalamalar

Analitik ortalamalar, bir veri kümesindeki tüm birimlerin ortalama hesabına katıldığı ve çözüme dayalı olarak hesaplanan ortalamalardır.

Veri setinde yer alan bütün birimler ortalama formülünde kullanılarak ortalama hesabı yapılır. Dolayısıyla, veri, kümesinde yer alan her bir birim ortalama hesabında kullanılmaktadır.

Analitik ortalamalar, veri kümesinde yer alan bütün birimlerin ortalama hesabında kullanılması sebebiyle duyarlı ortalamalar olarak da adlandırılır. Veri içinde yer alan bütün birimlerin ortalama üzerinde etkisi olması nedeniyle duyarlı ortalamalar kavramı kullanılmaktadır.

Duyarlı ortalamalar arasında en çok kullanılan, diğer deyişle kullanım alanı en yaygın olan üç tanesini inceleyeceğiz.

Bunlar; aritmetik ortalama, geometrik ortalama ve kareli ortalama olarak adlandırılan ortalamalardır.

5.1.2. Analitik Olmayan Ortalamalar

Analitik olmayan ortalamalar; bir seride yer alan bütün gözlem değerleri üzerinden hesaplanmayan, başka bir deyişle hesabında seride yer alan bütün gözlem değerlerinin kullanılmadığı ortalamalardır.

Analitik olmayan ortalamalar, seride yer alan bütün birimlerden hesaplanmadığı için, ortalama hesabına dâhil edilmeyen birimlere karşı duyarsızdır. Bu sebeple bu ortalamalara duyarsız ortalamalar da denmektedir.

Analitik olmayan ortalamalar başlığında Mod ve Medyan'ı öğreneceğiz.

5.2. Aritmetik Ortalama

Aritmetik ortalama analitik ortalamalar içinde en yaygın kullanım alanına sahip olan ortalama ve ortalama dendiğinde de akla aritmetik ortalama gelir.

Aritmetik ortalama, günlük yaşamımız içinde hepimizin karşılaştığı, hatta basit bir hesaplama hesaplayabildiği bir ortalama değildir.

İstatistik ile hiç ilgisi olmayan biri bile not ortalamasının 85 olmasının ne anlama geldiğini bilir. Ya da bir ay içinde beş kez farklı fiyatlar üzerinden peynir alan biri, o ay için ortalama peynir fiyatını kolaylıkla hesaplayabilir.

Aritmetik ortalama; anlaşılması, hesaplanması ve yorumlanması bakımından en kullanışlı ve yaygın ortalama değildir.

Basit bir örnek üzerinden aritmetik ortalama mantığını ve buna bağlı olarak nasıl hesaplandığını görelim.

Bir öğrencinin bir dönemde 5 ders aldığını ve dönem sonu itibarıyla notlarının 55, 65, 65, 70 ve 90 olduğunu varsayarak öğrencinin not ortalamasını hesaplayalım.

Öğrencinin beş dersi ve dolayısıyla beş tane notu olduğundan aritmetik ortalama hesaplayabilmek için beş derste öğrencinin aldığı notları toplayarak öğrencinin toplam ders sayısı olan beşe bölmek gerekecektir. Daha açık bir ifade ile 5 dersin notlarını toplamak ve kaç tane değeri topladıysak, toplamı o sayıya bölmemiz gerekecektir.

$$\text{Aritmetik Ortalama} = \frac{55 + 65 + 70 + 70 + 90}{5} = 70$$

Gözlem değerlerinin X_i biçiminde ifade edilmesi halinde aritmetik ortalamanın sembolü $\bar{X}$ olmaktadır.

Seride kaç adet gözlem değeri olduğunu ifade etmek üzere de n sembolü kullanılmaktadır. Daha önce de değindiğimiz gibi, bir seride yer alan birimlerin tekrar sayılarına **frekans** denilmektedir. Toplam frekans sayısı yani toplam gözlem sayısı da n sembolü ile gösterilmektedir.

Sembolleri kullanarak aritmetik ortalamanın formülünü yazmak istersek, serideki en küçükten en büyüğe n adet birimi toplamak ve toplam birim sayısı olan n 'e bölmek gerekecektir.

Aritmetik ortalamanın hesabına yönelik oluşturduğumuz bu tanımı bir formül ile ifade edelim:

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

Yukarıdaki formül, aritmetik ortalama hesaplayabilmek için, seride yer alan ilk sıradaki birimden son sırada yer alan, n. sıradaki birime kadar tüm birimleri toplamak ve toplam birim sayısına bölmek gerektiğini göstermektedir.

Formülde yer alan $\sum$ sembolü, toplama işlemcisi olup birimlerin toplanmasını ifade eder. $\sum$ sembolünün altında yer alan değer toplama işlemine başlanacak değeri, üzerinde yer alan değer de toplama işlemine dâhil edilecek son değeri gösterir. Nitekim yukarıdaki formül de $i=1$ 'den n 'e kadar toplama işlemi yapılacağını, yani birinci birimden başlanarak n. birime, yani seride yer alan sonuncu birime kadar toplama işlemine devam edileceğini göstermektedir.

Yukarıdaki formülden de takip edilebileceği gibi $\sum$ işaretinin sağında yer alan başlangıç ve bitiş değerleri formülün anlaşılabilirliğini zorlaştırmaktadır. Oysa, aksi belirtilmedikçe $\sum$ işareti ilk değerden son değere kadar seride yer alan bütün birimlerin toplanacağını ifade etmektedir. Başka bir deyişle, $\sum$ toplama işlemcisinin altına ve üstüne sınır değerlerinin yazılmaması durumunda, toplama işleminin serideki ilk birimden son birime kadar bütün değerleri kapsayacağı anlaşılmaktadır. Bu yaklaşımı benimseyerek, aritmetik ortalama formüllerini daha yalın halde kullanmayı tercih edeceğiz.

5.2.1. Basit Serilerde Aritmetik Ortalamanın Hesabı

Bu basit yaklaşımla, basit seride aritmetik ortalama formülü;

$$\bar{X} = \frac{\sum X_i}{n}$$

şeklini almaktadır.

Yukarıdaki formül, birimleri küçükten büyüğe doğru sıralanmış bir basit seride, ilk birimden son birime kadar tüm birimleri toplayıp, elde edilen toplamı; toplam birim sayısına bölmek suretiyle hesaplanacak aritmetik ortalama formülüdür.

Basit seride aritmetik ortalama hesabına yönelik olarak bir örnek daha görelim:

Örnek:

Bir otomobil markası yeni ürettiği bir modelin şehir içi yakıt tüketimini tespit etmek istiyor. Beş araç, İstanbul trafiğinde çeşitli güzergâhlarda 100 km yol kat ediyor. Tespit edilen yakıt tüketim miktarları aşağıda verilmiştir.

Araçlar	X_i (100 km’de yakıt tüketimi)
1. Araç	8,3
2. Araç	8,5
3. Araç	8,7
4. Araç	8,7
5. Araç	8,8
	$\sum_{i=1}^n X_i = 43$

Tablodan da takip edilebileceği gibi, beş araç şehir içinde 100 km mesafeyi 8,3 - 8,5 - 8,7 - 8,7 - 8,8 litre tüketerek kat etmiştir. Amacımız söz konusu aracın 100 km mesafeyi ortalama olarak kaç litre benzin tüketerek kat ettiğini tespit etmektir.

Buna göre, basit seride aritmetik ortalama formülünü kullanarak, yani birimlerin değerlerini toplayıp toplam birim sayısına bölerek, aritmetik ortalamayı kolayca hesaplayabiliriz.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{8,3+8,5+8,7+8,7+8,8}{5} = \frac{43}{5} = 8,6 \text{ litre}$$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{43}{5} = 8,6 \text{ litre}$$

Elde ettiğimiz sonuca göre, söz konusu araç şehir içinde 100 km yolu ortalama olarak 8,6 litre benzin tüketerek kat etmektedir.

5.2.2. Tasnif Edilmiş Serilerde Aritmetik Ortalamanın Hesabı

Önceki bölümde de değindiğimiz gibi, bölünme (frekans) serileri basit seri, tasnif edilmiş seri ve sınıflanmış seri olarak üç başlıkta ele alınmaktadır. Basit seride aritmetik ortalama hesabının formülü ve uygulamasını öğrendik. Şimdi de sırasıyla, tasnif edilmiş ve sınıflanmış serilerde formüllerin nasıl olduğunu görelim ve bununla ilgili uygulama yapalım.

Hatırlanacağı gibi, tasnif edilmiş seri; veri kümesi içinde yer alan gözlemleri, tekrar eden gözlem değerlerini bir araya toplamak suretiyle tasnif etmek ve her bir gözlemden kaç adet bulunduğunu gösteren frekansları kullanarak düzenlenen seri türüdür.

Tasnif edilmiş seriler, gözlem değerlerinin neler olduğunu gösteren bir sütun ve bunun karşısında da her bir gözlem değerinden kaç tane bulunduğunu gösteren frekansların yer aldığı sütun olmak üzere iki sütundan oluşan tablolar yardımıyla gösterilir. Frekans sütunu, her bir gözlem değerinden kaç tane bulunduğunu gösterdiğinden tüm matematiksel işlemlerde kullanılmalıdır.

Tasnif edilmiş serilerde kullanılacak aritmetik ortalama formülü aşağıdaki gibidir:

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i}$$

Aritmetik ortalama formülüne baktığımızda, her bir gözlem değerinin; o gözlem değerinden kaç tane olduğunu gösteren kendi frekansı ile çarpılarak toplanıldığını ve bu toplamın veri kümesinde yer alan toplam gözlem sayısına bölündüğünü görmekteyiz.

Tasnif edilmiş seride aritmetik ortalamanın nasıl hesaplandığını bir örnek yardımıyla görelim:

Örnek:

Bir sertifika programına katılan 20 öğrencinin program sonunda yapılan değerlendirme sınavından aldığı puanlar aşağıda gösterilmiştir:

Puanlar X_i	Frekanslar n_i	$n_i X_i$
10	2	20
20	1	20
30	1	30
40	1	40
50	3	150
60	2	120
70	4	280
80	2	160
90	3	270
100	1	100
	$\sum n_i = 20$	$\sum n_i X_i = 1190$

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{1190}{20} = 59,5$$

Tasnif edilmiş seriye yönelik olarak yukarıda verdiğimiz örneği birlikte inceleyelim. Frekans tablosundan da kolayca görüleceği gibi, değerlendirme sınavından sertifika programına katılan katılımcılardan iki tanesi 10, bir tanesi 20, bir tanesi 30, bir tanesi 40, üç tanesi 50, iki tanesi 60, dört tanesi 70, iki tanesi 80, üç tanesi 90 ve bir tanesi 100 almıştır.

Bir an için yukarıdaki aritmetik ortalama formülünü unutalım. Aritmetik ortalamanın serideki tüm birimleri toplayıp, toplam birim sayısına bölerek hesaplandığını biliyoruz.

O halde, yukarıda açık olarak yazdığımız iki tane 10, bir tane 20, bir tane 30, bir tane 40, üç tane 50, iki tane 60, dört tane 70, iki tane 80, üç tane 90 ve bir tane 100 bulunan bir serinin toplamını nasıl alırız?

Çok basit bir yaklaşımla, her bir değeri tekrar sayısıyla çarpar ve sonuçları toplarız. Yani;

$$(2.10)+(1.20)+(1.30)+(1.40)+(3.50)+(2.60)+(4.70)+(2.80)+(3.90)+(1.100)$$

yaklaşımıyla serideki tüm gözlem değerlerinin toplamını hesaplamış oluruz.

Dikkat edilecek olursa, burada her bir değeri kendi frekansı ile çarptık. Yani ($n_i \cdot X_i$) çarpımlarını hesapladık. Sonra, bu çarpım sonuçlarını toplayarak, serinin toplam değerine ulaştık.

Tasnif edilmiş seride aritmetik ortalama formülünü bir kez daha yazalım ve yukarıda yaptığımız açıklamaları göz önüne alarak formülü değerlendirelim.

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i}$$

Formülden de görüleceği gibi yukarıda detaylıca açıklamaya çalıştığımız gözlem değerlerinin frekanslar ile çarpılarak toplanması yaklaşımı formülde sembollerle ifade edilmiştir. Her bir gözlem değeri frekansı ile çarpılarak toplanmakta ve toplam frekans sayısına bölünmektedir.

Dolayısıyla, aritmetik ortalamanın tanım formülüyle de çözüm yaptığımızda, ortalamayı, yine,

$$\bar{X} = \frac{2 \cdot 10 + 1 \cdot 20 + 1 \cdot 30 + 1 \cdot 40 + 3 \cdot 50 + 2 \cdot 60 + 4 \cdot 70 + 2 \cdot 80 + 3 \cdot 90 + 1 \cdot 100}{2 + 1 + 1 + 1 + 3 + 2 + 4 + 2 + 3 + 1}$$

$$\bar{X} = \frac{1190}{20} = 59,5$$

olarak hesaplıyoruz.

Tasnif edilmiş serilerde aritmetik ortalama hesabında sıkça rastlanan bir hata, bir veri kümesinde birimlerin tekrar sayısını gözden kaçırmak ya da toplam gözlem sayısını hesaplarken toplam frekans değerini almak gerekirken kaç değişik gözlem değeri varsa bu değeri toplam frekans gibi düşünülmesi hatasıdır.

Yukarıdaki tablodan da görüleceği gibi, serimizde 10 değişik gözlem değeri vardır ama gözlem değerlerinin bir kısmından birkaç tane bulunmaktadır ve toplam gözlem sayımız 20'dir. 10 değişik gözlem değeri olması sebebiyle toplam gözlem sayısının 10 gibi düşünülmesi ve aritmetik ortalama formülünde 20'ye değil 10'a bölerek işlem yapmak hesapladığımız aritmetik

ortalamayı yanlış hesaplamamıza yol açacaktır. Bu konuya özellikle dikkat çekmek istediğimizi ifade edelim.

Tasnif edilmiş seride aritmetik ortalamanın nasıl hesaplandığına yönelik bir örnek daha görelim.

Örnek:

Bir süpermarket sattığı çilekleri, ürünün yapısının ve tazeliğinin zaman içinde bozulması sebebiyle farklı fiyatlama yaklaşımı ile satıyor. Aşağıda çileklerin satış fiyatları ve kaç kilo çileğin bu fiyattan satıldığını gösteren tasnif edilmiş seri verilmektedir.

Buna göre, çileğin ortalama satış fiyatı ne olmuştur?

Çilek Fiyatı (TL) X_i	Satılan Miktar (kg) n_i	$n_i X_i$
10	80	800
9	75	675
8	60	480
7	35	245
	$\sum n_i = 250$	$\sum n_i X_i = 2200$

Çileğin ortalama satış fiyatını bulmak için hangi fiyattan kaç kilo satıldığını ve ne kadar gelir elde edildiğini bulup, bunu tüm fiyatlar için yapıp sonuçları toplamak ve satılan miktara bölmek gerekecektir.

Aritmetik ortalama formülüyle çileğin ortalama fiyatının ne olduğunu kolayca hesaplayabiliriz:

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{2200}{250} = 8,8 \text{ TL}$$

Görüldüğü gibi, çileğin ortalama satış fiyatı 8,8 TL olmuştur. Yani çilek kilosu ortalama olarak 8,8 TL'den satılmıştır.

5.2.3. Sınıflanmış Seride Aritmetik Ortalamanın Hesabı

Basit ve tasnif edilmiş seride aritmetik ortalamanın nasıl hesaplandığını gördük. Şimdi de sınıflanmış serilerde aritmetik ortalamanın nasıl hesaplandığını öğrenelim.

Hatırlanacağı gibi, veri kümesi içinde birbirine yakın gözlem değerleri homojen sınıflar içinde toplanmak suretiyle sınıflanmış frekans serileri oluşturuyorduk. Dolayısıyla, bu serilerde

gözlem değerlerini ayrı ayrı görmek mümkün olmamakta, gözlem değerlerinin yer aldığı sınıfları ve bu sınıflara düşen gözlem sayısının ne olduğunu, yani frekansını görmek mümkün olmaktadır.

Aşağıda daha önceki bölümde düzenlediğimiz bir sertifika programına katılan 20 katılımcının değerlendirme sonuçlarını düzenlediğimiz sınıflanmış frekans serisi verilmiştir:

Puan Sınıfları	Sınıf Orta Noktaları (m_i)	Frekanslar (n_i)	$n_i m_i$
0-20	10	2	20
20-40	30	2	60
40-60	50	4	200
60-80	70	6	420
80-100	90	6	540
		$\sum n_i = 20$	$\sum n_i m_i = 1240$

Sınıflanmış frekans serilerinde aritmetik ortalama hesaplayabilmek için basit bir varsayıma ihtiyaç duymaktayız. Bu varsayım, bir sınıfta bulunan gözlem değerlerini temsil etmek üzere o sınıfın merkez noktasının, yani sınıf orta noktasının kullanılmasıdır.

Tablodan da görüleceği gibi, sınıfları temsil etmek üzere her sınıf için bir temsil noktası hesaplıyor ve bu sınıf içinde yer alan birimlerin tamamı sanki bu değeri almış gibi düşünüyoruz.

Sınıf orta noktalarının nasıl hesaplandığını önceki bölümde açıklamıştık. Burada hatırlatmakla yetinelim.

Sınıf orta noktası;

$$m_i = \frac{\text{Sınıf üst sınır değeri} + \text{Sınıf alt sınır değeri}}{2}$$

formülüyle hesaplanmaktadır. Dolayısıyla, (0-20) sınıfının sınıf orta noktası, $(20+0)/2=10$ şeklinde hesaplanmaktadır. Benzer şekilde, (20-40) sınıfının sınıf orta noktası da $(40+20)/2=30$ olarak hesaplanmaktadır. Diğer sınıfların sınıf orta noktaları tablo içeriğinde verilmiştir.

Sınıf orta noktalarını hesapladıktan sonra, cebirsel işlemlerde bu değerler kullanılacağı için aritmetik ortalama çözümümüz tasnif edilmiş seri ile benzer bir formül ile yapılacaktır. Tek farklılık, tasnif edilmiş seride gözlem değerlerinin kendisi kullanılırken, burada basit bir varsayımla, sınıfların sınıf orta noktaları gözlem değeriymiş gibi değerlendirilmektedir.

Sınıflanmış seride aritmetik ortalama formülü,

$$\bar{X} = \frac{\sum n_i m_i}{\sum n_i}$$

şeklinde olmaktadır ve burada m_i değerleri daha önce de ifade ettiğimiz gibi sınıfların orta noktalarını göstermektedir.

Yukarıdaki sınıflanmış frekans serisini tekrara düşmek pahasına konunun daha iyi anlaşılması için bir kez daha yalın şekliyle veriyoruz.

Tablodan da görüleceği gibi, tablo içeriği tasnif edilmiş seri görüntüsüne benzemektedir. Dolayısıyla, aritmetik ortalama hesabı aynı yaklaşımla çözülecektir.

Sınıf Orta Noktaları (m_i)	Frekanslar (n_i)	$n_i m_i$
10	2	20
30	2	60
50	4	200
70	6	420
90	6	540
	$\sum n_i = 20$	$\sum n_i m_i = 1240$

$$\bar{X} = \frac{\sum n_i m_i}{\sum n_i} = \frac{1240}{20} = 62 \text{ puan}$$

Sertifika programına katılan 20 katılımcının değerlendirme sınavı ortalaması 62 puan olarak hesaplanmıştır.

Konuyu pekiştirmek adına bir örnek çözüm daha vermek istiyoruz.

Örnek:

Aşağıda önceki bölümde oluşturduğumuz 100 işyerinde çalışan personel sayısına yönelik sınıflanmış frekans dağılımı verilmiştir. Buna göre, işletme başına ortalama personel sayısını hesaplayalım.

Sınıflar	Sınıf orta noktaları m_i	Frekans n_i	$n_i m_i$
20-27	23,5	5	117,5
27-34	30,5	8	244
34-41	37,5	29	1087,5
41-48	44,5	18	801
48-55	51,5	12	618
55-62	58,5	11	643,5
62-69	65,5	10	655
69-76	72,5	7	507,5
		$\sum n_i = 100$	$\sum n_i m_i = 4674$

Sınıf orta noktalarını belirledikten sonra, her bir sınıf orta noktası sanki bir gözlem değeriymiş gibi düşünülerek, bu farazi (varsayımsal) değerin sınıf içinde kaç kere tekrar ettiğini gösteren frekans değerleriyle çarpılıp toplanması ve toplam frekans sayısına bölünmesi suretiyle aritmetik ortalama hesaplanmaktadır.

Yukarıdaki sınıflanmış frekans serisinin yer aldığı tablonun içeriğinden de aritmetik ortalama hesabında izlenen aşamalar kolaylıkla takip edilebilir.

İşletme başına ortalama çalışan sayısı,

$$\bar{X} = \frac{\sum n_i m_i}{\sum n_i} = \frac{4674}{100} = 46,74 \cong 47 \text{ çalışan}$$

şeklinde hesaplanmaktadır.

5.2.4. Ağırlıklı Aritmetik Ortalama

Aritmetik ortalama başlığı altında verdiğimiz formüller, gözlem değerleri ya da birimlerin önem derecelerinin aynı kabul edildiği durumlarda kullanılmaktadır. Oysa bazı hallerde birimler arasında bazılarının diğerlerine kıyasla daha önemli olduğu durumlar da olabilir. Söz gelimi, bir öğrencinin bir yıl boyunca aldığı çeşitli dersler içinde haftalık ders saati yani kredisi yüksek olan mesleki dersleri de vardır; daha az ders saatine sahip kültür dersleri de vardır.

İşte, gözlem değerleri arasındaki bu önem farkının ortalama hesabına yansıtılması istendiğinde **ağırlıklı ya da tartılı ortalama** adı verilen ortalama tercih edilir.

Ağırlıklı ortalama ile ilgili olarak bir öğrencinin bir dönem boyunca aldığı derslerin yılsonu notlarına yönelik ağırlıklı ortalamayı ve aradaki farkı görmek üzere klasik aritmetik ortalamayı hesaplayalım.

Örnek:

Bir öğrencinin aldığı 7 dersin yılsonu başarı notları aşağıdaki tablo ile düzenlenmiştir. Buna göre, öğrencinin aritmetik not ortalamasını ve ağırlıklı not ortalamasını hesaplayınız.

Ders Adı	Haftalık Ders Saati t_i	Başarı Notu X_i	$t_i X_i$
Mikro İktisat	4	50	200
Makro İktisat	4	70	280
İşletme	3	75	225
Muhasebe	3	75	225
İstatistik	4	60	240
Siyaset Bilimi	2	75	150
Güzel Sanatlar	2	75	150
	$\sum t_i = 22$	$\sum X_i = 480$	$\sum t_i X_i = 1470$

Öncelikle öğrencinin bütün derslerin aynı önemde değerlendirildiği, yani kredisi yüksek ya da düşük tüm derslerin eşit kabul edildiği durum için basit aritmetik ortalamasını yani bilinen adıyla not ortalamasını hesaplayalım.

Burada, sadece notlardan oluşan tek sütunluk bir veri söz konusudur ve dolayısıyla basit seri gibi düşünerek aritmetik ortalama hesaplanmalıdır.

Basit seri için aritmetik ortalama formülü kullanılarak öğrencinin not ortalaması,

$$\bar{X} = \frac{\sum X_i}{n} = \frac{480}{7} \cong 68,57$$

hesaplanmıştır.

Öğrencinin 7 adet dersi olduğu için not toplamının 7'ye, yani ders sayısına bölündüğüne dikkat ediniz.

Şimdi de, derslerin haftalık ders saatlerini yani kredilerini ağırlık kabul ederek ağırlıklı not ortalaması hesaplayalım. Buradaki temel mantık, kredisi yüksek olan derse ortalama hesabında daha yüksek pay vermektir. Kredisi 4 olan dersten alınan not, sanki 4 kez alınmış gibi düşünülmekte, kredisi iki olan dersin notu ise iki kez alınmış gibi düşünülmektedir.

Ağırlıklı ortalama formülüyle öğrencinin ağırlıklı not ortalaması,

$$\bar{X} = \frac{\sum t_i X_i}{\sum t_i} = \frac{1470}{22} \cong 66,82$$

olarak hesaplanmıştır.

Ağırlıklı ortalama hesabının formülleri, basit ortalama formüllerine ağırlıkları ilave etmekle bulunur. Yukarıdaki örnekte, öğrencinin not serisi basit seri özelliği göstermektedir ve ağırlıkların ilavesiyle formül, tasnif edilmiş seri için kullanılan formüle benzemiştir. Bunun sebebi kredi sayılarının alınan notun tekrar sayısı gibi değerlendirilmesidir.

Örneğimizdeki öğrencinin basit not ortalaması ile ağırlıklı not ortalaması birbirinden farklı çıkmıştır. Basit not ortalaması 68,57 iken, ağırlıklı not ortalaması 66,82 çıkmıştır. Bunun sebebi öğrencinin kredisi yüksek olan derslerden daha düşük not almış olmasıdır. Daha önce de ifade ettiğimiz gibi, öğrenci kredisi yüksek olan mikro iktisat dersinden 50 alırken, kredisi düşük olan siyaset bilimi dersinden 75 almıştır. Bu iki dersin kredisi kadar tekrar ettiği varsayıldığından mikro iktisattan aldığı 50 puan 4 kez, siyaset biliminden aldığı 75 puan ise 2 kez alınmış gibi düşünülmüştür.

Ağırlıklı ortalama kullanımı not ortalaması gibi basit seri özelliği gösteren serilerde kullanıldığından diğer seri türlerinde nasıl uygulandığına değinilmeyecektir.

5.2.5. Aritmetik Ortalamanın Özellikleri

Aritmetik ortalama, analitik/duyarlı ortalamalar içinde en yaygın kullanım alanına sahip ortalamadır. Aritmetik ortalamanın uygulamada neden bu kadar yaygın kullanıldığını ve buna bağlı olarak aritmetik ortalamanın özelliklerini kısaca açıklamak istiyoruz.

- Aritmetik ortalama bir serideki bütün gözlem değerleri üzerinden hesaplandığı için genel olarak bakıldığında seriyi temsil etmek üzere tek bir değer belirlenmesi amaçlanıyorsa aritmetik ortalama bu amaçla kullanılabilecek en uygun ortalamadır.
- Aritmetik ortalama cebirsel işlemlere elverişli olması, kolay hesaplanması ve anlaşılması sebebiyle de yaygın kullanım alanına sahiptir. Nitekim ortalama dendiğinde akla sadece aritmetik ortalama gelmektedir.
- Cebirsel işlemlere elverişli olması, aritmetik ortalamanın pek çok istatistik analiz yönteminin temelinde veya içeriğinde yer almasına sebep olmuştur.
- Aritmetik ortalama, artış miktarı sabit aritmetik dizi özelliği gösteren serilerde daha temsili nitelik taşımaktadır. Başka bir deyişle, aritmetik ortalama bu tür serileri daha iyi temsil etmektedir.
- Aritmetik ortalamanın kullanılabilmesi için verinin en az aralık ölçeğinde toplanması gerekir. Nominal ve ordinal ölçekte toplanmış veriler için aritmetik ortalama kullanılması uygun değildir.
- Gözlem değerleri arasında aşırı küçük ya da aşırı büyük değerlerin varlığı aritmetik ortalamanın temsili olma özelliğini ortadan kaldırır. Aritmetik ortalama aşırı değerlere karşı en hassas ortalamadır.

- Aritmetik ortalama seride, serinin yapısını bozan aşırı değerlerin olması durumunda temsili olma özelliğini kaybetmesi sebebiyle kullanılmaz. Bunun yerine aritmetik ortalama kadar aşırı değerlerden etkilenmeyen geometrik ortalama tercih edilir.
- Aritmetik ortalama toplamı sıfır olan serilerin ortalaması hesaplanmak istenildiğinde de kullanılamamaktadır. Toplamı sıfır olan serilerde ise kareli ortalama kullanılmaktadır.

Aritmetik ortalamanın kullanılmadığı yerlerde kullanılan geometrik ve kareli ortalamaların nasıl hesaplandığını bir sonraki bölümde ele alacağız.

Bu Bölümde Neler Öğrendik?

Bu bölümde ortalamalar başlığı altında bir seriyi nasıl özetleyeceğimizi ya da o seriyi temsil etmek üzere tek bir değer belirleme işleminin nasıl yapıldığını gördük.

Bir veriyi özetlemek amacıyla kullanılabilecek ortalamaların analitik ve analitik olmayan ortalamalar şeklinde ayrıldığını öğrendik.

Analitik ortalamalar kapsamında bu bölümde aritmetik ortalamayı detaylı olarak inceledik ve çeşitli seri türlerine göre nasıl hesaplandığı üzerinde durduk. Öte yandan, aritmetik ortalamanın özelliklerini teker teker inceleyerek, aşırı değer içeren serilerde temsili olma özelliğini yitirdiğini gördük.

Bölüm Soruları

- 1- Bir seriyi özetlemek, akılda kalıcı tek bir değer olarak ifade etmek amacıyla kullanılan istatistik ölçüler olarak adlandırılır.
 - a- Mod
 - b- Medyan
 - c- Ortalama
 - d- Ortanca
 - e- Egemen değer
- 2- Aşağıdaki ortalamalardan hangisi analitik ortalamalar arasında yer almaz?
 - a- Mod
 - b- Aritmetik ortalama
 - c- Geometrik ortalama
 - d- Kareli ortalama
 - e- Duyarlı ortalama
- 3- Seride bulunan aşırı değerlere karşı en duyarlı ortalama aşağıdakilerden hangisidir?
 - a- Mod
 - b- Medyan
 - c- Aritmetik ortalama
 - d- Geometrik ortalama
 - e- Hepsi
- 4- Bir seride yer alan gözlem değerleri arasında önem farkı bulunması halindeortalama tercih edilir.
 - a- Ağırlıklı
 - b- Ağırlıksız
 - c- Mod
 - d- Medyan
 - e- Egemen değer

- 5- Bir sürücü kursu katılımcısı girdiği ehliyet yazılı sınavında ilkyardım sınavından 85, motor bilgisi sınavından 80 ve trafik sınavından 90 almıştır. Buna göre, söz konusu katılımcının ortalama sınav notu nedir?
- a- 80
 - b- 85
 - c- 90
 - d- 95
 - e- 98
- 6- TEOG sınavına hazırlanmakta olan bir ortaöğretim öğrencisi ardarda girdiği 3 deneme sınavından 90, 93, 98 almıştır. Buna göre öğrencinin ortalama başarı puanı nedir?
- a- 93,67
 - b- 92
 - c- 91
 - d- 95
 - e- 95,5
- 7- Bir ilde haziran ayı gündüz sıcaklıkları bir hafta boyunca gözlenerek günlük ortalama sıcaklık seviyeleri tespit ediliyor. Bu değerler 22, 24, 25, 25, 25, 26, 26 şeklindedir. Buna göre söz konusu hafta için ortalama sıcaklık nedir?
- a- 23
 - b- 24
 - c- 24,714
 - d- 25
 - e- 25,523
- 8- Bir seride aşırı değer bulunması halinde..... tercih edilir.
- a- Aritmetik ortalama
 - b- Tartılı aritmetik ortalama
 - c- Geometrik ortalama
 - d- Kareli ortalama
 - e- Duyarlı ortalama

- 9- Bir serinin toplam değerinin sıfır olması durumunda tercih edilir.
- a- Aritmetik ortalama
 - b- Kareli ortalama
 - c- Geometrik ortalama
 - d- Tartılı aritmetik ortalama
 - e- Duyarlı ortalama
- 10- Aritmetik ortalamanın hesaplanabilmesi için verinin en azölçüm düzeyinde toplanmış olması gerekir.
- a- Nominal
 - b- Kategorik
 - c- Ordinal
 - d- Sıralı
 - e- Aralık

Cevaplar

- 1- c
- 2- a
- 3- c
- 4- a
- 5- b
- 6- a
- 7- c
- 8- c
- 9- b
- 10- e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

6. ANALİTİK ORTALAMALAR – GEOMETRİK VE KARELİ ORTALAMALAR

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Daha önceki bölümlerde verilerin toplanmasını ve düzenlenerek sunulmasını, toplanmış ve düzenlenmiş verilerin ve verilerden elde edilen bilgilerin ortalamalar ile özetlenebileceğini öğrendik.

Önceki bölümde analitik ortalamalar içinde en yaygın kullanım alanına sahip olan aritmetik ortalamayı inceledik. Bu bölümde de analitik ortalamalar kapsamında geometrik ve kareli ortalamayı öğrendik.

Geometrik ve kareli ortalamaların nasıl hesaplandığı üzerinde durarak hangi durumlarda tercih edilmeleri gerektiğini ele aldık.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Ortalaması hesaplanacak seride aşırı değer bulunması durumunda aritmetik ortalama kullanmak doğru mudur?
- 2) Amaç temsili ortalama hesaplamak olduğunda aritmetik ortalama dışında analitik ortalamalar da var mıdır?
- 3) Gözlem değerleri arasında negatif değer varsa ortalama hesaplamak mümkün müdür?
- 4) Toplamı sıfır olan serinin ortalaması hesaplanabilir mi?
- 5) Geometri dizi özelliği gösteren, artış oranı az-çok sabit serilerde hangi ortalama kullanılmalıdır?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Seride aşırı değer bulunması halinde hangi ortalama tercih edilmeli.	Aşırı değer içeren serilerde geometrik ortalamanın aritmetik ortalamaya göre daha temsili bir ortalama olduğunu bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Geometrik ortalama hesabı	Geometrik ortalamanın ne zaman tercih edilmesi gerektiğini ve nasıl hesaplandığını bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Geometrik ortalamanın özellikleri	Geometrik ortalamanın neden önemli bir ortalama olduğunu bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Kareli ortalama hesabı	Kareli ortalamaya neden gerek duyulduğunu ve nasıl hesaplandığını bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.
Kareli ortalamanın özellikleri	Kareli ortalamanın özelliklerini ve önemini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik ortalamalar ve aritmetik ortalama ile ilgili kısımlarından yararlanılabilir.

Anahtar Kavramlar

- Aşırı değer: Bir veri grubu içinde verinin genel yapısına uymayan çok büyük ya da çok küçük gözlem değerleri.
- Geometrik ortalama: Seride aşırı değer bulunması halinde tercih edilen ve aritmetik ortalamaya kıyasla daha temsili olma özelliği taşıyan analitik ortalama.
- Temsili ortalama: bir seriyi tek bir değer olarak en iyi temsil etme durumu.
- Kareli ortalama: Seride negatif değer olması halinde veya toplamı sıfır olan serilerin ortalaması hesaplanmak istendiğinde kullanılabilecek analitik ortalama.

6.1. Geometrik Ortalama

Geometrik ortalama da aritmetik ortalama gibi seride yer alan bütün birimlerin ortalama hesabına dâhil edildiği analitik ya da duyarlı ortalamalar grubu içinde yer alan bir ortalamadır.

Geometrik ortalama, aritmetik ortalama gibi yaygın kullanım alanı olan bir ortalama olmayıp, özellikle tercih edildiği bazı spesifik durumlar söz konusudur.

Aritmetik ortalama başlığı altında da ifade ettiğimiz gibi, aritmetik ortalama artış miktarı sabit, başka bir deyişle aritmetik dizi özelliği gösteren seriler için uygun bir ortalamadır ve özellikle bu tür serilerde ait olduğu seriyi iyi temsil etme kabiliyetine sahiptir.

Ortalamanın, ait olduğu seriyi tek bir değer ile özetlemesini ve temsil etmesini hedeflemek ortalama hesaplamanın en temel amacıdır. Oysa, seride aşırı değer bulunması halinde aritmetik ortalamanın temsil kabiliyeti ciddi derecede zedelenmekte ve aritmetik ortalama, ortalama olma özelliğini yitirerek aşırı değerlere doğru çekilmektedir. Benzer şekilde serinin artış miktarının sabit olduğu aritmetik dizi değil de, artış oranı sabit, hızla büyüyen geometrik dizi özelliği gösterdiği durumlarda da aritmetik ortalama için benzer bir sıkıntı oluşmakta ve aritmetik ortalama seriyi temsil etme özelliğini büyük ölçüde yitirmektedir.

Yukarıda belirttiğimiz gibi, serinin özelliğine bağlı olarak aritmetik ortalamanın tercih edilmediği durumlarda geometrik ortalamayı kullanmak söz konusu olmaktadır.

Öte yandan, geometrik ortalamanın da aritmetik ortalama gibi aralık ya da oran ölçüm düzeyinde ölçülmüş veriler için kullanılabilecek bir ortalama olduğunu ifade edelim.

Geometrik ortalamanın nasıl hesaplandığına yönelik olarak uygulamada kullanılmakta olan formülünü vereceğiz.

Geometrik ortalama, bir seride yer alan gözlem değerlerinin logaritmalarının ortalamasıdır. Yani, geometrik ortalama; önce birimlerin 10 tabanında logaritmalarını almak, daha sonra bu değerlerin aritmetik ortalamasını hesaplamak ve en son olarak da bulunan sonucun antilog işlemiyle orijinal seriye dönüştürülmesiyle hesaplanır.

Bir kez daha ifade etmek gerekirse, geometrik ortalama, seride yer alan birimlerin logaritma değerlerinin aritmetik ortalamasıdır.

Basit seri, tasnif edilmiş seri ve sınıflanmış serilerde geometrik ortalamanın nasıl hesaplandığına yönelik formülleri vererek örnek çözümlerle uygulamasını yapalım.

6.1.1. Basit Seride Geometrik Ortalamanın Hesabı

Basit seri, gözlem değerlerinin küçükten büyüğe (ya da büyükten küçüğe) doğru sıralanarak oluşturulan en basit seri türüdür.

Basit seride geometrik ortalama formülünü vermeden önce, aritmetik ortalama formülünü hatırlamak ve bu formülden geometrik ortalama formülüne erişmek istiyoruz.

Basit seride aritmetik ortalama formülünün,

$$\bar{X} = \frac{\sum X_i}{n}$$

şeklinde olduğunu hatırlayalım.

Daha önce de ifade ettiğimiz gibi, geometrik ortalama gözlem değerlerinin logaritmalarının aritmetik ortalamasıdır. Dolayısıyla formülde (X_i) gördüğümüz yere ($\log X_i$) yazarsak, aritmetik ortalama formülünü geometrik ortalama formülüne dönüştürmüş oluruz.

Yani geometrik ortalama formülü,

$$\log GO = \frac{\sum \log X_i}{n}$$

olacaktır.

Şüphesiz burada hesaplanan ortalama, logaritmik değerler cinsinden hesaplandığı için elde edilen ortalama da geometrik ortalamanın logaritmasıdır. Dolayısıyla, son bir işlem olarak ilgili değerin antilog'u alınarak geometrik ortalamanın bulunması gerekir.

Antilog almanın unutulması halinde bulunulacak ortalama logaritmik bir değer olacağı için seride yer alan birimlere kıyasla oldukça küçük bir değere sahip olacaktır ki bu durum uygulamacılara da bir uyarı gibi düşünülebilir.

Geometrik ortalamanın basit serilerde nasıl hesaplandığına yönelik bir örnek çözüm görelim:

Örnek:

Bir şirkette birlikte çalışan 20-25 yaş aralığında 5 arkadaş kamu spotlarından etkilenerek kontrol amaçlı kan şekeri ölçümü yaptırıyor. Ölçüm değerlerinin düzenlenmesiyle elde edilen tablo aşağıda gösterilmiştir:

İsimler	Açlık Kan Şekeri Ölçüm Değeri X_i	$\log X_i$
Ayşe	85	1,9294
Esmâ	88	1,9445
Banu	90	1,9542
Tuğba	92	1,9638
Fatma	278	2,4440
	$\sum X_i = 633$	$\sum \log X_i = 10,2359$

Yukarıda verilen geometrik ortalama formülünü kullanarak, geometrik ortalamanın logaritmasını,

$$\log GO = \frac{\sum \log X_i}{n} = \frac{10,2359}{5} = 2,04718$$

2,04718 olarak hesaplıyoruz. Bu değerin geometrik ortalamanın logaritması olduğunu bir kez daha vurgulayalım. Nitekim, seride yer alan birimlerin 85-268 aralığında değer aldığına dikkat edecek olursak, böyle bir serinin ortalamasının 2,04718 olması mümkün değildir. Bu aşamada antilog alma işlemi uygulayarak Geometrik ortalamaya ulaşmamız gerekiyor.

$$\text{Antilog}(2,04718) = 111,4756 \cong 111,48$$

Böylelikle 5 iş arkadaşının açlık kan şekeri ortalamasını geometrik ortalama kullanarak 111,48 olarak hesaplamış olduk.

Aynı örnek için serinin aritmetik ortalamasını da hesaplayalım ve geometrik ortalaması ile kıyaslayalım.

Aritmetik ortalama formülünden,

$$\bar{X} = \frac{\sum X_i}{n} = \frac{633}{5} = 126,6$$

olarak hesaplanıyor.

Beş iş arkadaşının açlık kan şekeri ölçüm değerlerinin aritmetik ve geometrik ortalamalarını hesapladığımızda, aritmetik ortalamayı 126,6; geometrik ortalamayı ise 111,48 buluyoruz.

Açlık kan şekeri ölçüm değerlerini incelediğimizde, beş iş arkadaşının 4 tanesinin kan şekeri ölçüm değerlerinin birbirine yakın olduğunu sadece bir çalışanın yüksek açlık şekeri düzeyine sahip olarak diğer dört kişiden ciddi derecede ayrıştığı anlaşıyor.

Aşırı gözlem değeri olarak niteleyeceğimiz 278 değerinin varlığı aritmetik ortalamayı kendisine doğru çekmekte ve yükseltmektedir. Nitekim bulduğumuz aritmetik ortalama değeri 126,6'dır.

Geometrik ortalamaya baktığımızda ise, onun aşırı değer olan 278 değerinden aritmetik ortalama kadar etkilenmediğini ve geometrik ortalamanın 111,48 olarak hesaplandığını görüyoruz.

Buradaki kıyaslamadan da kolaylıkla görülebileceği gibi, geometrik ortalamanın özel olarak kullanıldığı ya da tercih edildiği durumlar, serinin içinde aşırı değer olarak nitelenebilecek diğer gözlem değerlerinden ayrılan çok büyük ya da çok küçük değerlerin var olduğu durumlardır. Böyle durumlarda, aşırı değerlerden aritmetik ortalamaya kıyasla daha az etkilenen ve bu sebeple daha temsili olma özelliği taşıyan geometrik ortalamanın tercih edildiğini görüyoruz.

6.1.2. Tasnif Edilmiş Seride Geometrik Ortalamanın Hesabı

Tasnif edilmiş serilerde geometrik ortalamanın formülünü yine aritmetik ortalama formülünden yola çıkarak vermek istiyoruz. Geometrik ortalamanın, seride yer alan gözlem değerlerinin logaritmalarının aritmetik ortalaması olduğu düşünüldüğünde, aritmetik ortalama formülünde gözlem değerlerinin logaritmalarına yer verdiğimizde, geometrik ortalamanın formülüne erişiyoruz. Ancak, bulduğumuz sonucun geometrik ortalamanın logaritması olduğunu unutmayarak en son aşamada antilog işlemiyle dönüşümü sağladığımızda geometrik ortalama ulaşmış oluyoruz.

Tasnif edilmiş serilerde aritmetik ortalama formülünün,

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i}$$

biçiminde olduğunu hatırlayalım. Daha önce de ifade ettiğimiz gibi, geometrik ortalama gözlem değerlerinin logaritmalarının aritmetik ortalamasıdır. Dolayısıyla formülde (X_i) gördüğümüz yere ($\log X_i$) yazarsak, aritmetik ortalama formülünü geometrik ortalama formülüne dönüştürmüş oluruz. Bu durumda formülümüz,

$$\log GO = \frac{\sum n_i \log X_i}{\sum n_i}$$

olmaktadır. Hesapladığımız değer, gözlem değerlerinin logaritmaları üzerinden hesaplandığından, bulduğumuz değerın antilog dönüşümünü sağlayarak geometrik ortalama ulaşmış oluruz.

Tasnif edilmiş seride geometrik ortalamanın nasıl hesaplandığını bir örnek üzerinde görelim.

Örnek:

Bir klinikte takibi yapılan 80 gebenin tamamladıkları haftalara göre doğum yaptıkları haftalar ve kaç kadının bu haftada doğum yaptığına bağlı olarak düzenlenen tasnif edilmiş frekans dağılımı aşağıdaki tabloda verilmiştir.

Buna göre, verinin yapısını da dikkate alarak uygun ortalama karar vererek, ortalama doğum haftasını hesaplayalım.

Gebelik haftası	Frekans (n _i)	n _i X _i	log X _i	n _i logX _i
34	10	340	1,5315	15,315
35	12	420	1,5441	18,529
36	30	1080	1,5563	46,689
40	19	760	1,6021	30,439
41	9	369	1,6128	14,515
	$\sum n_i = 80$	$\sum n_i X_i = 2969$		$\sum n_i \log X_i = 125,487$

Ortalama doğum haftasını belirlemek üzere, veri kümesinin içinde hatırı sayılır ölçüde geç doğum haftası verisi bulunduğu ve bu veriler hassas bir ölçü olan aritmetik ortalamayı kendilerine doğru çekerek, büyük çıkmasına yol açacaklarından geometrik ortalama kullanılarak daha temsili bir ortalama elde etmek tercih edilmiştir.

Geometrik ortalama formülü kullanılarak önce geometrik ortalamanın logaritması hesaplanmış ve daha sonra bulunan sonucun anti logaritması alınarak geometrik ortalamaya ulaşılmıştır.

$$\log GO = \frac{\sum n_i \log X_i}{\sum n_i} = \frac{125,487}{80} \cong 1,566$$

$$\text{Antilog}(1,566) \cong 36,80 \text{ hafta}$$

Ortalama doğum haftası geometrik ortalama kullanılarak yaklaşık 36,80 hafta bulunmuştur. Yukarıdaki tabloya dikkatlice bakacak olursak, verinin yaklaşık 2/3'ünün 36. hafta ve öncesinde doğum yaptığı görülmektedir. Oysa, yoğun olarak gözlenen doğum haftası bu haftalar olmasına rağmen 40. ve 41. haftalıkken doğan bebeklerin varlığı ortalama doğum haftasını yukarıya doğru çekmiştir.

Yukarıdaki veri için aritmetik ortalama tercih etseydik, ortalamamız;

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{2969}{80} \cong 37,11 \text{ hafta}$$

olacak ve geometrik ortalamaya kıyasla yukarıya doğru kayacaktı.

6.1.3. Sınıflanmış Seride Geometrik Ortalamanın Hesabı

Basit ve tasnif edilmiş seride geometrik ortalamanın formülünü vererek örnek çözümlerini yaptık. Şimdi de sınıflanmış bir dağılım söz konusu olduğunda geometrik ortalamanın nasıl hesaplandığını görelim.

Sınıflanmış dağılımlarda gözlem değerlerini görmek mümkün olmamakta ve gözlem değerlerinin bulunduğu sınır değerleri bilinmektedir. Dolayısıyla ilk olarak sınıfları temsil edecek sınıf orta noktalarının belirlenmesi gerekecektir.

Geometrik ortalama gözlem değerlerinin logaritmaları üzerinden hesaplandığı için, daha sonraki işlem sınıf orta noktalarının logaritmalarını almak ve daha sonra sanki bir tasnif edilmiş seri ile çözüm yapıyormuş gibi çözümün sürdürülmesidir.

Sınıflanmış seride, geometrik ortalama formülünü yine aritmetik ortalama formülünü hatırlayarak düzenleyelim.

Sınıflanmış seriler için aritmetik ortalama formülü,

$$\bar{X} = \frac{\sum n_i m_i}{\sum n_i}$$

şeklinde. Geometrik ortalamanın gözlem değerlerinin logaritmalarının aritmetik ortalaması olduğunu hatırlayarak, aritmetik ortalama formülünde m_i gördüğümüz yere $\log m_i$ yazarak geometrik ortalama formülüne ulaşalım. Yine bu formülle hesaplanacak sonucun bir logaritmik değer olacağını hatırlatalım ve antilog alma işleminin son aşamada unutulmaması gerektiğini bir kez daha yineleyelim.

Sınıflanmış serilerde kullanılacak geometrik ortalama formülü de,

$$\log GO = \frac{\sum n_i \log m_i}{\sum n_i},$$

$$\text{Antilog}(\log GO) = GO$$

şeklinde düzenlenmektedir.

Sınıflanmış dağılımlarda geometrik ortalama çözümünü bir örnek yardımıyla açıklamaya çalışalım.

Örnek:

Bir hastanede bir gün içinde dünyaya gelen bebeklerin doğum ağırlıkları gözlenerek sınıflanıyor ve elde edilen sınıflanmış frekans dağılımı aşağıda veriliyor. Bu verilere göre, geometrik ortalama kullanarak bebeklerin ortalama doğum ağırlığını belirleyelim.

Doğum ağırlığı (gr)	Frekanslar	Sınıf orta noktası m_i	$n_i m_i$	$\log m_i$	$n_i \log m_i$
2500-2700	2	2600	5200	3,4150	6,8300
2700-2900	2	2800	5600	3,4472	6,8944
2900-3100	4	3000	12000	3,4771	13,9084
3100-3300	5	3200	16000	3,5051	17,5255
3300-3500	3	3400	10200	3,5315	10,5945
	$\sum n_i = 16$		$\sum n_i m_i = 49000$		$\sum n_i \log m_i = 55,7528$

Yukarıdaki işlemler üzerinden, hastanede doğan bebeklerin ağırlıklarının geometrik ortalaması,

$$\log GO = \frac{\sum n_i \log X_i}{\sum n_i} = \frac{55,7528}{16} \cong 3,4846$$

$$\text{Antilog}(3,4846) \cong 3052 \text{ gr}$$

şeklinde hesaplanmaktadır.

Bebeklerin doğum ağırlığı ortalamasını belirlemede aritmetik ortalamayı tercih etseydik, ortalama bebek ağırlığını $49000 / 16 = 3062$ gr hesaplayacaktık.

Bu örnekte aritmetik ortalama ile geometrik ortalama arasında çok büyük bir fark bulunmamaktadır. Bunun sebebi, bebeklerin doğum ağırlıklarını gösteren tabloda, gözlem değerleri arasında aşırı büyük değerler bulunmamasıdır.

Bu açıdan bir kez daha hatırlatalım ki, geometrik ortalamanın özellikle seride aşırı değer içeren durumlarda tercih edilmesi anlamlıdır. Aksi halde, daha kolay ve anlaşılır bir yöntem olan aritmetik ortalamanın tercih edilmesi daha doğru bir yaklaşımdır.

Böylelikle analitik ortalamalar başlığı altında geometrik ortalama konusunu da öğrenmiş ve tamamlamış olduk.

6.1.4. Geometrik Ortalamanın Özellikleri

Geometrik ortalama gözlem değerleri arasında aşırı büyük ya da aşırı küçük gözlem değerleri bulunması durumunda, aritmetik ortalamaya kıyasla daha temsili ortalama olma özelliği taşıdığından böyle hallerde tercih edilen bir ortalama olmaktadır.

Geometrik ortalama gözlem değerlerinin logaritmalarının aritmetik ortalamasıdır. Dolayısıyla gözlem değerleri arasında negatif değer ya da sıfır varsa geometrik ortalama hesaplanamamaktadır.

6.2. Kareli Ortalama

Analitik ortalamalar başlığı altında inceleyeceğimiz son ortalamamız da kullanım alanı aritmetik ve geometrik ortalamaya kıyasla çok sınırlı olan, kareli ortalama.

Kareli ortalama, gözlem değerlerinin karesi alınmak suretiyle hesaplanan bir ortalama değildir. Başka bir deyişle kareli ortalama, gözlemlerin karelerinin aritmetik ortalaması alınarak hesaplanan ve son aşamada karekök alma işlemiyle orijinal değerlere dönüşüm sağlandığı bir ortalama değildir.

Kareli ortalama, kullanım alanı oldukça kısıtlı bir ortalama olarak karşımıza çıkmaktadır. Özellikle, bir serinin gözlem değerlerinin toplamı sıfır ise bu durumda aritmetik ortalama kullanılamaz. Benzer şekilde seride negatif değerler bulunması halinde de aritmetik ortalama kullanmak anlamlı bir ortalama hesabına izin vermez. Öte yandan, seride negatif değer bulunması halinde geometrik ortalamanın da kullanımı mümkün değildir. İşte böyle durumlarda analitik bir ortalama kullanılacaksa kareli ortalama tercih edilmektedir.

Kareli ortalama hesabı yapılabilmesi için verinin aralık ya da oran ölçüm düzeyinde ölçülmüş olması gerektiği unutulmamalıdır.

6.2.1. Basit Seride Kareli Ortalama Hesabı

Gözlem değerlerinin küçükten büyüğe doğru sıralandığı en basit veri düzenleme biçimi olan basit seride kareli ortalamanın formülünü yukarıda bahsettiğimiz yaklaşımla aritmetik ortalama formülü üzerinden oluşturmaya çalışalım.

Öncelikle, basit seride aritmetik ortalama formülünün;

$$\bar{X} = \frac{\sum X_i}{n}$$

şeklinde olduğunu hatırlayalım.

Kareli ortalama, gözlem değerlerinin kareleri alınarak hesaplanan bir aritmetik ortalama gibi düşünülebilir demistik. O halde, aritmetik ortalama formülünde X_i yerine X_i^2 yazarak,

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

Kareli ortalamanın karesine ve karekök alma işlemiyle,

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

kareli ortalama formülüne ulaşmış oluruz.

Burada dikkat edilmesi gereken nokta, gözlem değerlerinin kareleri alınarak hesaplanan bir ortalama söz konusu olduğundan, hesaplanan sonucun karekökü alınarak serinin orijinal verisine dönüşün sağlanması gereğidir. Karekök alınmaması halinde hesaplanan sonuç, kareli ortalama değil, kareli ortalamanın karesidir.

Kareli ortalamanın nasıl hesaplandığını bir örnek üzerinden görelim:

Örnek:

Trakya’da bulunan dört ilde aynı gün içinde gözlenen gündüz hava sıcaklıkları aşağıdaki tablo ile düzenlenmiştir. Buna göre, günlük ortalama sıcaklığı kareli ortalama ile hesaplayalım.

İller	Günlük sıcaklık X_i	X_i^2
Edirne	-2	4
Tekirdağ	0	0
Kırklareli	1	1
İstanbul	3	9
		$\sum X_i^2 = 14$

Tablodan da takip edilebileceği gibi, sıcaklık değerlerinin karelerini alıp toplayarak ve daha sonra da çıkan sonucun karekökünü alarak kareli ortalamaya,

$$KO = \sqrt{\frac{\sum X_i^2}{n}} = \sqrt{\frac{14}{4}} \sqrt{3,5} \cong 1,87$$

ulaşmış oluyoruz.

Yukarıdaki örnekte gözlem değerleri arasında negatif değer olması ve sıfır bulunması bu örnekte aritmetik ortalama ve geometrik ortalama kullanılamaması anlamına gelmektedir. Bu tür serilerde kullanılabilecek duyarlı ortalama kareli ortalamadır.

Tasnif edilmiş ve sınıflanmış frekans serilerinde de kareli ortalama hesaplamak mümkündür. İlgili formüller aşağıda verilmektedir.

6.2.2. Tasnif Edilmiş Serilerde Kareli Ortalamanın Hesabı

Tasnif edilmiş serilerde kullanılacak aritmetik ortalama formülü,

$$KO = \sqrt{\frac{\sum n_i X_i^2}{\sum n_i}}$$

biçimindedir.

6.2.3. Sınıflanmış Frekans Serilerinde Kareli Ortalamanın Hesabı

Sınıflanmış frekans serilerinde gözlem değerlerinin değil, oluşturulmuş sınıflarda yer alan kaç adet gözlem değeri bulunduğunun gösterildiğini hatırlayalım. Bu durumda ilk yapılması gereken işlem sınıfları temsil etmek üzere sınıf orta noktalarını (m_i) belirlemek olacaktır.

Sınıf orta noktaları belirlendikten sonra, bu değerler sanki gözlem değerleriymiş gibi düşünülerek tasnif edilmiş seri için uygulanan çözüm, sınıf orta noktaları kullanarak uygulanacaktır.

Dolayısıyla, sınıflanmış frekans dağılımlarında; kareli ortalama çözümlerinde kullanılacak formül,

$$KO = \sqrt{\frac{\sum n_i m_i^2}{\sum n_i}}$$

olmaktadır.

Kareli ortalama; duyarlı ortalamalar içinde kullanım alanı en sınırlı, en spesifik ortalama olduğundan tasnif edilmiş ve sınıflanmış seri uygulamalarına bu başlık altında yer verilmemiştir.

Kareli ortalama bir sonraki bölümde değineceğimiz dağılma ölçüleri başlığı altında bir kez daha karşımıza çıkacak olup bu başlık altında ayrıntılı olarak incelenecektir.

Bu Bölümde Neler Öğrendik?

Bu bölümde analitik ortalamalar kapsamında yer alan geometrik ortalama ile kareli ortalamayı öğrendik.

Gözlem değerleri arasında aşırı büyük değer bulunması halinde geometrik ortalamanın, aritmetik ortalamaya kıyasla daha temsili olduğunu ve bu sebeple aşırı değer içeren serilerde tercih edilen bir ortalama olduğunu öğrendik.

Gözlem değerleri arasında negatif değer bulunması durumunda ya da toplamı sıfır olan bir seri söz konusu olduğunda ise kareli ortalamanın tercih edildiğini öğrendik.

Bölüm Soruları

- 1- Aşağıdaki ortalamalardan hangisi analitik, duyarlı bir ortalama değildir?
 - a- Mod
 - b- Medyan
 - c- Geometrik ortalama
 - d- Ortanca
 - e- Egemen değer
- 2- Aşağıdaki ortalamalardan hangisi birimlerin logaritmaları üzerinden hesaplanır?
 - a- Geometrik ortalama
 - b- Kareli ortalama
 - c- Aritmetik ortalama
 - d- Medyan
 - e- Mod
- 3- Aşağıdaki ortalamalardan hangisi birimlerin kareleri üzerinden hesaplanır?
 - a- Mod
 - b- Medyan
 - c- Kareli ortalama
 - d- Geometrik ortalama
 - e- Aritmetik ortalama
- 4- Toplamı sıfır olan ya da negatif (–) işaretli değer bulunduran serilerde ortalama hesaplanmak istendiğinde kullanılır.
 - a- Kareli ortalama
 - b- Aritmetik ortalama
 - c- Geometrik ortalama
 - d- Toplamı sıfır olan seride ortalama hesaplanmaz.
 - e- Negatif değer bulunan seride ortalama hesaplanamaz.

5- Geometrik dizi özelliği gösteren ya da aşırı değer içeren serilerde hangi analitik ortalamanın kullanılması daha uygundur?

- a- Kareli ortalama
- b- Aritmetik ortalama
- c- Geometrik ortalama
- d- Mod
- e- Medyan

6- Aşağıdaki serinin geometrik ortalaması nedir?

X_i
50
100
150
200

- a- 125
- b- 120,25
- c- 120
- d- 115,55
- e- 110,66

7- Gözlem değerleri 2, 5, 9 olan bir serinin kareli ortalaması nedir?

- a- 36,67
- b- 5,33
- c- 7,22
- d- 6,055
- e- 4,48

- 8- Geometrik ortalama hesaplayabilmek için verinin en azölçüm düzeyinde ölçülmüş olması gerekir.
- a- Nominal
 - b- Sınıflayıcı
 - c- Sıralayıcı
 - d- Aralık
 - e- Ordinal
- 9- Kareli ortalama hesaplayabilmek için verinin veya ölçekte ölçülmüş olması gerekir.
- a- Sınıflayıcı-sıralayıcı
 - b- Aralık-oran
 - c- Nominal-ordinal
 - d- Kategorik-ordinal
 - e- Nominal-sıralayıcı
- 10- Bir ilde Mart ayı sıcaklıkları -2, 1, -1, 2 şeklinde ölçülmüştür. Bu serinin ortalaması hesaplanmak istendiğinde hangi duyarlı, analitik ortalamanın kullanılması uygun olacaktır?
- a- Aritmetik ortalama
 - b- Ağırlıklı aritmetik ortalama
 - c- Geometrik ortalama
 - d- Medyan
 - e- Kareli ortalama

Cevaplar

- 1- c
- 2- a
- 3- c
- 4- a
- 5- c
- 6- e

7- d

8- d

9- b

10- e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

7. VERİLERİN ÖZETLENMESİ: ANALİTİK OLMAYAN ORTALAMALAR

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Daha önceki bölümlerde analitik ortalamalar başlığı altında aritmetik, geometrik ve kareli ortalamaların nasıl hesaplandığını ve hangi durumda hangi ortalamanın tercih edilmesi gerektiğini öğrendik.

Bu bölümde de analitik olmayan ortalamalar başlığı altında yer alan mod ve medyanın nasıl hesaplandığını ve kullanıldığını öğreneceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Ortalama hesabında bütün birimleri kullanmak zorunlu mudur?
- 2) Sadece ortalama hesaplanmak istendiğinde, kolay yolla hesaplanabilecek ortalamalar var mıdır?
- 3) Bir seride en çok tekrar eden değer ortalama olarak alınabilir mi?
- 4) Gözlem değerlerini sıralasak, tam ortadaki değer ortalama kabul edilebilir mi?
- 5) Analitik çözüme dayanmayan ortalamalar temsili olabilir mi?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Analitik olmayan ortalamalar	Analitik olmayan ortalama kavramını bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik olmayan ortalamalar ile ilgili kısımlarından yararlanılabilir
Analitik olmayan ortalamalar ve temsili olma	Analitik olmayan ortalama kavramı ile temsili ortalama hesaplamak ilişkisini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik olmayan ortalamalar ile ilgili kısımlarından yararlanılabilir.
Mod (Egemen değer)	Seride en çok tekrar eden değer olarak modu ve özelliklerini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik olmayan ortalamalar ile ilgili kısımlarından yararlanılabilir.
Medyan (Ortanca)	Sıralanmış bir seride tam ortada yer alan değer olarak medyanı ve özelliklerini bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, analitik olmayan ortalamalar ile ilgili kısımlarından yararlanılabilir
Aritmetik ortalama, Mod ve Medyanın aldığı değere bakarak serinin simetrisini belirleme	Aritmetik ortalama, mod ve medyanın aldığı değerlerin dizilişinin serinin simetrik, pozitif ve negatif asimetric olmasını belirlemede nasıl kullanıldığını bilir.	Verilerin özetlenmesine yönelik bilgilerin yer aldığı temel düzeyde istatistik kitaplarından, ortalamalara dayalı eğiklik ölçüleri ilgili kısımlarından yararlanılabilir

Anahtar Kavramlar

- Analitik olmayan ortalamalar: ortalama hesabında tüm gözlem değerlerinin değil, bazı gözlem değerlerinin kullanıldığı ve duyarsız ortalamalar olarak adlandırılan ortalamaların genel adı.
- Mod: bir seride en çok tekrar eden değer.
- Medyan: sıralanmış seride tam ortada yer alan değer.

Giriş

Analitik olmayan ortalamalar; bir seride yer alan bütün gözlem değerleri üzerinden hesaplanmayan, başka bir deyişle hesabında seride yer alan bütün gözlem değerlerinin kullanılmadığı ortalamalardır.

Analitik olmayan ortalamalar, seride yer alan bütün birimlerden hesaplanmadığı için, ortalama hesabına dâhil edilmeyen birimlere karşı duyarsızdır. Bu sebeple bu ortalamalara duyarsız ortalamalar da denmektedir.

Analitik olmayan ortalamalar başlığında Mod ve Medyan'ı öğreneceğiz.

7.1. Mod (Egemen Değer)

Analitik olmayan ortalamalar grubu içinde yer alan mod, seride en çok tekrar eden değeri ifade etmektedir. Yani bir seride en çok tekrar eden değer, varsa, serinin modu o değerdir.

Mod, moda sözcüğü ile ilişkilendirilebilir. Nasıl ki moda; yaygınlık, çokça görülme anlamını özünde barındırıyorsa mod da çokça görülen anlamında bir ortalama olarak düşünülebilir.

Mod'un basit seride ve tasnif edilmiş serilerde hesabı yoktur. Zira, yapılması gereken her birimin kaç kez tekrar ettiğine bakmak ve en çok tekrar eden değeri mod değeri olarak almaktır.

Her serinin modu olmayacağı gibi, bazı seriler de birden fazla moda sahip olabilir. Örneğin, bir seride her bir gözlem değerinden birer tane varsa, en çok tekrar etme ya da en çok gözlenen olma durumu hiçbir gözlem değeri için geçerli değildir. Böyle seriler için, modu yoktur denir.

Benzer şekilde bazı serilerde de aynı sayıda gözlenen yani frekansı ya da tekrar sayısı aynı olan birden fazla gözlem değeri bulunmaktadır. Bu tür serilere de çok modlu seriler denmektedir.

7.1.1. Basit Seride Mod'un Belirlenmesi

Basit seriler, gözlem değerlerinin küçükten büyüğe doğru sıralanmasıyla elde edildikleri için, bu diziliş içinde en çok tekrar eden gözlem değerinin ne olduğuna bakılarak mod belirlenir. En çok tekrar eden gözlem değeri yoksa, yani her birimden birer tane varsa modun olmadığına karar verilir ya da birden fazla aynı sayıda tekrar eden gözlem değeri varsa birden fazla mod olduğu söylenir.

Şimdi bu anlattıklarımızı basit bir örnek üzerinden göstermeye çalışalım:

Örnek:

Aşağıda, üç öğrencinin bir dönemde aldığı 6 dersin on üzerinden dönem sonu başarı notları verilmektedir. Buna göre, öğrencilerin mod değerleri nedir?

	Başarı Notları		
Dersler	Ahmet	Mehmet	Umut
Matematik	9	6	5
Fen Bilgisi	9	6	6
Sosyal Bilgiler	8	7	7
Din Kül. Ve Ahlâk Bil.	8	8	8
Resim	10	9	9
Müzik	9	9	10
	Mod=9	Mod= 6 ve 9	Mod yoktur

Şimdi öğrencilerin notlarına ayrı ayrı bakalım:

Ahmet üç dersinden 9, iki dersinden 8 ve bir dersinden de 10 almıştır. Buna göre, en çok gözlenen ya da tekrar eden not 9 olduğu için Ahmet'in notlarının modu 9'dur.

Mehmet'in iki adet 6'sı, iki adet 9'u, bir adet 7'si ve bir adet 8'i vardır. Dolayısıyla Mehmet açısından en çok gözlenen not 6 ve 9 olmaktadır ve her ikisi de ikişer tanedir. Mehmet'in notları çift modludur ve mod değerleri 6 ve 9'dur.

Umut'un notlarına baktığımızda ise, bütün notlardan sadece birer tane olduğu ve hiçbir notun tekrar etmediği anlaşılmaktadır. Dolayısıyla, Umut'un notlarının modu yoktur.

7.1.2. Tasnif Edilmiş Seride Modun Belirlenmesi

Tasnif edilmiş serilerde, gözlem değerleri ile her gözlem değerinin kaç kez tekrar ettiğini gösteren frekans sütunundan oluştuğunu hatırlayalım. Frekans sütununa bakarak tespit edeceğimiz en yüksek frekans değeri bize modun bu frekansa sahip olan gözlem değeri olduğunu söyleyecektir.

Tasnif edilmiş seriler, tekrarlı ve çok sayıda gözlem değeri bulunması halinde tercih edildiklerinden, bu tür serilerde mod bulunmaması söz konusu olmamaktadır. Ancak tasnif edilmiş serilerde aynı frekansa sahip birden fazla sayıda gözlem değeri bulunabileceğinden çok modlu seriler görülebilmektedir.

Aşağıda, anlattığımız bu durumlara yönelik örnek çözümler verilmiştir:

Örnek:

Aşağıda ayakkabı üreterek satan bir firmanın kadın ve erkekler için hangi ayakkabı ölçüsünde kaç adet satış yaptığı görülmektedir. Buna göre, kadın ve erkek için ayakkabı ölçülerinin modu nedir?

KADIN		ERKEK	
Ayakkabı Ölçüsü X_i	Satılan Miktar n_i	Ayakkabı Ölçüsü X_i	Satılan Miktar n_i
35	120	40	150
36	180	41	175
37	355	42	230
38	640	43	455
39	450	44	455
40	320	45	45
Mod= 38		Mod= 43 ve 44	

Ayakkabı satış değerlerini tablodan inceleyelim.

Kadın ayakkabısı grubunda en yüksek frekans 640 olup, 38 numara için geçerlidir. Yani firma, en çok 38 numara ayakkabı satmıştır ve dolayısıyla, kadın ayakkabısı için mod değeri 38'dir.

Erkek ayakkabısı grubunda ise satış değerlerini incelediğimizde, en yüksek frekans değerinin 455 olduğunu ve bunun da 43 ve 44 numaralı ayakkabılarda gözlemlendiğini görüyoruz. Bu durumda, erkek ayakkabısı için iki mod söz konusudur ve mod değerleri 43 ve 44'dür.

7.1.3. Sınıflanmış Frekans Serilerinde Modun Belirlenmesi

Sınıflanmış serilerde hatırlanacağı gibi, gözlem değerlerinin sınıflandığı sınıflar ve her bir sınıfta kaç gözlem değeri bulunduğunu gösteren frekanslar olmak üzere veriler iki sütun ile gösterilmektedir. Mod değerinin ne olduğunu belirlemede yine ilk olarak frekans sütununda en yüksek değeri bulmak gerekmektedir. Frekans sütunundaki en yüksek değer, ait olduğu sınıfın en yüksek frekansa sahip sınıf olduğunu, başka bir deyişle en çok gözlem değerinin bu sınıfta yer aldığını gösterecektir.

Dolayısıyla, frekans sütununda en yüksek değeri bularak, mod sınıfını yani modun içinde yer aldığı sınıfı kolaylıkla belirlemek mümkündür. Pekala, bu durumda mod sınıfı içinde yer alan hangi değer mod değeridir?

Mod sınıfı içinde hangi değer mod değeri olduğunu belirlemede ilk akla gelen şüphesiz mod sınıfının orta noktasıdır. Yani, mod sınıfının alt ve üst sınır değerlerini toplayıp ikiye bölerek hesaplanabilecek olan sınıf orta noktasını, mod değeri olarak kabul etmek ilk yaklaşım olabilir. Ancak, uygulamada mod değeri bu şekilde belirlenmemektedir.

Mod değerinin, mod sınıfı içinde hangi değer olabileceğini tespit etmede basit bir varsayım yapılmakta ve modun tespiti bu varsayım üzerine kurgulanmaktadır.

Mod sınıfı içinde yer alan gözlem değerlerinin, mod sınıfından önceki ve sonraki frekanslara bakılarak, hangi frekans daha yüksek ise, mod sınıfı içindeki gözlem değerlerinin yüksek olan frekansın olduğu sınıfa doğru daha yoğun olarak yerleştiği varsayılmaktadır.

Bu basit varsayım ile oluşturulan sınıflanmış frekans dağılımlarında modun ne olduğunu belirlemeye yönelik olarak geliştirilmiş formül,

$$MOD = l_{alt} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot c$$

biçimindedir.

Formülde yer alan sembollere baktığımızda,

l_{alt} mod sınıfının alt sınır değerini,

Δ_1 , mod sınıfının frekansı – bir önceki sınıfın frekansını,

Δ_2 , mod sınıfının frekansı – bir sonraki sınıfın frekansını ve

c ise sınıflanmış dağılımdaki sınıf genişliğini

ifade etmektedir.

Sınıf genişliğini gösteren c sembolünün altında herhangi bir indis bulunmamaktadır. Dolayısıyla, c dağılımdaki sınıf genişliğini ifade etmekte ve sınıf genişliğinin eşit olduğunu da örtülü olarak göstermektedir.

Sınıflanmış frekans serilerinde mod hesabı yapılabilmesinin ön koşulu, sınıfların eşit aralıklı düzenlenmiş olmasıdır. Zira, sadece eşit aralıklı olarak düzenlenmiş serilerde bir sınıfta diğerlerine kıyasla daha çok gözlem değeri bulunduğunu söylemek anlamlı olacaktır.

Bir an için elimizde sınıf genişlikleri eşit olmayan bir dağılım bulunduğunu düşünelim. Böyle bir dağılımda, sınıf genişliği daha fazla olan sınıfa sınıf genişliği daha dar olan sınıfa kıyasla daha fazla gözlem değeri düşmesi olağan ve beklenen bir durumdur. Böyle bir durumda modun sınıf genişliği fazla olan sınıfta çıkması tamamen sınıf genişliğinin fazla olmasından kaynaklanır.

Özetleyecek olursak, sınıflanmış frekans dağılımlarında, modun anlamlı olabilmesi ve mod hesabı yapılabilmesi için sınıf genişliğinin eşit olması şartı vardır. Sınıf genişliği eşit olmayan serilerde mod hesaplanamaz.

Sınıflanmış dağılımlarda modun nasıl hesaplandığını bir örnek çalışmayla görelim:

Örnek:

İstatistik dersini ilk defa alan 120 öğrencinin notları aşağıdaki sınıflanmış dağılım ile düzenlenmiştir. Öğrencilerin notları için mod değerini hesaplayalım.

Not Sınıfları	Frekanslar
0-20	12
20-40	21
40-60	28
60-80	46
80-100	13
Toplam	120

Yukarıdaki sınıflanmış dağılımın frekans sütununa baktığımızda en yüksek frekansın 46 olduğunu ve bu frekansın da 60-80 sınıfına ait olduğunu görüyoruz. Dolayısıyla, mod sınıfımız 60-80 sınıfıdır ve en çok not alan öğrencinin bulunduğu sınıf bu sınıftır.

60-80 sınıfı içinde mod hangi değer olabilir?

Daha önce yaptığımız temel varsayımı hatırlayalım. Mod sınıfından bir önceki ve bir sonraki sınıfların frekanslarına bakılarak hangi frekans daha yüksekse modun o sınıfa daha

yakın olduğu kabul edilir. Bu örnekte, mod sınıfından bir önceki sınıfın frekansı 28; mod sınıfından bir sonraki sınıfın frekansı ise 13'tür. Bu durumda, 60-80 sınıfı içinde yer alan mod frekansı daha yüksek olan önceki sınıfa yaklaşacaktır.

Mod formülünü kullanarak modu belirleyelim:

$$MOD = l_{alt} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot c$$

$$MOD = 60 + \frac{(46 - 28)}{(46 - 28) + (46 - 13)} \cdot 20$$

$$MOD = 60 + \frac{18}{18 + 33} \cdot 20$$

$$MOD = 60 + \frac{18}{51} \cdot 20$$

$$MOD \cong 67,06$$

Görüldüğü gibi, yukarıdaki seri için mod 67,06 olarak hesaplanmıştır. Modun, 60-80 sınıfı içinde yer alacağını başlangıçta öngörmüştük ve önceki sınıfın frekansının daha yüksek olması sebebiyle modun alt sınır değerine doğru meyledeceğini de ifade etmiştik. Nitekim 67,06 değeri de bunu doğrulamaktadır.

Sınıflanmış dağılımlarda mod bulunmaması durumu söz konusu değildir. Zira, çok sayıda gözlem değeri içeren dağılım türü olması sebebiyle bir sınıfın frekansının diğer sınıfların frekanslarından daha fazla olması olasıdır. Ancak bazen, en yüksek frekans iki farklı sınıfta olabilir. Yani iki farklı sınıfın frekansları aynı olup serideki en yüksek frekansı oluşturuyor olabilir. Böyle durumlarda, frekansları eşit olan bu iki sınıf birbirinden uzak sınıflar ise o takdirde serinin çift modlu seri olduğuna karar verilir. Dolayısıyla seri, biri düşük değerlerde diğeri yüksek değerlerde olmak üzere iki moda sahip olacaktır. Sözgelimi, bir lisede eğitim gören öğrencilerin boy uzunlukları kız-erkek öğrenci ayrımı gözetmeksizin aynı seride gösterilirse, kız öğrencilerin ve erkek öğrencilerin boylarının mod değerleri ayrı ayrı ortaya çıkacak ve seri iki modlu bir görüntü verecektir. Böyle bir durumda, veriyi kızlar ve erkekler olarak ayırmak daha doğru bir yaklaşım olacaktır ki böylelikle her iki seri de tek modlu seriye dönüşecektir.

Ancak, bazı durumlarda da mod sınıflarının ard arda iki sınıf şeklinde ortaya çıktığı da görülmektedir. Başka bir deyişle, en yüksek frekans arka arkaya gelen iki sınıfa ait olmaktadır. Bu tür serilere, en yüksek frekanslı sınıflar ardışık sınıflar olduğundan, çift modlu seri denmemektedir. Böyle durumlarda, sınıf aralığı değiştirilmek suretiyle serinin tek bir mod sınıfının olması sağlanmalıdır. Aksi halde seriyi, çift modlu bir seri gibi düşünmek hatalı bir yaklaşım olacaktır.

A graph showing a bimodal distribution curve. The curve has two distinct peaks of equal height and a central dip, indicating two different modes or clusters within the data.

148

Medyan bir seride yer alan gözlem değerlerinin küçükten büyüğe ya da büyükten küçüğe doğru sıralanması sonucunda oluşan sıralamada tam ortada yer alan değer olarak tanımlanmaktadır ve seriyi iki eşit parçaya bölen değer olması bakımından da $Q_{1/2}$ şeklinde gösterilmektedir.

Öte yandan, seriyi tam ortadan iki eşit kısma ayıran ve bu sebeple serinin tam ortasında bulunan değer olması nedeniyle medyan, *ortanca* ya da *ortalayan* olarak da bilinmektedir.

Medyan, sıralanmış bir seride tam ortadaki değer olduğundan verinin sıralanabilir özellik taşıması gerekir. Bu açıdan bakıldığında, mod'a göre kullanım alanının daha sınırlı olduğunu söylemek mümkündür. Örneğin cinsiyet vasfının sıklıkları olan kadın – erkek sonuçları sıralanabilir sonuçlar değildir. Dolayısıyla cinsiyet ortalaması için medyan kullanılamaz. Benzer şekilde İstanbul Trafik Sicil Müdürlüğünde kayıtlı araçların ortalaması hesaplanmak istendiğinde, en çok frekansa sahip olan aracın mod olduğu kolaylıkla tespit edilerek söylenebilir. Ancak, araçlar sıralanabilir özellik göstermediğinden bu veri için medyan hesaplanamaz.

Medyan hesaplayabilmek için serinin mutlaka sıralanmış olması gerektiğini bir kez daha vurgulamakta fayda görüyoruz. Sıralanmamış veri söz konusu ise medyan hesaplanamaz.

7.2.1. Basit Seride Medyan Hesabı

Basit seri özelliği gereği verilerin küçükten büyüğe ya da büyükten küçüğe sıralandığı seridir. Böyle bir seride sıralamanın tam ortasında yer alan değer medyandır.

Basit seride medyanın hangi sırada yer alan değer olduğunu belirlemek üzere aşağıdaki basit formül kullanılmaktadır:

$$\text{Medyanın sıra değeri} = \frac{n + 1}{2}$$

Yukarıdaki basit formül bize medyanın sıra değerini verir. Bu sıradaki değer ne olduğunu tespit ettiğimizde medyana ulaşmış oluruz.

Şüphesiz yukarıdaki formülle elde edilen sıra değerinin tam sayı olması durumunda medyan kolaylıkla tespit edilebilir. Serideki gözlem sayısı 3, 5, 7 vb. tek sayı ise bu durumda $(n+1)/2$ değeri daima tam sayı olacak ve medyanın sıra değeri de kolaylıkla belirlenecektir.

Ancak serideki gözlem sayısı 2, 4, 6 gibi çift sayı ise $(n+1)/2$ değeri tam sayı olmayıp ondalık bir değer alacaktır. Sıra numaraları kesikli değerler olup sadece tam sayı şeklinde ifade edilebilir. Bu sebeple, çift sayıda gözlem içeren serilerde bulunan ondalıklı sıra değerinin öncesi ve sonrasındaki tam sayı şeklindeki sıra değerlerinde yer alan gözlemlerin ortalaması alınarak medyana ulaşılır.

Gerek tek gerekse çift gözlem sayısı içeren serilerde medyanın nasıl belirlendiğini bir örnek ile inceleyelim:

Örnek:

Aşağıda Ahmet ve Mehmet isimli iki öğrencinin dönem sonu başarı notları verilmiştir. Ahmet dönem boyunca 7 ders almıştır. Mehmet ise rahatsızlığı sebebiyle Beden Eğitimi dersinden muafır ve dolayısıyla 6 ders almaktadır.

Ahmet ve Mehmet'in notlarının medyan değerlerini hesaplayalım.

	Başarı Notları	
Dersler	Ahmet	Mehmet
Matematik	9	6
Fen Bilgisi	9	6
Sosyal Bilgiler	8	7
Din Kül. Ve Ahlak Bil.	8	8
Resim	10	9
Müzik	9	9
Beden Eğitimi	8	Muaf

Öncelikle, tabloda yer alan notların sıralanmamış olduğuna dikkat edelim. Dolayısıyla bu şekilde medyanyı bulmak mümkün değildir ve öncelikle notların küçükten büyüğe doğru sıralanması sağlanmalıdır.

Önce toplam 7 ders alan Ahmet için medyanyı belirleyelim. Ahmet'in notlarını sıraladığımızda aşağıdaki tablo ortaya çıkıyor:

Dersler	Başarı Notları	Sıra no:
Sosyal Bilgiler	8	1.
Din Kül. Ve Ahlâk Bil.	8	2.
Beden Eğitimi	8	3.
Matematik	9	4.
Fen Bilgisi	9	5.
Müzik	9	6.
Resim	10	7.
	n=7	

Bu durumda, medyanın bulunduğu sıra değeri;

$$\text{Medyanın sıra değeri} = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4. \text{ sıradaki değer}$$

olarak belirlenmektedir.

Tabloda 4. sıradaki değerin 9 olduğunu kolaylıkla tespit edebiliyoruz. Bu durumda, Ahmet'in notlarının medyan değeri 9 dur. Yani, $Q_{1/2} = 9$ şeklinde medyanyı gösterebiliriz.

Şimdi de Mehmet'in derslerini sıralayalım:

Dersler	Başarı Notları	Sıra no:
Matematik	6	1.
Fen Bilgisi	6	2.
Sosyal Bilgiler	7	3.
Din Kül. Ve Ahlâk Bil.	8	4.
Resim	9	5.
Müzik	9	6.
	n=6	

Mehmet, Beden Eğitimi dersinden muaf olduğundan toplam 6 ders almaktadır. Dolayısıyla, medyanın sıra değeri;

$$\text{Medyanın sıra değeri} = \frac{n+1}{2} = \frac{6+1}{2} = \frac{7}{2} = 3,5 \text{ .sıradaki değer}$$

3,5. sıra değeri olarak belirlenmektedir. 3,5 değeri 3. ve 4. değerlerin ortasındadır. Dolayısıyla, medyanı bulmak için 3. ve 4. gözlem değerlerini bulup ortalamasını almamız gerekecektir. Tablodan görüleceği gibi, 3. gözlem değeri 7, 4. gözlem değeri ise 8'dir. Dolayısıyla medyan,

$$Q_{1/2} = \frac{7+8}{2} = 7,5$$

olarak hesaplanmaktadır.

7.2.2. Tasnif Edilmiş Seride Medyan Hesabı

Tasnif edilmiş seriler, bir seri içinde tekrar eden gözlem değerlerinin bir araya toplanarak tasnif edilmesi ve her gözlemin tekrar sayısının frekans olarak nitelenerek düzenlenmesiyle elde edilen serilerdir.

Dolayısıyla, tasnif edilmiş seriler sıralanmış serilerdir ve bu sıralı seride tam ortada yer alan değeri tespit etmek medyanı belirlemek için yeterli olacaktır.

Basit seride medyanın belirlenmesine yönelik olarak verdiğimiz,

$$\text{Medyanın sıra değeri} = \frac{n+1}{2}$$

formülü burada da geçerlidir.

Gözlem sayısının tek sayı olması halinde $(n+1)/2$ formülüyle elde edilen sırada yer alan gözlem değeri bize doğrudan medyanı vermektedir. Gözlem sayısının çift sayı olması halinde ise $(n+1)/2$ formülüyle elde edilen sıra değeri ondalık bir değer olacağından, bu ondalık değer

öncesinde ve sonrasında bulunan sıralarda yer alan gözlem değerlerinin ortalaması alınarak medyana ulaşılacaktır.

Örnek:

Aşağıda, bir ayakkabı üreterek satışını yapan bir firmanın, bir sezon süresince sattığı bayan ayakkabılarının ayak ölçülerine göre dağılımı verilmektedir. Medyan değerini bulalım.

Ayakkabı Ölçüsü X_i	Satılan Miktar n_i	Birikimli Frekanslar	
35	120	120	120
36	180	300	120+180=300
37	355	655	300+355=655
38	640	1295	655+640=1295
39	450	1745	1295+450=1745
40	320	2065	1745+320=2065
	$\sum n_i = 2065$		

Tablodan da görüleceği gibi, serideki toplam gözlem sayısı 2065’dir. Bu durumda, medyanın sıra değerini bulmak için ilgili formülü kullanarak, medyanın sıra değerini,

$$\text{Medyanın sıra değeri} = \frac{n + 1}{2} = \frac{2065 + 1}{2} = 1033$$

olarak hesaplıyoruz. O halde, serideki 1033. değer hangi değerdir?

Tasnif edilmiş serilerde frekans sütununda gözlem değerlerinin tekrar sayıları verildiğinden hesaplanan sıra değerinde yer alan medyan değerini bulmak kolay olmamaktadır. Bunun için birikimli frekansların oluşturulması gerekecektir.

Birikimli frekans oluşturma işlemi, ilk frekans değeri ile başlayıp, her aşamada bir sonraki frekans değerinin üzerine ilave edilmesi, yani üst üste biriktirilmesiyle aşama aşama gerçekleştirilen bir işlemdir. Yukarıdaki tabloda son sütunda birikimli frekanslar verilmiştir.

Daha açıklayıcı olması ve kolay kavranması bakımından biriktirme işlemi aşağıda bir kez daha aşama aşama gösterilmiştir:

Ayakkabı Ölçüsü X_i	Satılan Miktar n_i	Birikimli Frekanslar	
35	120	120	120
36	180	300	120+180=300
37	355	655	300+355=655
38	640	1295	655+640=1295
39	450	1745	1295+450=1745
40	320	2065	1745+320=2065
	$\sum n_i = 2065$		

Birikimli frekans sütununun ilk gözlem değerinin frekansı ile başladığına ve toplam frekansa eşit bir değerle sonlandığına dikkat ediniz. Bunun sebebi son aşamada seride yer alan bütün gözlem değerlerinin toplanmış olmasıdır.

Örneğimize dönersek, 2065 gözlem değeri içeren bu seride 1033. değerin medyan olduğunu belirlemiştik. Şimdi, birikimli frekans sütununda 1033. değeri kapsayan birikimli frekansı bulalım. 1033. değer, 1245 birikimli frekans değerinin yer aldığı gözlem değeri olan 38 numaralı ayak ölçüsünü göstermektedir.

Dolayısıyla, yukarıdaki serinin medyanı,

$$Q_{1/2} = 38$$

olarak tespit edilmiştir.

7.2.3. Sınıflanmış Seride Medyan Hesabı

Sınıflanmış seride medyan hesabı da, tasnif edilmiş seride olduğu gibi öncelikle tam ortada yer alan birimin kaçınıcı sıradaki birim olduğunun belirlenmesi ile yapılır.

Sınıflanmış bir dağılım söz konusu olduğunda tam ortadaki birimin hangi birim olduğu değil, bu birimin hangi sınıfın içinde yer aldığı belirlenebilir. Dolayısıyla ilk olarak, medyan değerinin yer alacağı sınıf tespit edilerek bu sınıf medyan sınıfı olarak adlandırılır.

Daha sonra da aşağıdaki formül yardımıyla medyanın bu sınıf içindeki değerinin ne olabileceği hesaplanır.

Sınıflanmış bir frekans dağılımının medyanı,

$$Q_{1/2} = l_{alt} + \frac{\frac{n}{2} - n_a}{n_q} \cdot c_q$$

formülüyle hesaplanır.

Formülde yer alan sembollere baktığımızda,

l_{alt} medyan sınıfının alt sınır değerini,

$\frac{n}{2}$, medyanın sıra değeri, yani kaçınıcı birim olduğu,

n_a , medyan sınıfından bir önceki sınıfın birikimli frekansı,

n_q , medyan sınıfının frekansı,

c_q ise medyan sınıfının sınıf genişliğini

ifade etmektedir.

Yukarıdaki formülde medyanın kaçınıcı değeri olduğu $n/2$ şeklinde hesaplanmaktadır. Hatırlanacağı gibi, basit seride ve tasnif edilmiş seride ise $(n+1)/2$.sırada yer alan değeri medyan kabul etmiştik. Sınıflanmış frekans dağılımında $n/2$.sıradaki değeri medyan kabul etmemizin sebebi, birimlerin kendisinin değil bulundukları sınıfları görebilmektir. Dolayısıyla, medyanın sıra değerini $n/2$ veya $(n+1)/2$ almanın sonuç üzerinde önemli bir etkisi olmamaktadır.

Örnek:

İstatistik dersini ilk defa alan 120 öğrencinin notları aşağıdaki sınıflanmış dağılım ile düzenlenmiştir. Öğrencilerin notları için medyan değerini hesaplayalım:

Not Sınıfları	Frekanslar	Birikimli Frekanslar
0-20	12	12
20-40	21	12+21=33
40-60	28	33+28=61
60-80	46	61+28=107
80-100	13	107+13=120
Toplam	120	

Yukarıdaki tabloda toplam 120 gözlem değeri bulunmaktadır ve bu durumda medyanın sıra değeri $n/2 = 120/2 = 60$ olmaktadır.

Tablonun sağ sütununda birikimli frekanslar görülmektedir. Birikimli frekans sütununda 60. değerin hangi sınıfta olduğunu araştırıyoruz. 61. birim 40-60 sınıfında olduğuna göre, 60. birim de aynı sınıftadır. Dolayısıyla, medyan sınıfı 40-60 sınıfıdır.

Medyanın kaç eşit olduğunu bulabilmek için yukarıdaki verilen formülü kullanıyoruz:

$$Q_{1/2} = l_{alt} + \frac{\frac{n}{2} - n_a}{n_q} \cdot c_q$$

$$Q_{1/2} = 40 + \frac{60 - 33}{28} \cdot 20$$

$$Q_{1/2} = 40 + \frac{27}{28} \cdot 20$$

$$Q_{1/2} = 40 + 19,29 = 59,29$$

Medyan notu 59,29 olarak hesaplıyoruz.

7.2.4. Medyanın Özellikleri

Medyan, değerleri küçükten büyüğe doğru sıralanmış bir seride tam ortada yer alan değeri ifade etmektedir. Dolayısıyla, bir veride medyan hesaplayabilmek için serinin sıralanabilir sıklara sahip bir vasıf için düzenlenmiş bir seri olması gerekecektir. Aksi takdirde, medyan hesaplamak mümkün değildir.

Medyan, analitik olmayan yani duyarsız ortalamalar grubu içinde yer alan bir ortalama değildir. Medyan hesabında serideki tüm değerler dikkate alınmadığından, cebirsel işlemlere elverişli bir ortalama değildir. Dolayısıyla, amaç bir seriyi özetlemek ya da tek bir değer ile temsil etmek ise medyan uygun bir ortalama olarak düşünülebilir. Ancak, veriye yönelik bir takım istatistiksel analizler yapılması planlanıyorsa, cebirsel işlemlere elverişli bir ortalama olmaması sebebiyle kullanılamayacaktır.

Medyan, küçükten büyüğe doğru sıralanmış bir seride tam ortadaki değere karşılık geldiğinden, seride yer alan aşırı küçük ya da aşırı büyük değerler sıralamanın başında ve sonunda yer alacağından medyan üzerinde etkili olmayacaktır. Medyan, özellikle aşırı değer içeren serilerde bu özelliği bakımından en temsili ortalama olmaktadır.

7.3. Ortalamalar Arasındaki İlişkiler ve Bir Dağılımın Simetri Durumunun Değerlendirilmesi

Ortalamalar başlığı altında, analitik (duyarlı) ortalamalar kapsamında, aritmetik, geometrik ve kareli ortalamayı; analitik olmayan (duyarsız) ortalamalar başlığı altında da mod ve medyanı öğrendik.

Aritmetik, geometrik ve kareli ortalama serinin değerlerinin farklı versiyonlarını referans olarak hesaplandıklarından, sonuçları birbirinden farklı olmaktadır.

Örneğin, aritmetik ortalama gözlem değerlerinin kendisini, geometrik ortalama gözlem değerlerinin logaritmalarını ve kareli ortalama ise gözlem değerlerinin karelerini referans olarak hesaplanmaktadır.

Ortalamaların referans aldıkları değerlerin büyüklük sıralaması tahmin edileceği gibi,

$$\log X_i < X_i < X_i^2$$

şeklinde olmaktadır. Nitekim, bir değerin logaritması daima kendisinden küçük, - tam sayı olması halinde- karesi ise kendisinden büyüktür.

Bu diziliş ortalamalar için de söz konusudur. Bir serinin ortalamaları,

$$\text{Geometrik Ortalama} < \text{Aritmetik Ortalama} < \text{Kareli Ortalama}$$

$$GO < \bar{X} < KO$$

şeklinde büyüklük-küçüklük ilişkisi içinde bulunmaktadır.

Sözü edilen üç analitik ortalama, serideki birimlerin homojen dağılıma sahip olduğu ölçüde birbirine yaklaşır, homojenlikten uzaklaştığı ölçüde birbirinden uzaklaşır. Nitekim, seride aşırı büyük değerlerin bulunması durumunda aritmetik ortalamanın bu durumdan çok etkilenecek, büyük değerlere doğru kaydığını ifade etmiştik. Seride bulunan aşırı değerlerin kareli ortalamaya etkisi daha da fazla olmaktadır. Zira kareli ortalama, gözlem değerlerinin kareleri ile hesaplanan bir ortalama ve aşırı büyük gözlem değerlerinin karelerinin kullanılması ortalamaya olan etkisini çok daha fazla kılmaktadır.

Benzer şekilde, geometrik ortalama, gözlem değerlerinin logaritmaları kullanılarak hesaplandığından, değerler küçülmekte, aşırı değerlerin geometrik ortalamaya etkisi de aritmetik ortalamaya kıyasla daha az olmaktadır.

Bu başlıkta son olarak bir seride Aritmetik Ortalama, Mod ve Medyanın aldığı değerlere bakarak serinin simetri durumuyla ilgi nasıl bilgi sahibi olunabileceği üzerinde duracağız.

Simetrik Dağılım:

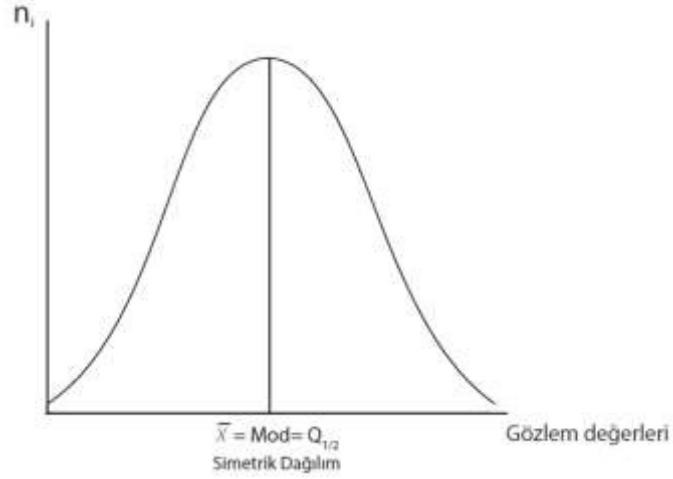
Bir dağılımın simetrik olması halinde, Aritmetik Ortalama, Mod ve Medyan birbirine eşit değer alır.

Başka bir deyişle,

$$\text{Aritmetik Ortalama} = \text{Mod} = \text{Medyan}$$

$$\bar{X} = \text{Mod} = Q_{1/2}$$

eşitliği söz konusu ise, dağılımın simetrik bir dağılım olduğu anlaşılır ve dağılımın grafiği aşağıdaki gibi olur:



Grafik 7.3: Simetrik Dağılım

Bir dağılım simetrik olma halinden uzaklaştığı ölçüde bu ortalamalar birbirinden uzaklaşır. Ortalamaların birbirinden uzaklaştığı ölçüde de dağılımın asimetri durumu artar.

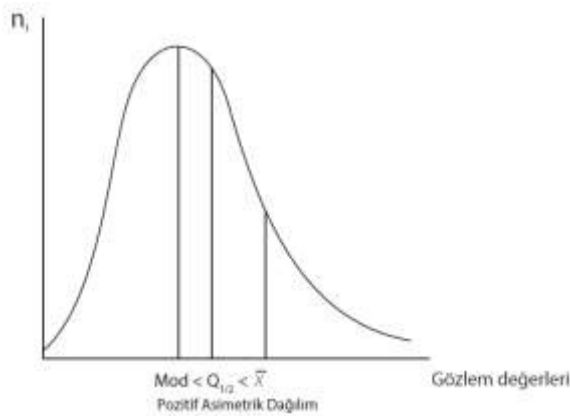
Pozitif Asimetrik Dağılım:

Seride büyük değerleri çoğunlukta olması halinde aritmetik ortalama bu değerlerin etkisiyle, büyük gözlemlerine doğru yönelir, yani yukarıya doğru kayar. Mod daima bir dağılımda en çok gözlenen değerdir ve en yüksek frekansı ifade eden tepe noktasında yerleşir. Medyan ise, küçükten büyüğe sıralanmış bir seride tam ortada yer alan değeri ifade ettiği için serinin ortasında yer alır.

Bu durumda, ortalamaların dizilişi şu şekilde gerçekleşir ve dağılımın şekli aşağıdaki gibi olur:

$$\text{Mod} < \text{Medyan} < \text{Aritmetik Ortalama}$$

$$\text{Mod} < Q_{1/2} < \bar{X}$$



Grafik 7.4: Asimetrisi pozitif dağılım

Yukarıdaki grafiğe dikkat edecek olursak, grafiğin tepe noktası modun yerleştiği yeri ifade etmektedir. Aritmetik ortalama, grafiğin sağ tarafa doğru uzaması sebebiyle, bu kısımda bulunan aşırı değerlerin etkisiyle moddan uzaklaşmış ve sağa doğru kaymıştır. Medyan da serinin tam ortasında yerleşmiştir. Medyanın, aritmetik ortalama ile mod arasında olması durumu söz konusudur ancak, iki ortalamanın tam ortasında yer aldığı düşünülmemelidir.

Yukarıda örneği görülen, aşırı değerlerin etkisiyle şeklin sağ kuyruğunun uzadığı durumlar pozitif asimetri durumu olarak adlandırılır. Başka bir deyişle, bu durumda asimetrisi pozitif bir dağılımdan söz edilmektedir.

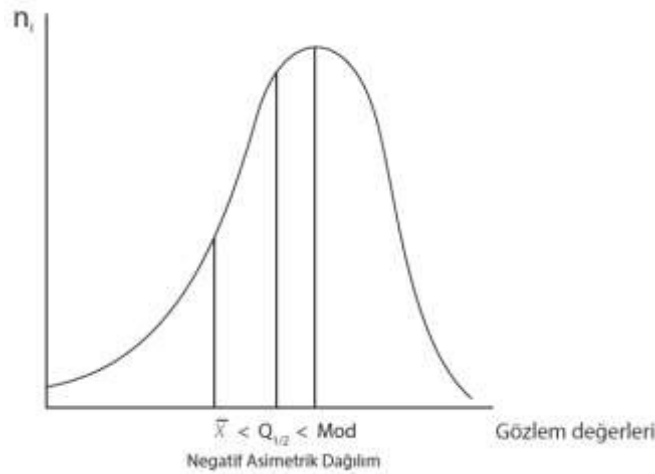
Negatif Asimetrik Dağılım:

Seride aşırı küçük değerlerin var olması halinde ise, aritmetik ortalama küçük değerlerin etkisiyle küçülür ve sola doğru kayarak serinin modundan uzaklaşır. Mod, yine en yüksek frekansa sahip gözlem değeridir ve dağılımın tepe noktasında konumlanır. Medyan ise, serinin tam ortasında yer alan değer olarak dağılımın ortasındaki yerini alır.

Böyle bir durumda ise ortalamaların dizilişi şu şekilde gerçekleşir ve dağılımın şekli aşağıdaki gibi olur:

$$\text{Aritmetik Ortalama} < \text{Medyan} < \text{Mod}$$

$$\bar{X} < Q_{1/2} < \text{Mod}$$



Grafik 7.5: Asimetrisi Negatif Dağılım

Yukarıdaki grafiği incelersek, grafiğin sol kuyruk kısmının sol tarafa doğru uzamış olduğunu görürüz. Bu durum seride küçük değerlerin fazla olduğunu ifade eder. Küçük değerlerin etkisiyle aritmetik ortalama küçülerek sola doğru kayar ve moddan uzaklaşır. Mod, maksimum frekansa sahip gözlem değeri olarak grafiğin tepe noktasında konumlanır. Medyan ise, tam ortada yer alan gözlem değerini ifade ettiğinden dağılımın ortasında yer alan değer olmaktadır.

Bu Bölümde Neler Öğrendik?

Bu bölümde analitik olmayan ortalamalar başlığı altında mod ve medyan hesabına yer vererek, bu iki ortalamanın kullanım alanlarından ve hangi durumlarda tercih edilmekte olduklarından söz ettik.

Ortalamalar arasındaki ilişkilerden ve aritmetik ortalama, mod ve medyanın aldığı değerlerin dizilişine bakarak bir serinin simetrik olup olmadığının nasıl belirlendiğine değindik.

Ortalama, bir veriyi özetlemek açısından son derece önemli bir ölçü olmakla beraber; tek başına yeterli olabilen bir ölçü değildir ve seriyle ilgili doğru ve detaylı bilgi sahibi olabilmek için serinin dağılma ölçüsünü de bilmek gerekir.

İzleyen bölümde de bu amaçla kullanılan ikinci ölçü grubunu oluşturan dağılma ölçüleri ele alınacaktır.

Bölüm Soruları

- 1- Aşağıdaki ortalamalardan hangisi analitik olmayan (duyarsız) bir ortalamaadır?
 - a- Aritmetik ortalama
 - b- Geometrik ortalama
 - c- Medyan
 - d- Kareli ortalama
 - e- Duyarlı ortalama
- 2- Seride yer alan gözlem değerleri arasında en çok tekrar eden gözlem değerine denir.
 - a- Mod
 - b- Aritmetik ortalama
 - c- Geometrik ortalama
 - d- Kareli ortalama
 - e- Duyarlı ortalama
- 3- Sıralanmış bir seride tam ortada yer alan değeredenir.
 - a- Aritmetik ortalama
 - b- Geometrik ortalama
 - c- Medyan
 - d- Kareli ortalama
 - e- Duyarlı ortalama
- 4- Bir sınıfta bulunan 10 öğrencinin ağırlıkları 65, 70, 70, 60, 55, 60, 70, 60, 65, 60 olarak ölçülmüştür. Buna göre ağırlıkların mod değeri nedir?
 - a- 60
 - b- 65
 - c- 70
 - d- 55
 - e- Modu yoktur.
- 5- Gözlem değerleri (2, 3, 5, 6, 7, 7, 8, 8, 8, 9, 9, 10) olan ve 12 gözleminden oluşan serinin medyanı nedir?

- a- 8
- b- 7
- c- 7,5
- d- 9
- e- 8,5

6- Aşağıda verilen serinin modu nedir?

X_i	n_i
10	5
20	7
25	9
30	7
40	5

- a- 25
- b- 20
- c- 9
- d- 7
- e- 30

7- Simetrik bir seri için aşağıdakilerden hangisi doğrudur?

- a- Aritmetik Ortalama > Medyan > Mod
- b- Aritmetik Ortalama < Medyan < Mod
- c- Aritmetik Ortalama = Medyan = Mod
- d- Aritmetik Ortalama = Mod > Medyan
- e- Aritmetik Ortalama > Mod = Medyan

8- Pozitif asimetri bir seri için aşağıdakilerden hangisi doğrudur?

- a- Ortalama = Medyan=Mod
- b- $\text{Mod} > \text{Medyan} > \text{Aritmetik Ortalama}$
- c- $\text{Mod} < \text{Medyan} < \text{Aritmetik Ortalama}$
- d- $\text{Ortalama} = \text{Mod} > \text{Medyan}$
- e- $\text{Ortalama} > \text{Mod} = \text{Medyan}$

9- Aşağıda verilen sınıflanmış seride mod sınıfı hangi sınıftır?

Sınıflar	n_i
0-10	4
10-20	9
20-30	8
30-40	7
40-50	6

- a- 0-10 sınıfı
- b- 10-20 sınıfı
- c- 20-30 sınıfı
- d- 30-40 sınıfı
- e- 40-50 sınıfı

10- Aşağıda verilen sınıflanmış seride medyan sınıfı hangi sınıftır?

Sınıflar	n_i
0-10	4
10-20	9
20-30	8
30-40	7
40-50	6

- a- 0-10 sınıfı
- b- 10-20 sınıfı
- c- 30-40 sınıfı
- d- 40-50 sınıfı
- e- 20-30 sınıfı

Cevaplar

- 1- c
- 2- a
- 3- c
- 4- a
- 5- c
- 6- a
- 7- c
- 8- c
- 9- b
- 10- e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

8. DAĞILMA ÖLÇÜLERİ

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde bir dağılım hakkında bilgi sahip olmak konusunda ortalamaların neden yeterli olmadığını ve bu amaçla ihtiyaç duyduğumuz dağılma ölçülerini öğreneceğiz.

Bu bağlamda önce dağılma ve değişkenlik kavramlarının neler ifade ettiğini üzerinde duracak, daha sonra da dağılma ve değişkenliğin belirlenmesinde kullanılabilecek ölçülerin neler olduğunu göreceğiz.

Bu bölümde, birimlerin birbirine ya da ortalamaya kıyasla nasıl dağıldığını belirlemede kullanılan değişim aralığı, ortalama sapma, standart sapma ve varyans ölçülerini ayrıntılı olarak örneklerle el alacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Çok sayıda gözlem değerinden oluşan bir veri setini özetlemede ortalama hesaplamak neden yeterli değildir?
- 2) Dağılma ve değişkenlik ne demektir?
- 3) Dağılma ölçüleri nelerdir?
- 4) Dağılma ölçüleri arasında en çok tercih edilen hangi ölçüdür?
- 5) Dağılmanın az ya da çok olması neyin göstergesidir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Dağılımları belirlemede ortalamalar tek başına yeterli midir?	Dağılımları belirlemede ortalamalar yanında neden dağılma ölçülerine de ihtiyaç duyulduğunu bilir.	Dağılım ve değişkenlik konusunda çok sayıda temel düzeyde istatistik kitabı bulunmaktadır ve her biri detaylı bilgi sahibi olmak için kullanılabilir.
Dağılma ve değişkenlik ölçüleri: kavramsal ve kuramsal çerçeve	Dağılma ve değişkenlik kavramlarının anlamını bilir.	Temel düzeyde istatistik kitaplarının dağılma ve değişkenlik bölümleri.
Mutlak dağılma ölçüleri nelerdir?	Bir dağılımın dağılma değişkenlik derecesini ortaya koymak üzere geliştirilmiş birbirinden ya da aritmetik ortalamadan uzaklığı dikkate alan ölçüleri bilir.	Dağılma ölçülerine yönelik olarak basit düzeyde yazılmış istatistik kitaplarında ilgili konulara ilişkin çok sayıda örnek çözüm bulunabilir.
Değişim Aralığı ve Ortalama Sapma	Mutlak dağılma ölçüsü olarak Değişim Aralığı ve Ortalama Sapmanın nasıl hesaplandığını ve özelliklerini bilir.	Dağılma ölçülerine yönelik olarak basit düzeyde yazılmış istatistik kitaplarında ilgili konulara ilişkin çok sayıda örnek çözüm bulunabilir.
Standart sapma ve varyans	Mutlak dağılma ölçüsü olarak standart sapma ve varyansın nasıl hesaplandığını ve özelliklerini bilir.	Dağılma ölçülerine yönelik olarak basit düzeyde yazılmış istatistik kitaplarında ilgili konulara ilişkin çok sayıda örnek çözüm bulunabilir.

Anahtar Kavramlar

- Dağılma ve değişkenlik kavramı: Bir dağılımda gözlem değerlerinin ne derece geniş bir alanda yayıldığı.
- Homojenlik: Bir dağılımda yer alan gözlem değerlerinin birbirinden çok uzaklaşmadığı durum.
- Ortalamanın temsil kabiliyeti: Hesaplanan ortalamanın ait olduğu seriyi temsil edebilme yeteneği.
- Mutlak Dağılma Ölçüleri: Değişim aralığı, ortalama sapma, standart sapma ve varyans.

Giriş

Daha önceki bölümlerde veri toplama aşamasından bahsettik. Bu kapsamda temel kavramlardan söz ettikten sonra veri toplama yöntemlerini, veri toplama aşamasında yapılabilecek hataları ve dikkat edilmesi gereken hususları inceledik. Daha sonra, ham veri olarak nitelenen toplanmış verilerin nasıl düzenlendiğini ve istatistik analizler yapılmaya nasıl uygun hale getirildiğini öğrendik.

Daha sonra düzenlediğimiz verileri tek bir değer ile özetleme ya da temsil etmenin öneminden bahsederek, bu amaçla kullanılmakta olan ortalamalardan değindik. Hangi veri türlerinde hangi ortalamanın kullanılabileceğini belirterek ortalamanın temsili olmasının önemi üzerinde durduk.

Bu bölümde de verileri özetlemede ortalamaların tek başlarına neden yeterli olmadığını açıklayarak, bu amaçla ihtiyaç duyulan ikinci bir ölçüye, dağılma ölçülerine değineceğiz.

Hatırlayacağınız gibi, ortalamalar, bir serinin gözlem değerlerini tek bir değer olarak özetleme, seriyi tek bir değer olarak temsil etme amacıyla hesaplanırlar. Ayrıca, ortalama hesaplanarak seriye ilişkin akılda kalıcı bilgi yaratma ve bu bilgi vasıtasıyla çeşitli serileri birbiri ile kıyaslayabilme imkânına sahip olmanın da amaçlandığını söyleyebiliriz.

Ortalamalar bir seriyi özetlemede çok önemli ölçüler olmakla beraber, serideki birimlerin ne derece dağıldığına yönelik bilgi sahibi olunmadığı takdirde, sadece ortalama ile söz konusu seri ile ilgili yeterli düzeyde bilgi sahibi olunduğu söylenemez.

Bu konuyu bir örnek üzerinden açıklamaya çalışalım.

İki öğrencinin belirli bir dersten, bir yıl boyunca girdikleri tüm sınavlardan aldıkları notlar aşağıdaki tabloda düzenlenmiş olsun.

Umut'un Notları X_i	Cemre'nin Notları Y_i
20	50
40	55
60	60
80	65
100	70
$\sum X_i = 300$	$\sum Y_i = 300$
$\bar{X} = \frac{\sum X_i}{n} = \frac{300}{5} = 60$	$\bar{Y} = \frac{\sum Y_i}{n} = \frac{300}{5} = 60$

Yukarıdaki örnekte, her iki öğrencinin de ders ortalamaları eşit olup her ikisinin de ortalaması 60'dır. Öğrencilerin notlarına baktığımızda, Umut'un notlarının daha geniş bir

alandaki (20-100 aralığında) dağıldığı, Cemre'nin notlarının ise birbirine daha yakın (50-70 aralığında) seyrettiği görülmektedir.

Hangi öğrencinin ortalaması, o öğrencinin yıl içindeki başarısını daha iyi temsil etmektedir sorusuna cevap aradığımızda, Cemre'nin notlarının ortalama not olan 60 değerine yakın olduğunu ve notların gerek birbirinden, gerekse ortalamadan çok uzaklaşmadığını görüyoruz. Bu açıdan bakıldığında Cemre'nin not ortalaması Cemre'nin başarı durumunu özetlemede ya da temsil etmede Umut'a kıyasla daha başarılıdır.

Umut'un notlarının ise gerek ortalama olan 60 değeri etrafında gerekse notların birbirlerine göre daha fazla dağıldığını ve bu sebeple Umut'un ortalamasının Cemre kadar temsili olmadığını söyleyebiliyoruz.

Görüldüğü gibi, her iki öğrencinin ortalama notunun 60 olmasına bakarak, sadece ortalamaları üzerinden öğrencilerin başarı düzeylerini belirlemek ve yorumlamak yetersiz bir yaklaşım olmaktadır.

Ortalamaların yanında, birimlerin ne derece dağıldığını bilmek de sağlıklı değerlendirme yapabilmek açısından son derece önemlidir.

Dolayısıyla, bir seriyi özetlemek için ortalamalar tek başına yeterli ölçüler olmayıp dağılma ölçülerinin de değerlendirmeye alınması gerekmektedir.

Şimdi dağılma ölçülerinin neler olduğunu ve nasıl kullanıldıklarını açıklayalım.

8.1. Dağılma Ölçüleri

Dağılma ölçüleri, bir seride yer alan birimlerin birbirlerinden ya da ortalamadan ne derece uzaklaştığını gösteren ve bir dağılımın yapısıyla ilgili bilgi veren olmazsa olmaz ikinci ölçü grubudur.

Giriş kısmında da ifade ettiğimiz gibi, bir dağılımın birinci karakteristiği birimlerin hangi noktada toplandığını gösteren ortalamadır. Ancak bir dağılımın ortalaması o dağılım hakkında bilgi veren önemli bir ölçü olmakla beraber birimlerin ne derece dağıldığını / yayıldığını bilmek de önemlidir. Nitekim, ortalaması aynı olan iki dağılımda birimlerin dağılımının birbirinden çok farklı olabileceğini yukarıdaki not örneğinde de gördük. İşte bu sebeple, bir dağılımı belirlemede ya da anlamada kullanılacak ikinci karakteristik ölçü dağılma ölçüsü olmaktadır.

Bir dağılımın dağılma ölçüsü, o dağılımın homojen olup olmadığını anlamada önemli bir araçtır. Dağılımın az olması birimlerin birbirine ya da ortalamaya daha yakın olacak şekilde dağıldığını ifade ederken, dağılımın fazla olması birimlerin birbirlerinden ya da ortalamadan uzaklaştığını ifade eder. Dolayısıyla, dağılımı daha az olan seri daha homojendir.

Bir dağılımın homojen olup olmaması da hesapladığımız ortalamanın, o dağılımını temsil etmede ne derece başarılı olduğunun göstergesidir. Şöyle ki, bir dağılım homojense; yani gözlem değerleri birbirine yakın seyrediyorsa, hesaplanan ortalama o dağılımı temsil etmek konusunda başarılıdır ve temsili bir ortalama hesaplanmıştır.

Oysa, bir dağılım homojen değilse, başka bir deyişle birimler birbirinden ya da ortalamadan uzaklaşıyorsa bu durumda hesaplanan ortalamanın seriyi temsil kabiliyeti zayıftır ve temsili bir ortalamadan söz etmek mümkün olamamaktadır.

Özetle, bir dağılımın dağılma derecesi arttıkça homojenliği azalır ve ortalama temsili olma özelliğini yitirir. Benzer şekilde, bir dağılımın dağılma derecesi azaldıkça homojenliği artar ve ortalamanın temsil kabiliyeti yükselir.

Dağılma ölçüleri birimlerin birbirine göre uzaklıklarını dikkate alabileceği gibi, birimlerin ortalamadan uzaklıklarını da kullanabilir. İşte, birimlerin birbirinden ya da ortalamadan uzaklığını temel alan ölçülere mutlak dağılma ölçüleri denir.

Şimdi uygulamada kullanılan mutlak dağılma ölçülerini en basitinden en yaygın kullanılabilecek doğru sıralamasıyla inceleyelim.

8.2. Değişim Aralığı

Değişim aralığı bir serinin ne derece dağıldığını belirlemede kullanılabilecek en basit ve anlaşılır ölçüdür.

Değişim aralığı, serideki en büyük değer ile en küçük değer arasındaki farkı ifade eder ve aşağıdaki şekilde formüle edilebilir:

$$\text{Değişim Aralığı} = X_{maks} - X_{min}$$

Formülde yer alan,

X_{maks} serideki en büyük değeri

X_{min} ise serideki en küçük değeri

ifade etmektedir.

Formülden de anlaşılacağı gibi, değişim aralığı bir serideki en küçük ve en büyük değeri dikkate almakta, seri bünyesinde yer alan diğer birimleri hesaba katmamaktadır. Bu açıdan bakıldığında değişim aralığı analitik olmayan yani duyarsız bir dağılma ölçüsüdür.

Değişim aralığının hesabına yönelik bir örnek üzerinden hem nasıl hesaplandığını görelim hem de zayıf yanlarının neler olduğunu tartışalım.

Örnek:

Bir öğrenci, üniversitelere giriş sınavı hazırlık aşamasında her hafta bir matematik deneme testine girmektedir. Öğrencinin son iki haftada girdiği 40 soruluk testlerdeki doğru cevap sayısı aşağıdaki tabloda düzenlenmiştir.

Öğrencinin başarısının hangi haftada daha homojen olduğu belirlenmek isteniyor. Değişim aralığı yaklaşımı ile öğrencinin başarısının homojenliğini değerlendirelim.

1. Hafta	2. Hafta
25	28
28	30
30	33
33	34
34	34
Değişim Aralığı =34-25=9	Değişim Aralığı =34-28=6

Değişim Aralığı, serideki en büyük değer ile en küçük değer arasındaki farka dayandığına göre,

$$\text{Değişim Aralığı} = X_{maks} - X_{min}$$

formülüyle her iki haftanın değişim aralığı değerlerini kolaylıkla hesaplayabiliriz.

1.Hafta için değişim aralığı,

$$\text{Değişim Aralığı}_1 = 34-25=9$$

2.Hafta için değişim aralığı,

$$\text{Değişim Aralığı}_2 = 34-28=6$$

şeklinde hesaplanmıştır.

Buna göre, $6 < 9$ olması sebebiyle öğrencinin 2.haftada gösterdiği matematik başarısının daha homojen olduğunu söylemek mümkündür. Başka bir deyişle öğrenci 2.haftada test sonuçları anlamında daha istikrarlı bir başarıya sahiptir.

8.2.1. Tasnif Edilmiş ve Sınıflanmış Frekans Serilerinde Değişim Aralığı

Tasnif edilmiş ya da sınıflanmış serilerde de değişim aralığı hesabı basit seride olduğu gibi yapılır. Seride yer alan ilk ve son gözlem değeri dışındaki gözlem değerlerinin değişim aralığı hesabında kullanılmaması sebebiyle, aradaki gözlemler ve frekanslar işleme alınmaz. Dolayısıyla, en büyük gözlem değeri ile en küçük gözlem değeri arasındaki fark değişim aralığını verecektir.

Sınıflanmış frekans dağılımı açısından da yine, en yüksek ve en düşük sınıf orta noktaları arasındaki fark alınmak suretiyle değişim aralığı hesaplanabilir.

Ancak değişim aralığının analitik olmayan bir ölçü olması yanında cebirsel işlemlere elverişsiz bulunmasının söz konusu ölçünün kullanım alanını büyük ölçüde sınırlamakta olduğunu görüyoruz. Öte yandan, tasnif edilmiş ya da sınıflanmış frekans dağılımları gibi gözlem sayısının göreceli olarak daha fazla olduğu uygulamalarda da; dağılımın ilk ve son değerini dikkate almak suretiyle hesaplanan değişim aralığının bir dağılma ölçüsü olarak tercih edilmediğini görmekteyiz.

8.2.2. Değişim Aralığının Özellikleri

Değişim aralığı, serideki en büyük değer ile en küçük değer arasındaki farka dayanan bir ölçüdür ve iki değer arasındaki fark ne derece büyükse, dağılımın o derece fazla olduğu düşünülür.

Değişim aralığının hesabında sadece seride yer alan en büyük ve en küçük değeri dikkate alması bu ölçünün analitik olmayan ölçüler yani duyarsız ölçüler sınıfında değerlendirilmesine yol açar. Hesabında seride yer alan tüm değerlerin kullanılmaması sebebiyle cebirsel işlemlere elverişli değildir ve kullanım alanı iki serinin dağılımı hakkında kabaca bir fikir sahibi olmakla sınırlıdır.

Değişim aralığının en zayıf yönü şüphesiz bir dağılımdaki en büyük ve en küçük değerlerin farkına dayanması ve bu dağılımda yer alan diğer gözlem değerlerini dikkate almamasıdır. Oysa, dağılımda yer alan tüm birimlerin dağılma ölçüsü hesabına dâhil edilmesi dağılma ölçüsünü daha güçlü bir ölçü haline getirir. Başka bir deyişle, sadece ilk ve son değer değil, diğer tüm gözlem değerlerinin de seri içinde dağılımını dikkate almak hesaplanan dağılma ölçüsünü daha anlamlı kılar. Bu sebeplerle, değişim aralığı anlaşılması basit ve kullanımı kolay bir ölçü olmakla beraber bir serinin dağılma derecesini belirlemek amacıyla tercih edilmekte olan ölçülerden olamamıştır.

Değişim Aralığının sözüne ettiğimiz zayıflıkları, bir serinin dağılma derecesini belirlemek üzere daha güçlü ölçüler geliştirmeyi gerektirmiş ve ortalama sapma ve standart sapma istatistik çalışmalarda yerini almıştır. Şimdi sırasıyla bu dağılma ölçülerini inceleyelim.

8.3. Ortalama Sapma

Ortalama sapma, bir serideki bütün değerleri dağılma ölçüsü hesabına dâhil eden analitik ölçülerden biridir.

Ortalama sapma, seride yer alan birimlerin, aritmetik ortalamadan farklarını hesaplayarak, tüm birimlerin ortalamadan, ortalama olarak ne kadar saptığını belirlemeyi amaçlar. Kısaca ortalama sapma, birimlerin aritmetik ortalamadan ortalama sapma miktarıdır.

Ortalama sapma formülüne geçmeden önce, bir örnek seri üzerinde ortalamadan sapma miktarlarını hesaplamak istiyoruz. Aritmetik ortalamadan sapma miktarlarını değerlendirdikten sonra ortalama sapma formülünü aşama aşama oluşturacağız.

Aşağıda diyabet hastası olduğu şüphesiyle hekime başvuran 35 yaşında bir erkeğin, 4 günlük süre içinde yapılan açlık kan şekeri ölçümleri veriliyor:

Kan Şekeri Ölçüm Değerleri X_i	Ortalamadan Sapmalar $X_i - \bar{X}$
120	120-130= -10
127	127-130= -3
133	133-130= +3
140	140-130=+10
$\sum X_i = 520$ ve $\bar{X} = \frac{\sum X_i}{n} = \frac{520}{4} = 130$	$\sum (X_i - \bar{X}) = 0$

Ortalama sapmanın, birimlerin aritmetik ortalamadan sapmalarına dayalı bir ölçü olduğunu söylemiştik. O halde öncelikle yukarıdaki tabloda görülen kan şekeri ölçüm düzeylerinin aritmetik ortalamasını hesaplayalım:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{520}{4} = 130$$

Açlık kan şekeri ortalamasını 130 birim olarak hesapladık. Şimdi de, her bir gözlem değerinin aritmetik ortalamadan uzaklıklarını yani aritmetik ortalamadan farklarını hesaplayalım. Yukarıdaki tablonun ikinci sütununda, gözlem değerlerinin aritmetik ortalamadan sapmaları verilmiştir.

Aritmetik ortalamadan sapmaların ortalamasını hesaplamak ve her bir birimin ortalamadan ortalama olarak ne kadar saptığını belirlemek istediğimize göre, hesapladığımız sapmalar toplamını birim sayısına bölmemiz gerekecektir. Ancak, tablodan da görüleceği gibi; sapma serisi toplamı sıfırdır.

Aritmetik ortalamadan sapmaların toplamının sıfır olması sadece bu örneğe özgü bir durum değildir. Aritmetik ortalamadan sapmalar toplamı, istisnasız tüm seriler için daima sıfır çıkar. Dolayısıyla, sapma serisinin ortalamasını hesaplamak istediğimizde toplamı sıfır olan bir serinin ortalamasını hesaplamak durumundayız.

İşte bu noktada sapmalar serisi toplamını sıfır olmaktan kurtaracak iki yaklaşım benimsenebilir. Birincisi, + ve – işaretli olarak gerçekleşen sapma miktarlarının mutlak değerlerini almak, diğeri ise sapmaların karelerini almak.

İkinci yaklaşımı bir sonraki başlıkta incelemek üzere, birinci yaklaşımın nasıl uygulandığını açıklayalım.

Toplamı sıfır olan aritmetik ortalamadan sapma değerlerinin mutlak değerlerini alarak toplamı sıfır olmaktan kurtarabiliriz dedik. Bilindiği gibi, mutlak değer birimler arasında negatif ya da pozitif ayrımı gözetilmeksizin tüm birimlerin pozitif olarak değerlendirilmesidir.

Aritmetik ortalamadan sapma değerlerinde, mutlak değer alma işlemi,

$$|X_i - \bar{X}|$$

formülüyle yapılır ve hesaplanan mutlak sapma değerleri toplanıp, gözlem sayısına bölünerek ortalama sapma hesaplanmış olur.

Dolayısıyla ortalama sapma formülü;

$$\text{Ortalama Sapma} = \frac{\sum |X_i - \bar{X}|}{n}$$

şeklinde kolaylıkla oluşturulabilir.

Ortalama hesabı, aritmetik ortalamadan sapmaların mutlak değerleri alınarak hesaplandığından bu ölçüye ortalama mutlak sapma ya da mutlak ortalama sapma da denilmektedir.

Şimdi üç seri türünde ortalama hesabının nasıl yapıldığını öğrenelim:

8.3.1. Basit Seride Ortalama Sapmanın Hesabı

Basit seride ortalama sapma hesabını yukarıdaki kan şekeri örneği üzerinden açıklayalım.

Kan Şekeri Ölçüm Değerleri X_i	Ortalamadan Sapmalar $X_i - \bar{X}$	Mutlak Sapmalar $ X_i - \bar{X} $
120	120-130= -10	10
127	127-130= -3	3
133	133-130= +3	3
140	140-130=+10	10
$\sum X_i = 520$	$\sum (X_i - \bar{X}) = 0$	$\sum X_i - \bar{X} = 26$

Açlık kan şekeri ortalamasını 130 olarak hesaplamıştık. Aritmetik ortalamadan sapmaları hesaplamak üzere her bir gözlem değerinden 130'u çıkararak sapma değerlerini bulduk. Toplamı sıfır olan sapma serisinin toplamını sıfır olmaktan kurtarmak üzere hesapladığımız sapma değerlerinin mutlak değerlerini aldık.

Yukarıdaki tabloda ikinci sütunda ortalamadan sapmaları, üçüncü sütunda da ortalamadan sapmaların mutlak değerlerini görmekteyiz.

Şimdi de mutlak sapma değerlerinin ortalamasını hesaplayalım, Bunun için mutlak sapma değerlerini toplayarak toplam gözlem sayısına bölmek yeterlidir. Yani ortalama sapma formülünü uygulayarak,

$$\text{Ortalama Sapma} = \frac{\sum |X_i - \bar{X}|}{n} = \frac{26}{4} = 6,5$$

şeklinde ortalama sapmaya ya da diğer söylenişle mutlak ortalama sapmaya ulaşmış oluyoruz. Bu değer, gözlem değerlerinin kan şekeri ölçüm değerleri ortalaması olan 130 değerinden ortalama olarak 6,5 birim saptığını göstermektedir. Başka bir deyişle, her bir kan şekeri ölçüm değeri ortalamadan 6,5 birim sapmaktadır. Şüphesiz bu sapma değeri ortalama bir değerdir ve gözlem değerleri bu sapma miktarı kadar + ya da – yönde değişebilmektedir.

8.3.2. Tasnif Edilmiş Seride Ortalama Sapmanın Hesabı

Tasnif edilmiş serilerde, gözlem değerlerinin her bir gözlem değerinin kaç kez tekrar ettiğini gösteren frekanslarıyla birlikte düzenlendiğini hatırlayalım.

Nasıl ki her bir gözlem değeri frekansı kadar tekrar ediyorsa, her bir sapma değeri de yine frekansı kadar tekrar edecektir.

Frekansları da ortalama sapma formülüne dahil ettiğimizde, aşağıdaki formüle ulaşırız:

$$\text{Ortalama Sapma} = \frac{\sum n_i |X_i - \bar{X}|}{\sum n_i}$$

Yukarıdaki ortalama sapma formülünü incelediğimizde, birimlerin aritmetik ortalamadan sapmalarını hesaplayarak sapma değerlerinin mutlak değerlerini almak ve bu değerleri tekrar sayıları olan frekanslarla çarpıp toplamamız ve çıkan sonucu toplam gözlem sayısına bölmemiz gerektiğini anlıyoruz.

Tasnif edilmiş serilerde ortalama sapmanın nasıl hesaplandığını bir örnek uygulama ile görelim.

Örnek:

Bir klinikte, operasyon geçiren hastaların taburcu edilme süreçleri inceleniyor. Klinikte tedavi gören hastaların operasyon sonrası kaçınıcı günde taburcu edildikleri aşağıdaki tablo ile

düzenleniyor. Buna göre, hastaların ortalama taburcu olma süresini ve taburcu olma sürelerinin ortalama sapmasını hesaplayalım.

Taburcu olma süresi (X_i)	Frekans (n_i)	$n_i X_i$	$X_i - \bar{X}$	$ X_i - \bar{X} $	$n_i X_i - \bar{X} $
1	18	18	-1	1	18
2	18	36	0	0	0
3	10	30	+1	1	10
4	4	16	+2	2	8
	50	$\sum n_i X_i = 100$			$\sum n_i X_i - \bar{X} = 36$

Yukarıdaki tabloda ayrıntıları görebileceğiniz biçimde, öncelikle taburcu olma süresi için aritmetik ortalamayı hesaplayalım.

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{100}{50} = 2 \text{ gün}$$

Hastaların operasyon sonrası ortalama taburcu olma süresi 2 gündür.

Şimdi de taburcu olma sürelerinin ortalama sapmasını hesaplayalım. Bunun için tasnif edilmiş serilerde kullanılan ortalama sapma formülüyle,

$$\text{Ortalama Sapma} = \frac{\sum n_i |X_i - \bar{X}|}{\sum n_i} = \frac{36}{50} = 0,72 \text{ gün}$$

olarak hesaplanmaktadır.

Dolayısıyla, söz konusu operasyon sonrası hastalar ortalama olarak 2 günde taburcu edilmekte ve taburcu olma süresi 0,72 günlük bir sapma gösterebilmektedir. Şüphesiz bu sapma değeri ortalama olarak hem + hem – yönde ortaya çıkabilir. Yani hastalar, ortalama taburcu süresinden ortalama olarak sapma süresi kadar daha fazla ya da daha az bir sürede de taburcu olabilir.

8.3.3. Sınıflanmış Seride Ortalama Sapmanın Hesabı

Sınıflanmış serilerin, gözlem değerlerinin yer aldığı sınıflar ve her bir sınıfta kaç gözlem değeri bulunduğunu gösteren frekanslardan oluşan seriler olduğunu hatırlayalım.

Sınıflanmış serilerde, gözlem değerleri yerine alt ve üst sınırları belli olan sınıflar bulunduğundan, öncelikle her bir sınıfı temsil etmek üzere sınıf orta noktalarını belirlemek zorunda olduğumuzu da anımsayalım.

Ortalama sapma hesabında da, daha önceki analitik çözümlerde olduğu gibi, sınıf orta noktalarının (yani m_i 'lerin) bir gözlem değeri gibi düşünüldüğünü ve formüllerin bu şekilde oluşturulduğunu görüyoruz.

Sınıflanmış seride kullanılacak ortalama sapma formülü,

$$\text{Ortalama Sapma} = \frac{\sum n_i |m_i - \bar{X}|}{\sum n_i}$$

biçiminde olmaktadır.

Sınıflanmış bir frekans dağılımında ortalama sapmanın nasıl hesaplandığını da bir örnek yardımıyla inceleyelim.

Örnek:

Bir tüp bebek merkezinde, tedavi yoluyla bebek sahibi olan 80 kadının bebek sahibi oldukları yaşlar inceleniyor ve aşağıdaki tablo ile toplanan veriler düzenleniyor.

Bu kliniğe başvuran kadınların tüp bebek yöntemiyle bebek sahibi oldukları ortalama yaşı ve ortalama sapmasını hesaplayalım.

Yaş sınıfları	Frekanslar (n_i)	Sınıf orta noktaları (m_i)	$n_i m_i$	$m_i - \bar{X}$	$ m_i - \bar{X} $	$n_i m_i - \bar{X} $
34-36	29	35	1015	-2,225	2,225	64,525
36-38	25	37	925	-0,225	0,225	5,625
38-40	14	39	546	+1,775	1,775	24,85
40-42	12	41	492	+3,775	3,775	45,3
	80		$\sum n_i m_i = 2978$			$\sum n_i m_i - \bar{X} = 140,3$

Öncelikle kliniğe başvuran kadınların bebek sahibi oldukları ortalama yaşı hesaplayalım.

$$\bar{X} = \frac{\sum n_i m_i}{\sum n_i} = \frac{2978}{80} = 37,225 \text{ yaş}$$

Kadınların tedavileri sırasında bebek sahibi olabildikleri ortalama yaş 37,225'dir.

Şimdi de ortalama sapmayı hesaplayabilmek için, sınıf orta noktalarının aritmetik ortalamadan farklarını hesaplayalım ve sonra bu farkların mutlak değerlerini alalım (Bu işlemi

5.ve 6. sütundan takip edebilirsiniz). Son olarak da her bir mutlak farkı kendi frekansı ile çarpalım.

Sınıflanmış frekans dağılımları için ortalama sapma formülüne göre,

$$\text{Ortalama Sapma} = \frac{\sum n_i |m_i - \bar{X}|}{\sum n_i} = \frac{140,3}{80} \cong 1,75 \text{ yaş}$$

olmaktadır.

Sonuç olarak, bu kliniğe başvuran kadınların bebek sahibi olma yaşları ortalama olarak 37,225'dir ve bebek sahibi olma yaşları ortalamadan yaklaşık 1,75 yaş sapabilmektedir. Bu sapma miktarının, bebek sahibi olunan yaş üzerinde + ve – yönde etki edebileceğini de ilave edelim.

8.3.4. Ortalama Sapmanın Özellikleri

Ortalama sapma seride yer alan bütün gözlem değerlerinin ortalama hesabında kullanılması sebebiyle analitik ya da duyarlı dağılma ölçüleri grubunda yer almaktadır.

Ortalama sapma değişim aralığına kıyasla, dağılma derecesini belirlemede daha güçlü bir ölçüdür. Öte yandan hesabının kolay ve anlaşılır olması da ortalama sapmanın diğer olumlu taraflarını oluşturmaktadır.

Ancak, ortalama sapma hesabının mutlak değer almayı gerektirmesi ise ortalama sapmanın cebirsel işlemlere elverişli olmaması sonucunu doğurmaktadır ki bu ortalama sapmanın kullanılışını ciddi derecede kısıtlamaktadır.

Ortalama sapma, bir dağılımın ortalama etrafında ortalama dağılma derecesini göstermesi dışında daha ileri analizlerde kullanılabilir olma özelliği taşımaz. Dolayısıyla amaç, bir dağılımı tanımlamak üzere bir ortalama ve dağılma ölçüsü hesaplamaksa ortalama sapma tercih edilebilir. Ancak, söz konusu veriye daha ileri istatistik analizler uygulanacaksa serinin dağılımını ya da değişkenliğini göstermek üzere ortalama sapmanın kullanılması mümkün olmamaktadır. Böyle durumlarda, izleyen başlıkta ele alacağımız standart sapma tercih edilmektedir.

8.4. Standart Sapma ve Varyans

Bir dağılımın dağılma ya da değişkenlik derecesini gösteren ölçülerden biri de standart sapmadır. Standart sapma da, dağılma ölçüsü hesabında bütün gözlem değerlerini kullandığından; tıpkı ortalama sapma gibi analitik ya da duyarlı olarak nitelenen dağılma ölçüleri arasında yer alır.

Benzer şekilde standart sapma da, ortalama sapma gibi birimlerin aritmetik ortalamadan sapmalarının ortalamasını hesaplamayı amaçlar. Ancak, bir önceki başlıkta değindiğimiz gibi, aritmetik ortalamadan sapmaların cebirsel toplamı sıfırdır. Dolayısıyla, toplamı sıfır olan bir serinin ortalamasını hesaplayabilmek için toplamı sıfır olmaktan kurtaracak bir dönüşüm

yapmak gerekir. İşte bu noktada, ortalama sapma, sapma değerlerinin mutlak değerlerini alarak toplamı sıfır olmaktan kurtarırken; standart sapma bunu sapma değerlerinin karelerini alarak yapar.

Dolayısıyla, standart sapma, gözlemlerin aritmetik ortalamadan sapmalarının karelerinin ortalamasıdır ki bu ortalama, kareli ortalama olarak bilinir.

Kareli Ortalama formülünün,

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

şeklinde olduğunu hatırlayalım.

Standart sapma, $(X_i - \bar{X})$ şeklinde ifade edilen aritmetik ortalamadan sapmaların ortalamasını hesaplamayı amaçladığına göre, yukarıdaki kareli ortalama formülünde X_i gördüğümüz yere $(X_i - \bar{X})$ yazarsak,

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

standart sapma formülüne ulaşmış oluruz. Standart sapma, aritmetik ortalamadan sapmaların kareleri alınarak hesaplandığı için, yani özünde bir kareli ortalama olduğu için kareli ortalama sapma olarak da bilinir.

Ortalamaları ele aldığımız bölüm içinde kareli ortalamayı anlatırken özellikle toplamı sıfır olan serilerde tercih edilen bir ortalama olduğunu söylemiştik. Nitekim, aritmetik ortalamadan sapmalar serisinin de toplamı sıfırdır ve bu noktada sapma serisinin ortalamasını hesaplamada kareli ortalama önemli bir ölçü olarak karşımıza çıkmaktadır.

Şimdi basit, tasnif edilmiş ve sınıflanmış dağılımlarda standart sapmanın nasıl hesaplandığını örneklerle görelim.

8.4.1. Basit Seride Standart Sapma ve Varyansın Hesabı

Basit seride standart sapma hesabı, gözlem değerlerinin aritmetik ortalamadan sapmalarını hesaplayıp, bu sapma değerlerinin karelerini almak ve toplamak ve gözlem sayısına bölerek karekökünü hesaplayarak yapılır. Dolayısıyla, basit seride kullanılacak standart sapma formülü;

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

biçimindedir.

Örnek:

Bir üniversite öğrencisi, bir ay içinde 5 değişik günde sinemaya gitmiş ve farklı sinema salonlarında farklı ücretler ödemiştir. Öğrencinin ödediği sinema bilet ücretleri aşağıda gösterilmiştir. Buna göre, öğrencinin sinema bileti için ödediği ortalama ücret ve standart sapması nedir?

Sinema Bilet Ücreti X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
18	18-24=-6	36
20	20-24=-4	16
25	25-24=+1	1
27	27-24=+3	9
30	30-24=+6	36
$\sum X_i = 120$	$\sum (X_i - \bar{X}) = 0$	$\sum (X_i - \bar{X})^2 = 98$

Öncelikle öğrencinin sinema bileti için ödediği ortalama bilet ücretini hesaplayalım:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{120}{5} = 24 \text{ TL}$$

Dolayısıyla, öğrenci bir sinema biletine ortalama olarak 24 TL ödemiştir.

Şimdi de bilet ücretinin standart sapmasını hesaplayalım:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{98}{5}} = \sqrt{19,6} \cong 4,4 \text{ TL}$$

Sinema bilet ücretlerinin standart sapması 2,21 TL olarak hesaplanmıştır. Bu değer, bilet fiyatlarında ortalama bilet ücreti olan 24 TL'den ortalama olarak 4,4 TL'lik bir sapma söz konusu olduğunu ifade eder.

Burada, dağılma ölçüsü olarak kullanılan diğer bir ölçüden, varyanstan söz etmek istiyoruz. Varyans, başlı başına bir dağılım ölçüsü olmayıp, standart sapmanın karesidir. Dolayısıyla, yorumlanması da standart sapma gibidir.

Yukarıdaki örnekte $\sigma = 4,4$ hesaplanmıştı. Bu durumda varyans da standart sapmanın karesi olarak $\sigma^2 = 19,6$ olmaktadır.

Varyansın formülünü de yazmak istersek,

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

şeklinde olacaktır.

Görüldüğü gibi, standart sapma bir kareli ortalamadır; varyans ise onun karesidir.

8.4.2. Tasnif Edilmiş Serilerde Standart Sapma ve Varyansın Hesabı

Tasnif edilmiş seriler, gözlem değerleri ile her gözlem değerinin kaç kez tekrar ettiğini gösteren frekanslardan oluşur. Bu sebeple, standart sapma hesaplanırken her bir ortalamadan sapma değerinin hesaplanıp karesi alındıktan sonra ilgili frekans değeri ile çarpılması gerekir. Daha sonra, sapma değerlerinin kareleri toplanarak, toplam gözlem sayısına bölünür ve karekök alınarak standart sapmaya ulaşılmış olur.

Bu anlattıklarımızdan yola çıkarak tasnif edilmiş serilerde kullanılan standart sapma formülünü oluşturalım:

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

Aritmetik ortalamadan sapmaların karelerinin frekanslarla çarpıldığına lütfen dikkat edelim. Tasnif edilmiş serilerde kullanılacak varyans formülünü de yazmak istersek,

$$\sigma^2 = \frac{\sum n_i(X_i - \bar{X})^2}{\sum n_i}$$

formülüne ulaşırız.

Şimdi de tasnif edilmiş seride standart sapma ve varyans hesabının nasıl yapıldığına yönelik bir örnek uygulama yapalım.

Örnek:

Bir köyde yaşayan çocuk sahibi 44 ailenin sahip olduğu ortalama çocuk sayısı hesaplanmak isteniyor. Ailelerin sahip olduğu çocuk sayısı aşağıdaki tablo ile düzenlenmiştir. Buna göre, sahip olunan ortalama çocuk sayısını ve standart sapmayı hesaplayalım.

Çocuk Sayısı X_i	Frekans n_i	$n_i X_i$	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$n_i(X_i - \bar{X})^2$
1	15	15	-1	1	15
2	18	36	0	0	0
3	7	21	+1	1	7
4	4	16	+2	4	16
	$\sum n_i = 44$	$\sum n_i X_i = 88$			$\sum n_i(X_i - \bar{X})^2 = 38$

Öncelikle ailelerin sahip olduğu ortalama çocuk sayısını hesaplayalım. Tasnif edilmiş seriler için aritmetik ortalama formülünü kullanarak, ortalama çocuk sayısını,

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{88}{44} = 2 \text{ çocuk}$$

şeklinde hesaplıyoruz.

Standart sapmayı hesaplayabilmek için, öncelikle ailelerin çocuk sayısı ile ortalama çocuk sayısı arasındaki farkları yani, ortalamadan sapmalarını hesaplamamız gerekiyor. Daha sonra, hesapladığımız sapma değerlerinin karelerini alarak, her bir sapma değerini frekansı ile çarpıyoruz. Bu işlemler veri tablosunda 4, 5 ve 6. sütunlardan takip edilebilir.

Tasnif edilmiş seriler için standart sapma formülü kullanılarak, ailelerin çocuk sayısına ilişkin standart sapma,

$$\sigma = \sqrt{\frac{\sum n_i (X_i - \bar{X})^2}{\sum n_i}} = \sqrt{\frac{38}{44}} \cong \sqrt{0,8636} \cong 0,93 \text{ çocuk}$$

Olarak hesaplanmaktadır.

Dağılımın varyansı da;

$$\sigma^2 = \frac{\sum n_i (X_i - \bar{X})^2}{\sum n_i} = \frac{38}{44} \cong 0,8636 \text{ çocuk}$$

olarak hesaplanmaktadır.

Dolayısıyla, bu köyde yaşayan 44 aile için ortalama çocuk sayısı 2 çocuktur ve standart sapma miktarı da 0,93 çocuk şeklindedir. Yani 0,93 değerini yaklaşık olarak 1 kabul edersek, aileler ortalama olarak 2 çocuğa sahiptir ve çocuk sayısı ortalamadan, ortalama olarak + ve – yönde 1 çocuk kadar sapmaktadır.

8.4.3. Sınıflanmış Serilerde Standart Sapma ve Varyansın Hesabı

Sınıflanmış serilerde gözlem değerlerinin neler olduğunu görmenin mümkün olmadığını bunun yerine gözlem değerlerinin yer aldığı sınıfların ve bu sınıflara düşen kaç tane gözlem olduğunu gösteren frekansların yer aldığını hatırlayalım.

Sınıflanmış bir dağılım söz konusu olduğunda, ilk olarak sınıfları temsil etmek üzere sınıf orta noktalarının belirlenmesi gerekir. Daha sonra sınıf orta noktaları birer gözlem değeri gibi düşünülerek tasnif edilmiş serilerde uygulanan çözüm aşamaları takip edilir.

Tasnif edilmiş seriler için oluşturarak kullandığımız standart sapma ve varyans formüllerini, gözlem değerlerini gösteren X_i değerlerinin yerine sınıf orta noktalarını sembolize eden m_i değerlerini yazarak, sınıflanmış dağılımlar için yeniden düzenleyelim.

Böylelikle, sınıflanmış dağılımlarda kullanılacak standart sapma formülü,

$$\sigma = \sqrt{\frac{\sum n_i(m_i - \bar{X})^2}{\sum n_i}}$$

ve varyans formülü de;

$$\sigma^2 = \frac{\sum n_i(m_i - \bar{X})^2}{\sum n_i}$$

şeklinde olacaktır.

Şimdi de sınıflanmış dağılım örneği için standart sapma ve varyansın nasıl hesaplandığına yönelik bir örnek çözüm görelim.

Örnek:

Bir tüp bebek merkezine başvurarak bebek sahibi olan kadınların, başvurularından itibaren ortalama olarak kaç hafta sonra gebe kaldıkları araştırılıyor. Tedavi sonrasında gebe kalan kadınların başvuru tarihleri dikkate alınarak, gebe kaldıkları tarihe kadar geçen süre hafta cinsinden tespit ediliyor ve elde edilen veriler aşağıdaki tablo ile düzenleniyor.

Sınıflar (hafta olarak)	Frekanslar n_i	m_i	$n_i m_i$	$(m_i - \bar{X})$	$(m_i - \bar{X})^2$	$n_i(m_i - \bar{X})^2$
0-4	5	2	10	-9	81	405
4-8	8	6	48	-5	25	200
8-12	8	10	80	-1	1	8
12-16	10	14	140	+3	9	90
16-20	9	18	162	+7	49	441
	$\sum n_i = 40$		$\sum n_i m_i = 440$			$\sum n_i(m_i - \bar{X})^2 = 1144$

Öncelikle, kliniğe başvuran kadınların gebe kalana kadar geçirdiği ortalama süreyi hesaplayalım.

Sınıflanmış serilerde kullanılan aritmetik ortalama formülünü kullanarak, ortalama süreyi,

$$\bar{X} = \frac{\sum n_i X_i}{\sum n_i} = \frac{440}{40} = 11 \text{ hafta}$$

olarak hesaplıyoruz.

Şimdi de ortalama sürenin ortalama olarak ne kadar sapabileceğini de standart sapma kullanarak hesaplayalım.

Sınıflanmış serilerde kullanılan standart sapma formülü ile,

$$\sigma = \sqrt{\frac{\sum n_i(m_i - \bar{X})^2}{\sum n_i}} = \sqrt{\frac{1144}{40}} = \sqrt{28,6} \cong 5,35 \text{ hafta}$$

ve varyans formülü kullanılarak da, varyans;

$$\sigma^2 = \frac{\sum n_i(m_i - \bar{X})^2}{\sum n_i} = \frac{1144}{40} = 28,6 \text{ hafta}$$

olarak hesaplanmaktadır.

Elde edilen sonuçlardan, kliniğe başvurarak tüp bebek tedavisi gören kadınların, başvurularından itibaren ortalama olarak 11 haftada gebe kaldıkları ve bu sürenin ortalamadan ortalama olarak + ve – yönde olmak üzere 5,35 hafta saptığı anlaşılmaktadır.

8.4.4. Standart Sapma ve Varyans Hesabında Başka Bir Formül

Standart sapma ve varyans hesabında kullanılan, gözlem değerlerinin aritmetik ortalamadan farkları alınmak suretiyle oluşturulan ve sapma değerlerinin kareleri üzerinden hesaplanmasını ifade eden formülleri, standart sapma ve varyans tanımı üzerinden oluşturmayı ve çeşitli seri örnekleri ile uygulamayı öğrendik.

Standart sapma ve varyansın tanım formülleri olarak nitelenen bu formüller dışında bir formül ile de standart sapma ve varyans hesabı yapmak mümkündür.

Standart sapma ve varyans hesabında kullanılabilecek bu alternatif formül aritmetik ve kareli ortalama üzerinden varyans ve standart sapma hesabına olanak sağlayan formüldür.

Bir serinin aritmetik ve kareli ortalamasının bilinmesi durumunda varyansı,

$$\sigma^2 = \text{Kareli Ortalama}^2 - \text{Aritmetik Ortalama}^2$$

$$\sigma^2 = KO^2 - \bar{X}^2$$

ve standart sapmayı da;

$$\sigma = \sqrt{\text{Kareli Ortalama}^2 - \text{Aritmetik Ortalama}^2}$$

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

formülleriyle hesaplamak mümkün olmaktadır.

Şimdi kareli ve aritmetik ortalamayı kullanarak standart sapma ve varyans hesabının nasıl yapıldığına yönelik bir örnek çözelim.

Örnek:

Bir futbol takımının sezon açılışı öncesi yaptığı 5 hazırlık maçında attığı gol sayısı aşağıda verilmektedir. Buna göre, takımın attığı gol sayısı için, aritmetik ortalama, kareli ortalama, standart sapma ve varyansı hesaplayınız.

Puanlar X_i	X_i^2
2	4
2	4
3	9
3	9
5	25
$\sum X_i = 15$	$\sum X_i^2 = 51$

Buna göre, futbol takımının maç başına attığı ortalama gol sayısı, aritmetik ortalama ile;

$$\bar{X} = \frac{\sum X_i}{n} = \frac{15}{5} = 3 \text{ gol}$$

kareli ortalama ile;

$$KO^2 = \frac{\sum X_i^2}{n} = \frac{51}{5} = 10,2$$

$$KO = \sqrt{10,2} = 3,19 \text{ gol}$$

hesaplanmaktadır.

Varyans ve standart sapmayı da, aritmetik ve kareli ortalama değerlerini kullanarak hesaplayalım. Varyans,

$$\sigma^2 = KO^2 - \bar{X}^2$$

$$\sigma^2 = 10,2 - 9 = 1,2$$

ve standart sapmada;

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

$$\sigma = \sqrt{10,2 - 9} = \sqrt{1,2} = 1,095 \text{ gol}$$

şeklinde hesaplanmaktadır.

Bu çözüm, varyans ve standart sapmanın hesaplanmasında daha pratik bir yöntem olarak benimsenebilir.

Ancak, bu noktada iki önemli hususu hatırlatmakta fayda görüyoruz:

İlk olarak, kullanılacak aritmetik ve kareli ortalama formülleri hesaplama yapılacak seri türüne göre seçilmelidir. Burada basit seri örneği olduğundan, basit seri durumunda kullanılmakta olan aritmetik ve kareli ortalama formülleri kullanılmıştır.

İkinci olarak, kareli ve aritmetik ortalamalar ile hesaplanan standart sapmanın; $\sigma = \sqrt{KO^2 - \bar{X}^2}$ biçiminde olan formülünde ortalamaların kareleri ile karekök işleminin sadeleştirilemeyeceğidir. Zira, aradaki işaret – işaretidir ve bu durum sadeleştirmeye yani kare alma işleminin karekök alma işlemini yok etmesine sebep olamaz.

8.4.5. Standart Sapmanın Özellikleri

Standart sapma seride yer alan tüm gözlem değerlerini hesaba katan analitik ve duyarlı bir dağılma ölçüsü olarak uygulamada oldukça yaygın kullanıma sahiptir.

Standart sapmanın, ortalamadan sapma değerlerinin karesi alınarak hesaplanan bir kareli ortalama özelliği taşıması, onu cebirsel işlemlere elverişli kılmakta ve bu sebeple, standart sapma ileri istatistik analizlerin temelinde var olan bir dağılma ölçüsü olarak karşımıza çıkmaktadır.

Öte yandan, standart sapma, alt sınırı sıfır olan ancak, üst sınırı bulunmayan bir dağılma ölçüsüdür. Bir uç durum olarak, seride yer alan bütün birimlerin birbirine eşit olması halinde, ortalama da aynı olacak ve ortalamadan sapma değerleri serideki tüm gözlem değerleri için sıfır çıkacaktır. Dolayısıyla, standart sapmanın alt sınırı sıfırdır ve bu durumda gözlem değerleri tek bir değerde toplanmıştır ve böyle bir durumda herhangi bir dağılım söz konusu değildir.

Ancak, standart sapmayı yorumlarken asıl ihtiyacımız olan bir dağılımın en yüksek değişkenliğe sahip olduğu, en fazla dağılımın yaşandığı en uç durumu bilmektir ki standart sapmanın değerini üst sınır değeri ile kıyaslayarak dağılımın derecesi hakkında fikir sahibi olabilelim.

Alt sınırı belli ancak üst sınırı açık olan standart sapmanın değerine bakarak, yüksek bir dağılma derecesini mi yoksa düşük bir dağılma derecesini mi işaret ettiğini anlamak oldukça güçtür ve bunun için ancak serinin yapısına ve gözlem değerlerinin seyrine bakarak karar vermek mümkündür.

Standart sapma, hesaplandığı serinin ölçü birimiyle ifade edilmektedir. Seri hafta şeklinde düzenlenmiş bir zaman ifade ediyorsa, standart sapmanın birimi hafta, kilogram ile ölçülmüş bir veri için, kilogram, metre ile ölçülmüş bir veri söz konusu olduğunda ise metre cinsindendir.

Dolayısıyla, eğitim hayatımızın başından beri bildiğimiz o temel kurala, elmalarla armutların toplanamayacağı, kıyaslanamayacağı kuralına, ters bir durum söz konusudur. Ölçü birimleri farklı olan serilerin dağılma derecelerini standart sapma ile kıyaslamak mümkün değildir. Örneğin, elimizde iki seri olduğunu varsayalım ve bu serilerden biri kilogram ile diğeri ton ile ölçülmüş olsun. İkisi de ağırlık ölçüsü olmasına rağmen standart sapmaları kıyaslayarak dağılma derecelerini karşılaştırmak yanıltıcı olur. Zira, kilogram ile ölçülen serinin rakamsal büyüklükleri, ton ile ölçülen seriye kıyasla daha yüksektir ve bu durum kilogram ile ölçülmüş serinin standart sapmasının daha büyük çıkmasına sebep olacaktır.

Öte yandan, ölçü birimi aynı olsa dahi, gözlem değerlerinin rakamsal büyüklüğü fazla olan serilerin standart sapması, küçük olan serilere kıyasla daha büyük çıkacaktır.

Özetle, ölçü birimi aynı ve rakamsal büyüklükleri benzer olan serilerin standart sapma kullanılarak dağılma dereceleri kıyaslanabilir. Fakat, ölçü birimi ya da rakamsal büyüklükleri farklı olan serileri kıyaslamada standart sapma uygun bir ölçü değildir.

Bu Bölümde Neler Öğrendik?

Bir veriyi özetlemede ortalamaların tek başına yeterli ölçüler olamadığını ve ortalamaların yanında birimlerin birbirlerinden ya da ortalamadan ne kadar uzaklaştığını gösteren mutlak dağılma ölçülerine de ihtiyaç duyulduğunu öğrendik.

Mutlak dağılma ölçüleri kapsamında kullanım alanı kısıtlı olan değişim aralığı ve ortalama sapma ölçülerini ve bu ölçülerin kullanımının neden kısıtlı kaldığını ele aldık.

Dağılma ölçüleri konusuyla ilgili olarak istatistik çalışmalarda büyük yeri olan standart sapma ve varyans ölçülerini detaylı olarak inceledik. Standart sapmaya bakarak bir dağılımın homojenlik derecesinin nasıl değerlendirildiğini gördük.

Bölüm Soruları

1- Gözlem değerleri 12, 14, 16, 19, 25, 34 olan serinin değişim aralığı nedir?

a- 22

b- 12

c- 34

d- 23

e- 16

2- Aşağıdaki serinin ortalama sapması nedir?

X_i
6
7
9
10
13

a- 0

b- 2

c- 3

d- 4

e- 5

3- Standart sapması 11,5 olan serinin varyansı nedir?

a- 11,5

b- 21,5

c- 132,25

d- 22,5

e- 3,5

- 4- Varyansı 9 olan bir serinin standart sapması nedir?
- a- 81
 - b- 9
 - c- 18
 - d- 12
 - e- 3
- 5- Aritmetik ortalaması 6 kareli ortalaması 11 olan bir serinin varyansı nedir?
- a- 5
 - b- 17
 - c- 8,5
 - d- 85
 - e- 11
- 6- Aritmetik ortalaması 9 kareli ortalaması 15 olan serinin standart sapması nedir?
- a- 15
 - b- 12
 - c- 81
 - d- 225
 - e- 24

7- Aşağıdaki serinin varyansı nedir?

X_i
3
4
7
9
12

a- 10,8

b- 3,4

c- 9

d- 8,7

e- 7,9

8- Aşağıdaki tasnif edilmiş serinin varyansı nedir?

X_i	n_i
2	1
4	4
6	3
8	2

a- 1,83

b- 3,36

c- 11,29

d- 4,2

e- 6,3

9- Bir seride yer alan bütün birimlerin aynı olması ya da eşit değer alması durumunda standart sapma ne olur?

- a- Bir
- b- Aritmetik ortalamaya eşit olur.
- c- Sıfır
- d- Bir şey söylenemez.
- e- Ortalamanın yarısına eşit olur.

10- Aşağıdaki ifadelerden hangisi yanlıştır?

- I- Değişim aralığı serideki en büyük ve en küçük değer arasındaki farktır.
 - II- Ortalama sapma, birimlerin aritmetik ortalamadan mutlak sapmalarına dayalı bir ölçüdür.
 - III- Standart sapma bir kareli ortalamadır.
 - IV- Ölçü birimi farklı olan serilerin dağılma derecelerinin kıyaslanmasında standart sapma uygun bir ölçü değildir.
 - V- Ölçü birimleri farklı serilerin dağılma derecelerini kıyaslamada standart sapma uygun bir ölçüdür.
- a- I
 - b- II
 - c- III
 - d- IV
 - e- V

Cevaplar

- 1- a
- 2- b
- 3- c
- 4- e
- 5- d

6- b

7- a

8- b

9- c

10- e

Yararlanılan Kaynaklar

ALOBİ, K. B. (2003), *İstatistiksel Analiz Metotları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

9. NİSPİ DAĞILMA ÖLÇÜLERİ

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bir dağılımın dağılma derecesini belirlemede mutlak dağılma ölçülerinin yeterli olmadığı durumlar olabilmektedir. Bu bölümde mutlak dağılma ölçülerinin yetersizliklerini gidermek üzere geliştirilmiş nispi dağılma ölçülerinin neler olduğunu öğreneceğiz.

Birimlerin birbirine ya da ortalamaya uzaklıklarını temel alan mutlak dağılma ölçülerinin, serinin dağılma ve değişkenlik derecesini belirlemede yetersiz kaldığı durumlar da söz konusu olabilmektedir. Örneğin, ölçü birimleri ya da terim büyüklükleri farklı serilerin dağılma derecelerini kıyaslamak istediğimizde standart sapma yanıltıcı olmakta ve yetersiz kalmaktadır. Böyle durumlarda nispi dağılma ölçüleri içinde yer alan değişim katsayısı tercih edilmektedir.

Öte yandan, bir birimin ya da gözlem değerinin ait olduğu dağılım içinde yerini belirlemek için nispi dağılma ölçülerinden standart değerin nasıl kullanıldığını da bu bölümde inceleyeceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Mutlak dağılma ölçüleri, bir dağılımın dağılma derecesini ortaya koymak üzere her koşulda kullanılabilecek ölçüler midir?
- 2) Nispi dağılma ölçüleri nelerdir?
- 3) Ölçü birimi farklı olan serilerin dağılma dereceleri nasıl kıyaslanır?
- 4) Terim büyüklükleri farklı olan serilerin dağılma dereceleri nasıl kıyaslanır?
- 5) Bir birimin, ait olduğu dağılım içinde yeri nasıl belirlenir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Mutlak dağılma ölçülerinin yetersiz kaldığı durumlar.	Mutlak dağılma ölçülerinin ölçü birimi ya da terim büyüklüğü farklı serilerde kullanılamadığını bilir.	Dağılım ve değişkenlik konusunda çok sayıda temel düzeyde istatistik kitabı bulunmaktadır ve her biri detaylı bilgi sahibi olmak için kullanılabilir.
Standart sapmaya yönelik kısıtlar.	Ölçü birimi ya da terim büyüklüğü farklı serilerde standart sapmanın sakıncalarını bilir.	Dağılım ve değişkenlik konusunda çok sayıda temel düzeyde istatistik kitabı bulunmaktadır ve her biri detaylı bilgi sahibi olmak için kullanılabilir.
Değişim Katsayısı	Ölçü birimi ya da terim büyüklükleri farklı serilerde değişim katsayısının önemini bilir.	Dağılma ölçülerine yönelik olarak basit düzeyde yazılmış istatistik kitaplarında ilgili konulara ilişkin çok sayıda örnek çözüm bulunabilir.
Değişim katsayısının sınırları	Değişim katsayısının sınırları belli bir nispi dağılma ölçüsü olduğunu bilir.	Temel düzeyde istatistik kitaplarından değişim katsayısının kullanımına yönelik çeşitli örnekler incelenebilir.
Standart değer kavramı ve standart değer gerekliliği.	Özellikle çan eğrisi olarak öğrenciler tarafından bilinmekte olan not sistemi başta olmak üzere standart değer ne olduğunu ve nasıl kullanıldığını bilir.	Temel düzeyde yazılmış istatistik kitaplarda standart değer hesabına yönelik örnekler incelenebilir.

Anahtar Kavramlar

- Nispi Dağılma Ölçüleri: Birimlerin ortalamaya kıyasla dağılma derecesini belirlemeye yarayan ölçüler.
- Değişim Katsayısı: Ölçü birimi ya da terim büyüklüğü farklı olan serilerde kullanılan dağılma ölçüsü.
- Standart Değer: Bir dağılım içinde yer alan herhangi bir gözlem değerinin, dağılım içinde yerini belirlemede kullanılan ölçü.

Giriş

Daha önceki bölümde mutlak dağılma ölçülerini öğrendik. Mutlak dağılma ölçüleri, gözlem değerlerinin birbirlerinden ya da aritmetik ortalamadan ne kadar uzaklaştığını gösteren ölçülerdir.

Mutlak dağılma ölçüleri içinde en önemli yere sahip olan standart sapma ve varyans da istatistik çalışmalarda çok fazla kullanılan ve hemen her istatistik analiz yönteminin içinde yer bulan ölçülerdir.

Bu derece yaygın kullanılan ölçüler olmakla birlikte, standart sapma ve varyans terim büyüklüğünün ve ölçü biriminin etkisi altındadır. Bu özelliği sebebiyle de özellikle ölçü birimleri ya da terim büyüklükleri farklı olan serilerin dağılma dereceleri kıyaslanmak istendiğinde söz konusu iki ölçü yetersiz ve yanıltıcı olmaktadır. Nispi dağılma ölçüleri özellikle böyle durumlarda tercih edilen ve bahsedilen sıkıntıları ortadan kaldıran dağılma ölçüleridir.

Nispi dağılma ölçüleri kapsamında değişim katsayısını ve standart değer kavramını inceleyeceğiz.

9.1. Değişim Katsayısı

Değişim katsayısı, bir dağılımın dağılma derecesini nispi olarak ortaya koyan ve yine dağılma ölçüleri başlığı altında değerlendirilen bir ölçüdür.

Değişim katsayısı, bir dağılımın dağılma derecesini başka bir deyişle homojen olup olmadığını ortaya koymak üzere geliştirilmiş nispi bir dağılma ölçüsüdür.

Standart sapmanın ölçü birimi ve/veya rakamsal büyüklükleri farklı olan serilerin dağılma derecelerini kıyaslamada uygun bir ölçü olmadığını ilgili başlık altında dile getirmiştik.

Öte yandan, standart sapmanın sınırları belli olmayan bir dağılma ölçüsü olması sebebiyle, hesaplanan standart sapma değerinin nasıl bir dağılma derecesi ifade ettiğini, yani dağılma derecesi az bir seri ile mi yoksa dağılma derecesi yüksek bir seri ile mi karşı karşıya olduğumuzu anlamada yetersiz kaldığını da ifade ettik.

Standart sapmanın yukarıda açıklamaya çalıştığımız zayıflıklarını ortadan kaldırmak üzere Değişim Katsayısı geliştirilmiştir. Nispi bir dağılma ölçüsü olan Değişim Katsayısı, aşağıdaki formül vasıtasıyla hesaplanmaktadır:

$$\text{Değişim Katsayısı} = \frac{\sigma}{\bar{X}} \cdot 100$$

Değişim Katsayısı formülünü incelediğimizde, serinin standart sapmasının ortalamasına oranlandığını ve çıkan sonucun 100 ile çarpıldığını görüyoruz.

Dağılımın standart sapmasının dağılımın ortalamasına oranlanması ile her iki ölçü de aynı ölçü birimiyle ifade edildiğinden, bu oranlama sonrasında ölçü biriminden bağımsız bir katsayı elde etmiş oluruz. Öte yandan, serinin rakamsal büyüklüğü hem ortalama hem standart sapma üzerinde etkili olduğundan, oranlama işlemiyle rakamsal büyüklükten kaynaklanan yorumlamaya yönelik sorunlar da böylelikle ortadan kalkmış olmaktadır.

Değişim katsayısı bir dağılımın dağılma derecesi ile ilgili olarak oldukça sağlıklı karar vermemizi sağlayan bir ölçü olmakla beraber; kullanım alanı dağılımın özelliğini tanımlama amacıyla sınırlı kalmakta ve ileri istatistik analizler için uygun bir ölçü olamamaktadır.

Değişim Katsayısının istatistik çalışmalarındaki önemini şöyle bir örnekle açıklamaya çalışalım:

Örnek:

Aşağıda aynı veri kilogram ve ton ölçü birimleri kullanılarak düzenlenmiştir. Önce ortalama ve standart sapmalarını, daha sonra da değişim katsayılarını hesaplayalım.

Ton ile Ölçülen Seri (A)			Kilogram ile Ölçülen Seri (B)		
X_i (ton)	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	X_i (kg)	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	-1,5	2,25	1000	-1500	2250000
2	-0,5	0,25	2000	-500	250000
3	+0,5	0,25	3000	+500	250000
4	+1,5	2,25	4000	+1500	2250000
$\sum X_i = 10$		$\sum (X_i - \bar{X})^2 = 5$	$\sum X_i = 1000$		$\sum (X_i - \bar{X})^2 = 5000000$

İlk olarak ton ile ölçüm yapılan A serisinin ortalamasını ve standart sapmasını hesaplayalım.

Aritmetik ortalama,

$$\bar{X} = \frac{\sum X_i}{n} = \frac{10}{4} = 2,5 \text{ ton}$$

ve standart sapma da,

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{5}{4}} \cong 1,118 \text{ ton}$$

hesaplanmaktadır.

Şimdi de kilogram ile ölçüm yapılan B serisinin ortalamasını ve standart sapmasını hesaplayalım. Sırasıyla dağılımın ortalama ve standart sapması aşağıdaki gibi hesaplanmıştır.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1000}{4} = 2500 \text{ kg}$$

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{5000000}{4}} \cong 1118 \text{ kg}$$

Aynı verinin ton ve kilogram olarak farklı ölçü birimleriyle ölçülmesi neticesinde, serilerin gerek ortalamaları gerek standart sapmaları farklı çıkmıştır. Zira, hesaplanan ortalama ve standart sapmaların da ölçüleri farklıdır.

Standart sapmalar kıyaslandığında, B serisinin standart sapmasının 1118, A serisinin standart sapmasının ise 1,118 olması, B serisinin dağılım derecesi daha yüksek bir seri gibi değerlendirilmesine yol açar. Oysa, her iki seri de aynı veriye ilişkin serilerdir.

Şimdi her iki serinin değişim katsayılarını hesaplayalım.

A serisinin değişim katsayısı,

$$\text{Değişim Katsayısı} = \frac{\sigma}{\bar{X}} \cdot 100$$

$$\text{Değişim Katsayısı} = \frac{1,118}{2,5} \cdot 100 = \%44,72$$

B serisinin değişim katsayısı,

$$\text{Değişim Katsayısı} = \frac{\sigma}{\bar{X}} \cdot 100$$

$$\text{Değişim Katsayısı} = \frac{1118}{2500} \cdot 100 = \%44,72$$

olarak hesaplanmaktadır.

Aynı verinin farklı ölçü birimleriyle ölçüldüğü yukarıdaki iki seriye baktığımızda, standart sapmaların ölçü birimi ve rakamsal büyüklük etkisi altında kalarak farklı sonuçlar aldığını ancak, değişim katsayısının bu durumdan etkilenmeyerek aynı sonucu verdiği görülmektedir. Buna göre, her iki seride % 44,72 oranında bir dağılım derecesine sahiptir.

Değişim Katsayısının hesaplanmasına ve yorumlanmasına yönelik bir örnek uygulama daha yapalım:

Örnek:

Aşağıda iki ülkenin son 5 yıllık dönemde et tüketimine yönelik olarak hesaplanmış bazı istatistikler verilmiştir. Buna göre, hangi ülkede et tüketiminin daha homojen olduğu söylenebilir?

Ülkeler	Ortalama Et Tüketimi	Standart Sapma
A ülkesi	2350 kg	750kg
B Ülkesi	88450 kg	10250 kg

Her iki ülkenin de et tüketimi kg cinsinden olmakla beraber, değerlerin rakamsal büyüklükleri birbirinden çok farklıdır. Bu durumda, standart sapmaları kıyaslama yoluna giderek hangi ülkede et tüketiminin daha homojen olduğunu belirlemek istersek yanılırız. Zira, standart sapması büyük olan serinin daha çok dağıldığını düşünürüz oysa bu durum rakamsal büyüklük sebebiyle oluşmuştur.

Bu durumda, her iki ülkenin değişim katsayılarını hesaplamak suretiyle hangi dağılımın daha homojen olduğuna karar vermek daha doğru bir yaklaşım olacaktır.

Şimdi ülkelerin değişim katsayılarını hesaplayalım.

A ülkesinin değişim katsayısı,

$$\text{Değişim Katsayısı} = \frac{\sigma}{\bar{X}} \cdot 100$$

$$\text{Değişim Katsayısı} = \frac{750}{2350} \cdot 100 = \%31,91$$

B ülkesinin değişim katsayısı ise,

$$\text{Değişim Katsayısı} = \frac{\sigma}{\bar{X}} \cdot 100$$

$$\text{Değişim Katsayısı} = \frac{10250}{88450} \cdot 100 = \%11,59$$

olarak hesaplanmıştır.

Değişim katsayılarının sonuçlarına göre A ülkesinde et tüketimi % 31,91 oranında dağılmaktayken, B ülkesinin et tüketimi için dağılım oranı % 11,59'dur.

Bu durumda, % 11,59 dağılım oranı ile B ülkesinin et tüketiminin, A ülkesinin et tüketiminden daha homojen olduğunu söylemek doğru olacaktır.

9.2. Standart Değer

Dağılım ölçüleri bölümü içinde son olarak standart değer kavramına değinmek istiyoruz.

Standart değer kavramı şimdiye değin öğrendiğimiz dağılım ölçüleri gibi, bir dağılımın dağılım derecesi ya da homojenlik derecesini belirlemeye yönelik bir ölçü değildir. Standart değer, bir dağılım içinde yer alan herhangi bir gözlem değerinin dağılımın ortalamasına göre hangi noktada yer aldığını tespit etmek ve bu gözlem değerini farklı dağılımlarda bulunan farklı gözlem değerleriyle kıyaslayabilmek amacıyla hesaplanan bir ölçüdür.

Standart değer hesabı, gerçekte bir dönüşüm ya da dönüştürme işlemidir. Ortalamaları ve standart sapmaları farklı farklı olan dağılımlarda yer alan gözlem değerlerini, standart hale dönüştürerek, bu gözlem değerlerini sağlıklı ve doğru şekilde kıyaslayabilme olanağı verir.

Söz gelimi, farklı iki sınıfta sınava girmiş iki öğrencinin başarısını kıyaslamak istersek, öğrencilerin 100 üzerinden aldıkları notları karşılaştırmak ve yüksek not alanı başarılı kabul etmek mutlak başarıya dayalı bir değerlendirme ölçüsü olacaktır.

Oysa, öğrencileri sınava birlikte girdiği sınıfı içinde değerlendirmek daha doğru bir yaklaşımdır.

Örneğin bir sınıfta herkesin 60 aldığı bir sınavda alınan 75, herkesin 100 aldığı sınavdan alınan 90'dan daha başarılı bir sonucu ifade eder.

İşte standart değer kavramı, bir seride yer alan herhangi bir gözlem değerinin, dağılımın ortalamasına ve standart sapmasına göre nasıl konumlandığını gösteren bir ölçüdür.

Daha teknik bir ifade ile gözlem değerinin kendi ortalamasından kaç standart sapma uzakta yer aldığını gösterir. Yani işin özü bir standartlaştırma dönüşümüdür.

Bu işlem, standart bir dağılım yaratmak ve gözlem değerlerini farklı dağılımlara ait olsalar bile doğru ve sağlıklı bir şekilde kıyaslayabilmek amacıyla yapılır.

Standart değer dönüşümü,

$$z_i = \frac{(X_i - \bar{X})}{\sigma}$$

formülüyle yapılır. Böylelikle, ortalaması ve standart sapması farklı serilerde yer alan X_i değerleri; yukarıdaki dönüşüm formülüyle z_i değerlerine dönüştürülmüş olur.

z standart değerleri, ortalaması sıfır ve varyans ve standart sapması 1 olan standart bir dağılım oluşturur. Dolayısıyla, bu dönüşüm işlemiyle, ortalaması ve standart sapması farklı tüm dağılımlar, standart bir dağılıma dönüştürülmüş olmakta ve böylelikle her bir gözlem değeri diğer dağılımlardaki gözlem değerleriyle kıyaslanabilir hale gelmektedir.

z değerinin pozitif olması ortalamanın üzerinde bir değer olduğunu, negatif olması ise ortalamanın altında bir değer olduğunu gösterir. $z=0$ değeri de kolayca tahmin edilebileceği gibi ortalamaya karşılık gelen z değeridir.

Standart değer konusu öğrencilerin aşına olduğu not değerlendirme sistemi olan ve öğrenciler arasında bağıl not sistemi ya da çan eğrisi olarak adlandırılan başarı değerlendirme sisteminin özüdür.

Bu sistemde amaç, bir öğrencinin aldığı notu, sınıf ortalaması ve standart sapmasını kullanarak standardize etmek ve not hangi standart nota karşılık geliyorsa öğrenciye o harf notunu vermek şeklinde özetlenebilir.

Şimdi standart değer, nasıl hesaplandığını ve nasıl yorumlandığını bir örnek çözümle görelim.

Örnek:

Aşağıda bir öğrencinin matematik ve istatistik derslerinden aldığı notlar ve her iki dersin sınıf ortalamaları ve standart sapmaları veriliyor. Buna göre öğrencinin hangi derste daha başarılı olduğunu belirlemeye çalışalım.

Dersin Adı	Öğrencinin Notu	Sınıf Ortalaması	Sınıf Standart Sapması
Matematik	80	65	12
İstatistik	70	45	8

Öğrencinin sadece, iki ders için aldığı notları kıyasladığımızda, Matematik dersinde 80, İstatistik dersinde ise 70 aldığını görüyor ve matematikte daha yüksek not aldığı için bu derste daha başarılı olduğuna karar veriyoruz. Bu değerlendirme, öğrenciyi 100 üzerinden aldığı nota göre mutlak olarak değerlendirmedir.

Derslerin ortalamalarına baktığımızda, matematik dersinde ortalamanın 65 olduğunu ve bunun istatistik dersi ortalaması olan 45'den daha yüksek olduğunu görüyoruz. Dolayısıyla, sınıfın geneli açısından matematik sınavı istatistik sınavına kıyasla daha başarılı bir sınav olmuştur.

Öğrencimiz ise başarısı yüksek olan dersten yüksek not almıştır. Nitekim öğrenci, matematik dersinde ortalama nottan çok fazla yukarıya çıkamamış ancak sınıf ortalamasının düşük olduğu istatistik dersinde, sınıf ortalamasından bir hayli yüksek not almıştır.

Şimdi, öğrencinin her iki derste aldığı notları standart hale getirelim:

Standart değer dönüşüm formülünü uygulayarak,

$$z_i = \frac{(X_i - \bar{X})}{\sigma}$$

Matematik dersi için standart değeri,

$$z_{mat} = \frac{(80 - 65)}{12} = +1,25$$

İstatistik dersi içinse standart değeri,

$$z_{ist} = \frac{(70 - 45)}{8} = +3,125$$

hesaplıyoruz.

Standart değer sonuçlarına baktığımızda öğrencinin matematik dersinde aldığı not (80), sınıf ortalamasından 1,25 standart sapma yukarıda yer alırken, istatistik dersinde aldığı 70, sınıf ortalamasından 3,125 standart sapma yukarıda yer almaktadır. Dolayısıyla öğrenci, istatistik dersinde daha başarılı olmuştur.

Bu Bölümde Neler Öğrendik?

Mutlak dağılma ölçüleri kapsamında yer alan ve tüm istatistik çalışmalarda büyük yeri olan standart sapma ve varyans ölçülerinin kullanım alanını sınırlayan kısıtları öğrendik.

Standart sapma ve varyansın ölçü birimi ya da terim büyüklüğü farklı olan serilerin dağılma derecelerinin kıyaslanmasında yanıltıcı olabileceğini, gördük. Öte yandan, standart sapmanın üst sınırının bulunmamasının da ölçünün önemli bir zayıflığı olduğunu öğrendik.

Nispi dağılma ölçüsü olarak adlandırılan ölçüler standart sapmaya yönelik sözü edilen eleştirileri ortadan kaldırmak üzere geliştirilmiş ölçülerdir.

Nispi dağılma ölçülerini ele aldığımız bu bölümde değişim katsayısı ile standart değer nasıl hesaplandığını, yorumlandığını ve ne amaçla kullanıldıklarını öğrendik.

Bölüm Soruları

- 1-terim büyüklüğünün ve ölçü biriminin etkisi altında olmayan bir dağılma ölçüsüdür.
 - a- Değişim katsayısı
 - b- Değişim aralığı
 - c- Ortalama sapma
 - d- Standart sapma
 - e- Varyans
- 2- Herhangi bir gözlem değerinin bir dağılım içinde nerede yer aldığını belirlemedekullanılır.
 - a- Değişim katsayısı
 - b- Standart değer
 - c- Standart sapma
 - d- Varyans
 - e- Ortalama sapma
- 3- Ortalamaya eşit bir gözlem değerinin standart değeri nedir?
 - a- Birdir
 - b- Ortalama değeridir
 - c- Sıfırdır
 - d- Pozitif bir değerdir
 - e- Negatif bir değerdir
- 4- Bütün birimleri aynı olan bir dağılımın değişim katsayısı ne olur?
 - a- Bir
 - b- Ortalama değere eşit
 - c- Ortalamanın yarısına eşit olur
 - d- Pozitif bir değer olur
 - e- Sıfır

- 5- Standart sapması 5, aritmetik ortalaması 11 olan dağılımın değişim katsayısı nedir?
- a- % 2,2
 - b- % 0,4545
 - c- % 220
 - d- % 45,45
 - e- % 6
- 6- Ortalaması 65 ve standart sapması 12 olan bir sınavda 80 notunun standart değeri nedir?
- a- 2,3
 - b- 1,25
 - c- 1,4
 - d- 1,75
 - e- 1,35
- 7- Bir dağılım içinde ortalamanın üstünde yer alan bir birimin standart değeri nasıl değer alır?
- a- Pozitif değer alır
 - b- Negatif değer alır
 - c- Sıfır olur
 - d- Bir değerini alır
 - e- Bir şey söylenemez.
- 8- Bir dağılım içinde ortalamanın altında yer alan bir birimin standart değeri nasıl değer alır?
- a- Pozitif değer alır
 - b- Negatif değer alır
 - c- Sıfır olur
 - d- Bir değerini alır
 - e- Bir şey söylenemez

9- Bir dağılımın aritmetik ortalaması ve standart sapmasının aynı değere sahip olması durumunda değişim katsayısının değeri ne olur?

- a- % 0
- b- % 1
- c- % 100
- d- Ortalamaya eşit olur
- e- Standart sapmaya eşit olur

10- Aritmetik ortalaması 80 ve standart sapması 10 olan bir dağılımda 70'e karşılık gelen z değeri nedir?

- a- 70
- b- 80
- c- 0
- d- 1
- e- -1

Cevaplar

- 1- a
- 2- b
- 3- c
- 4- e
- 5- d
- 6- b
- 7- a
- 8- b
- 9- c
- 10- e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

10. OLASILIK – TEMEL KAVRAMLAR

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde öncelikle, olasılığa yönelik objektif ve subjektif olasılık tanımlarını öğrenecek ve günlük yaşamımız içinde olasılığın yerini değerlendireceğiz.

Bir tesadüfi deneyin sonsuz sayıda tekrarı sonucunda ortaya çıkabilecek limit durumda tanımlı olasılık kavramından söz ederek, basit ve bileşik olayların olasılıklarının nasıl hesaplandığı üzerinde duracağız. Öte yandan, bağdaşan ve bağdaşmayan olaylar ve olasılık hesabında önemine, koşullu olasılık, bağımlı ve bağımsız olay kavramlarına, permütasyon ve kombinasyon analizi kavramlarına yönelik bilgi sahibi olacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Geçmişini biliyoruz, geleceğimizi ise olasılıkları değerlendirerek planlıyoruz. O halde nedir bu olasılık?
- 2) Olasılıkları doğru değerlendirmek belirsizliği azaltır mı?
- 3) İhtimalleri doğru değerlendirmek başarı olasılığını artırır mı?
- 4) Olasılığın yüksek olması sürpriz sonuçlar doğurabilir mi?
- 5) Olasılığın artması ya da azalması karar süreçlerinde neden etkilidir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Olasılığın objektif ve sübjektif tanımı, bir limit durum olarak olasılık.	Hayatın içinden bir kavram olarak olasılık konusuna ilişkin temel bilgi sahibi olur.	Temel düzeyde istatistik kitaplarının olasılık bölümleri
Basit - bileşik olaylar	Olasılık konusu içinde yer alan temel düzeyde olasılık hesaplarına yönelik temel prensipler ve uygulamaları bilir.	Temel düzeyde istatistik kitaplarının olasılık bölümleri
Bağdaşan ve bağdaşmayan olaylar	Olasılık konusu içinde yer alan bağdaşan ve bağdaşmayan olay tanımına yönelik temel prensipler ve uygulamaları bilir.	Temel düzeyde istatistik kitaplarının olasılık bölümleri
Koşullu olasılık ve bağımsızlık kavramı, bağımlı olaylar.	Olasılık konusu içinde yer alan koşullu olasılık ve istatistik bağımsızlık kavramlarını bilir.	Temel düzeyde istatistik kitaplarının olasılık bölümleri
Permütasyon ve kombinasyon analizleri ve olasılık hesapları.	Permütasyon ve kombinasyon analizlerini bilir.	Temel düzeyde istatistik kitaplarının olasılık bölümleri

Anahtar Kavramlar

- Olasılık: Hayatın içinden bir kavram olarak, bugünü ve yarını programlamada etkin bir araç.
- Sübjektif Olasılık: Kişisel algı, beklenti ve değerlendirmelere göre belirlenen olasılık değeri.
- Objektif olasılık: elverişli sonuç sayısının toplam sonuç sayısına oranlanmasıyla elde edilen olasılık değeri, frekans olasılığı.

Giriş

İnsanoğlunun şu dünyada yaşamına yönelik tek gerçeklik, değişmezlik veya kesinlik tamamladığı geçmişinde. Gelecek ise, alternatif planlar, beklentiler, koşullanmalar vb. pek çok insani durum ve koşul altında şekillenen bir belirsizlikler hali.

Geleceğe yönelik planlar ve programlar yapılmış olsa bile, geleceğin belirsizliği o kesin gözüyle baktığımız plan ve programlar üzerinde bile etkili. Örneğin, uçak biletinizi aldınız, bir seyahate gidiyorsunuz, her şey hazır, ancak o sabah çalmayan bir saat, uçağa yetişememeniz ve uçak yolculuğu ile başlayacak tatilinizin ve o kesin gözüyle baktığınız tatil programınızın hiç hesapta olmayan değişimi.

Ya da benzer şekilde biletleri alınmış bir tiyatro oyununa gideceksiniz, her şey hazır, program yapılmış, önce dostlarla yemek yenecek, sonra tiyatro, derken gelen bir telefon mutlaka katılmanız gereken bir toplantı haberi. Hani her şey hazırdı, tiyatroya gidiyorduk.

Bunlar gibi örnekleri çoğaltabiliriz. Bu örneklerdeki ortak nokta geleceğin hiçbir zaman mutlak kesinlik taşımadığı ve olasılık kavramının içinde bulunduğumuz andan itibaren yaşamımızın değişmez bir unsuru olduğu.

O halde gelecek bu denli belirsizlik içindeyken insanoğlu yaşamını nasıl planlıyor? Bu noktada karşımıza sübjektif ve objektif olasılık tanımlarıyla olasılık kavramı çıkıyor.

10.1. Olasılık Kavramı

Olasılık kavramına yönelik olarak günlük hayatımız içinde iki farklı tanım ve yaklaşım olduğunu görüyoruz. Bunlar, sübjektif olasılık ve objektif olasılık olarak karşımıza çıkıyor. Şimdi olasılığı tanımlamada kullanılan ancak özünde çok başka anlamlar içeren bu iki yaklaşımı açıklamaya çalışalım:

10.1.1. Sübjektif Olasılık

İnsan zihni gün içinde devamlı bir olasılıklar hesaplama ve kendince en yüksek gördüğü ya da en olabilir bulunduğu duruma göre karar vermekte. Şüphesiz bu olasılık hesabı kâğıt kalem ile yapılan türden bir hesap değil.

Sözelimi sabah evden çıkıyorsunuz, gökyüzüne bakarak, havayı bulutlu gördüyseniz yağmur yağma olasılığını yüksek buluyor ve şemsiyenizi yanınıza alıyorsunuz. Yağmur yağarsa, olasılığı doğru tahmin etmiş oluyor ve ıslanmıyorsunuz; ancak eğer yağmur yağmazsa olasılık tahmininiz başarısız oluyor ve bütün gün şemsiyenizi boşuna taşımış oluyorsunuz.

Benzer şekilde, öğrencisiniz bir sınava gireceksiniz, çok fazla vaktiniz yok, bütün ders notlarına bakmanız mümkün değil. Öğretim üyesinin derste daha fazla üzerinde durduğu konuları potansiyel sınav sorusu olarak görüyor ve bu konulardan soru çıkma olasılığını yüksek olarak değerlendiriyor ve sadece bu konulara çalışıyorsunuz. Olasılık tahmininiz başarılıysa, sınavınız iyi geçiyor ya da tersi.

Tuttuğunuz futbol takımının maçı var, takımın daha önceki maçlarda performansı muhteşem. Rakip takımla yapılmış daha önceki maçlarda hep üstünlük sağlanmış. Bu maçı da kazanma olasılığı sizce çok yüksek. Bunlar hep içinde bulunulan mevcut durum ve geçmişin bir değerlendirmesi sonucunda sizde oluşan beklenti. Maçın sonucu ise ya beklediğiniz gibi galibiyetle sonuçlanır; ya da tersi.

Dikkat edilecek olursa hayatımız yukarıda verdiğimiz birkaç örnekte olduğu gibi, devamlı bir olasılıklar değerlendirmesi üzerinde ilerliyor.

Buradaki örneklerin içinde yer alan olasılık, sübjektif yani öznel olasılık. Kişiden kişiye değişebilen özellik taşıyor. Sözelimi, sabah siz gökyüzüne bakarak yağmur yağma olasılığını yüksek görürken; karşı komşunuz aynı gökyüzüne bakarak yağmur yağmama olasılığını yüksek bulabiliyor. Ya da siz futbol maçını, taraftarı olduğunuz takımın kazanma olasılığını yüksek bulurken, bir başka futbol sever diğer takımın galibiyet olasılığını daha yüksek görebiliyor.

Sübjektif olasılık, kişilerin deneyimlerine, inançlarına ve içinde bulundukları koşullara göre, tamamen bireysel kararlarıyla yarattıkları olasılıkları ifade eder. İnsanlar arasında farklılıklar gösterir yani değişkendir.

Sübjektif olasılık hayatımızın içinde olmakla beraber istatistiğin konusu içinde yer almaz.

İstatistik, objektif yani nesnel olasılık ile ilgilenir.

10.1.2. Objektif Olasılık - Frekans Olasılığı

Objektif olasılık, sübjektif olasılık gibi kişiden kişiye değişmez. Olasılık kuramı gereği, belirli bir olay için, belirli koşullar altında tek bir olasılık değeri vardır ve bu değer herkes için aynıdır.

Sözelimi, bir tarafı yazı diğer tarafı tura olan hilesiz bir parayı havaya atmanız halinde, yazı gelme olasılığının $\frac{1}{2}$ olduğunu düşünürsünüz. Bu durum sizin için de, dünya üzerinde yaşayan tüm insanlar için de aynıdır ve $\frac{1}{2}$ 'dir.

Objektif ya da nesnel olasılık hepimizin bildiği o klasik olasılık formülü ile ortaya çıkar:

$$\text{Olasılık} = \frac{\text{Elverişli sonuç sayısı}}{\text{Toplam sonuç sayısı}}$$

$$P = \frac{m}{n}$$

Hilesiz bir paranın atılması örneğinde, yazı gelme olasılığını ele alalım. Paranın yazı ve tura olmak üzere iki yüzü bulunmaktadır. Yazı gelme olasılığı ile ilgilendiğimize göre, elverişli sonuç yazı olma durumudur ve bir tanedir. Toplam sonuç sayısı ise, sadece yazı ve tura durumları söz konusu olabildiğinden ikidir.

Bu durumda, atılan paranın yazı gelme olasılığı;

$$P = \frac{m}{n} = \frac{1}{2}$$

olmaktadır.

Benzer şekilde, iki parayı aynı anda attığımızı varsayalım. Bu durumda, karşılaşılabilecek mümkün durumlar,

(Tura- Tura), (Tura-Yazı), (Yazı-Yazı), (Yazı-Tura)

biçiminde olacaktır.

İki paranın da tura gelmesi olayı (Tura-Tura durumu) ile ilgileniyorsak, olasılık,

$$P = \frac{m}{n} = \frac{1}{4}$$

ve benzer şekilde iki paranın da aynı yüzünün gelmesi olayı (Tura-Tura ve Yazı-Yazı durumları) ile ilgileniyorsak olasılık,

$$P = \frac{m}{n} = \frac{2}{4}$$

olacaktır.

Dikkat edilecek olursa, olasılık hesabı elverişli sonuçları sayma prensibine dayalı olarak hesaplanmaktadır. Başka bir deyişle, olasılık, elverişli sonuç özelliği taşıyan sonuçların frekansının, toplam sonuç sayısına yani toplam frekansa oranlanmasıyla hesaplanmaktadır. Bu sebeple, olasılığa **frekans olasılığı** da denmektedir.

10.2. Olasılık Kuramına Yönelik Temel Kavramlar:

Olasılık kuramı, tesadüfi deneyler ve bunların sonuçlarıyla ilgilidir. Tesadüfi deney, mümkün sonuçları bilinen, ancak hangi sonucun gerçekleşeceği bilinmeyen işlemlerdir. Örneğin bir zar atıldığında birden altıya kadar bir sayı geleceğini biliriz ancak hangi sayının geleceğini bilemeyiz ve bunu ancak deney sonuçlandığında öğrenebiliriz.

Benzer şekilde, para atıldığında, yazı ya da tura gelebileceğini biliriz ancak, hangisinin geleceğini bilemeyiz.

Burada söz konusu olan tesadüfi değişkenin alabileceği bütün değerlerin oluşturduğu kümeye **evrensel küme**, **örnek uzayı** ya da **temel küme** denir.

Söz gelimi, bir paranın atılması deneyinde evrensel küme;

$$E = \{Yazı, Tura\}$$

Bir zarın atılması deneyinde ise evrensel küme;

$$E = \{1,2,3,4,5,6\}$$

olmaktadır.

İki paranın birlikte atılması deneyinde evrensel küme;

$$E = \{YazıYazı, YazıTura, TuraYazı, TuraTura\}$$

İki zarın birlikte atılması deneyinde ise evrensel küme;

$$E = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), \dots, (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

6x6=36 mümkün sonuçtan oluşturmaktadır.

Bu bilgiler ışığında olasılığı tekrar formüle edelim:

$$Olasılık = \frac{\text{Elverişli sonuç sayısı}}{\text{Mümkün sonuç sayısı}}$$

$$P = \frac{m}{n}$$

Formülde yer alan,
P= olasılık değerini,
m=Elverişli sonuç sayısını,
n= Mümkün sonuç sayısını
göstermektedir.

10.2.1. Olasılık Bir Limit Durumdur

Olasılık, deney sayısı sonsuza giderken ortaya çıkacak bir limit durumu ifade eder. Söz gelimi, bir paranın atılması örneğinde % 50 olasılıkla yazı, % 50 olasılıkla tura gelmesini bekleriz. Bu durumda, parayı iki kez atsak bir kez yazı bir kez de tura gelmelidir. Oysa bu durum pratikte gerçekleşmeyebilir. Her iki atışta da yazı ya da her iki atışta da tura gelmesi mümkündür.

Para atma deneyinin sayısını arttırdığımızda, yazı ve tura sonuç sayılarının yarı yarıya gerçekleşme eğilimi içinde girdiğini görürüz. Örneğin, bir parayı 1000 kere attığımızda yazı ve tura sonuçları sayısı yarı yarıya gerçekleşmeye yaklaşmaktayken; deney sayısı çok çok arttığında, örneğin atış işlemi 1.000.000 kere tekrarlandığında, yazı ve tura gelme sayılarının her iki sonuca eşit olarak dağıldığını gözlemleriz. Dolayısıyla olasılık dediğimiz kavram, deney sayısı sonsuza yaklaştıkça ortaya çıkan bir limit durumu ifade etmektedir. Yani, daha doğru bir ifadeyle olasılık,

$$P = \lim_{n \rightarrow \infty} \frac{m}{n}$$

şeklinde olmaktadır.

10.2.2. Olasılığın Sınırları

Olasılık, alt ve üst sınırı belli olan bir kavramdır. Şöyle ki, bir olayın gerçekleşme olasılığı yoksa başka bir deyişle kesinlikle gerçekleşmeyeceği biliniyorsa, olasılık sıfırdır. Öte yandan, bir olayın gerçekleşmesi kesin ise olasılığı 1'dir. Bu iki uç durum, içinde kesinlik barındıran iki durumdur ve olasılık bu iki kesinlik hali arasında değer alabilir. Olasılığın bire yaklaşması gerçekleşme olasılığının yüksek olduğunu, sıfıra yaklaşması ise gerçekleşmeme olasılığının yüksek olduğunu gösterir.

Sonuç olarak olasılık;

$$0 \leq \frac{m}{n} \leq 1$$

$$0 \leq P \leq 1$$

sınırları arasında değer alabilecektir.

Olasılık değerinin kesinlik ifade ettiği durumlara yönelik birkaç örnek verelim:

- Bir zarın atılması deneyinde, üste gelen yüzün 1, 2, 3, 4, 5, 6'dan biri olma olasılığı 1'dir.
- Bir zarın atılması olayında üste gelen yüzün 7 olma olasılığı sıfırdır.
- Güneşin doğudan doğma olasılığı birdir.
- İnsanın ölümsüz olma olasılığı sıfırdır.

Örneklerden de görülebileceği gibi, olasılık kesin olarak gerçekleşme ve kesin olarak gerçekleşmeme hali şeklinde ifade edebileceğimiz iki uç durum arasında ortaya çıkabilecek bir sonuçtur.

10.3. Basit ve Birleşik Olaylar

10.3.1. Basit Olaylar ve Basit Olaylarda Olasılık Hesabı

Tesadüfi bir deneyde, ortaya çıkabilecek sonuçlardan her birine **basit olay** denir. Söz gelimi, para atılması olayında yazı sonucu elde etmek, zar atılması örneğinde 6 sayısını elde etmek birer basit olaydır.

Basit olaylarda olasılık hesabı, yukarıda da açıkladığımız olasılık formülü ile yani elverişli sonuç sayısının toplam sonuç sayısına oranlanmasıyla bulunur.

Zar atılması olayında üste gelen yüzün çift sayı olması olasılığı, çift sayı durumu 2, 4, 6 gelmesi halinde geçerli olacağı için elverişli sonuç sayısı 3 ve mümkün sonuç sayısı 6 adet olduğundan,

$$P = \frac{m}{n} = \frac{3}{6}$$

hesaplanacaktır.

Benzer şekilde, üste gelen yüzün ikiden büyük tek sayı olması durumu ile ilgileniyorsak, bu durum 3 ve 5 sonuçları için geçerli olacağından, olasılık,

$$P = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}$$

biçiminde hesaplanacaktır.

İki paranın aynı anda havaya atılması durumunda, iki yüzün de aynı gelmesi olasılığı,

$$P = \frac{m}{n} = \frac{2}{4} = 0,50$$

olmaktadır.

Verdiğimiz örneklerden her biri basit bir olay ve bu olayların olasılık hesabının nasıl yapıldığına yöneliktir.

10.3.2. Birleşik Olaylar ve Birleşik Olaylarda Olasılık Hesabı

Birden fazla basit olayın gerçekleşmesi söz konusu olduğunda **birleşik olaylardan** söz edilmektedir. Örneğin 52 karttan oluşan bir iskambil destesinde 13 adet kupa, 13 adet maça, 13 adet sinek ve 13 adet karo kart bulunur. Şimdi bu desteden tesadüfen bir kart çektiğimizi varsayalım ve çekilen kartın kupa ya da karo olması olasılığını araştıralım.

Burada, çekilen kartın kupa olması da, çekilen kartın karo olması da basit birer olaydır. Dolayısıyla verilen örnekte, iki basit olaydan oluşan bir birleşik olay tanımlanmıştır.

Çekilen kartın kupa ya da karo olması elverişli durumları oluşturur. Başka bir deyişle çekilen kartın kupa ya da karo olması istenen sonuç anlamında bir farklılık yaratmamaktadır.

A olayını çekilen kartın kupa gelmesi şeklinde tanımlarsak, A olayının olasılığı,

$$P(A) = \frac{m}{n} = \frac{13}{52}$$

B olayını da çekilen kartın karo gelmesi olarak tanımlarsak, B olayının olasılığı,

$$P(B) = \frac{m}{n} = \frac{13}{52}$$

olacaktır.

O halde, çekilen kartın kupa ya da karo olması olasılığını nasıl hesaplanacaktır?

Çekilen kartın kupa ya da karo olması elverişli sonuçlar kümesini oluşturur. Bir anlamda çekilen kartın kupa ya da karo olması istenen durum açısından bir farklılık yaratmaz.

Burada belirleyici nokta, çekilen kartın aynı anda hem kupa hem karo olmasının mümkün olup olmadığıdır.

İskambil destesi içinde aynı anda kupa ve karo olma özelliği taşıyan kart bulunmamaktadır. Başka bir deyişle, kupa ve karo kartların kesişim kümesi boş kümedir.

Bu durumda, birleşik olayların olasılık hesabı, basit olayların olasılıkları toplanarak yapılır.

Çekilen kartın kupa ya da karo olması, sırasıyla A ve B olayları şeklinde tanımlanan bu iki olayın fark gözetmeksizin gerçekleşmesi durumu olduğuna göre, olasılık,

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = \frac{13}{52} + \frac{13}{52} = \frac{26}{52} = 0,50$$

olarak hesaplanmaktadır.

Burada, çekilen kartın aynı anda kupa ve karo olma olasılığının bulunmadığını bir kez daha hatırlatmak isteriz.

Bu konuya ilişkin bir örnek çözüm daha yapalım.

Örnek:

Bir zarın atılması deneyinde sonucun 1 veya 6 gelme olasılığı nedir?

Zarın atılması deneyinde,

Üste gelen yüzün 1 olması durumunu A olayı, 6 olması durumunu ise B olayı şeklinde tanımlayalım. Bu durumda A ve B olaylarının olasılıkları;

$$P(A) = \frac{1}{6}$$

$$P(B) = \frac{1}{6}$$

ve üste gelen yüzün 1 ya da 6 olması arasında fark gözetilmediği durumda A veya B olaylarının gerçekleşme olasılığı da,

$$P(A \cup B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

hesaplanmaktadır.

Zar atılması örneğinde üste gelen yüzün 1 ya da 6 olmasının aynı anda gerçekleşemeyecek olaylar olduğunu hatırlatalım.

Yukarıda verdiğimiz iki örnekte de A ve B olaylarının aynı anda gerçekleşmesi mümkün değildir. Zira, A ve B olaylarının aynı anda gerçekleşmesini sağlayacak ortak elemanları bulunmamaktadır. Bu durum olasılıkta **bağdaşmayan** olaylar olarak adlandırılır.

10.3.2.1. Bağdaşmayan Olaylar

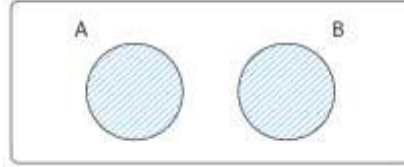
A ve B olayı olarak tanımlanan iki basit olayın aynı anda meydana gelmesi mümkün değilse, bu tür olaylara **bağdaşmayan olaylar** denir. İki olayın bağdaşmayan olaylar olması durumunda, olaylardan birinin meydana gelmesi diğerinin meydana gelmesini olanaksız kılar.

Yukarıda örneklerini verdiğimiz 52 karttan oluşan bir iskambil destesinde çekilen kartın aynı anda kupa ve karo olması mümkün değildir. Benzer şekilde zar atılması örneğinde de üste gelen yüz aynı anda 1 ve 6 olamaz. Dolayısıyla bu tür olaylar bağdaşmayan olaylara örnektir.

Burada temel nokta, A ve B olaylarının kesişim kümesinin boş küme olması başka bir deyişle ortak elemanlarının bulunmamasıdır. Bağdaşmayan olaylar söz konusu olduğunda, A veya B olayının gerçekleşme olasılığı, A ve B olaylarının olasılıklarının toplamı alınarak hesaplanır.

A ve B olaylarının bağımsız olaylar olması durumunu kümeler yaklaşımı ile gösterelim:

Aşağıda verilen Şekil 10.1’de görüleceği gibi, A ve B olaylarının ortak elemanlarının bulunmadığı ve dolayısıyla aynı anda veya birlikte gerçekleşmelerinin de söz konusu olmadığı anlaşılmaktadır.



Şekil 10.1: Bağımsız Olaylar

Bu durum A ve B olaylarının bağımsız olaylar olduğunu ve aynı anda gerçekleşemeyeceklerini ifade eder. Böyle bir durumda, A veya B olayının gerçekleşme olasılığı, A ve B olaylarının olasılıkları toplamı alınarak hesaplanır.

Yani bağımsız olaylarda olasılık hesabı,

$$P(A \cup B) = P(A) + P(B)$$

formülüyle yapılır.

Örnek:

52 karttan oluşan bir iskambil destesinden tesadüfen çekilen bir kartın, as ya da papaz olma olasılığı nedir?

İskambil destesinde dört tane as, dört tane papaz kart bulunmaktadır ve aynı anda hem as hem papaz olan kart bulunmamaktadır.

Bu durumda, çekilen kartın as olması olayını A, papaz olması olayını B olayı şeklinde tanımlayalım.

İskambil destesinde 4 tane as kart olduğuna göre ve destede toplam 52 kart bulunduğuna göre, çekilen kartın as olma olasılığı,

$$P(A) = \frac{4}{52} = \frac{1}{13}$$

Benzer şekilde, 4 papaz kart olduğuna göre ve yine destede toplam 52 kart bulunduğuna göre, çekilen kartın papaz olma olasılığı,

$$P(B) = \frac{4}{52} = \frac{1}{13}$$

ve çekilen kartın as ya da papaz olma olasılığı da, olasılıklar toplamından,

$$P(A \cup B) = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$$

hesaplanmaktadır.

Bağdaşmayan Olaylar Durumunda Ortaya Çıkan Bazı Önemli Sonuçlar:

Bağdaşmayan olaylar söz konusu olduğunda ortaya çıkan ve olasılık hesaplarında önem taşıyan bazı özellikler üzerinde durmak istiyoruz.

Bağdaşmayan olaylar söz konusu olduğunda, bir deney sonucunda ortaya çıkabilecek bütün mümkün sonuçların olasılıkları toplamı 1 olup bu tam sistem değerini vermektedir.

Örneğin, madeni bir para atıldığında, yazı ve tura şeklinde iki sonuç gelmesi mümkündür ve yazı ve tura sonuçlarının kesişim kümesi bulunmamaktadır. Dolayısıyla, paranın yazı ya da tura gelmesi olayları bağdaşmayan olaylardır.

Bu deneyde, yazı gelme olasılığı $\frac{1}{2}$ ve tura gelme olasılığı da yine $\frac{1}{2}$ 'dir. Yazı ve tura olasılıklarını topladığımızda 1 sonucunu elde ederiz ki bu durum, para atılması deneyinin tüm mümkün sonuçlarının göz önüne alındığını ifade eder.

Benzer şekilde, zar atılması deneyinde de mümkün sonuçlar 1, 2, 3, 4, 5, 6 şeklindedir ve sonuçların kesişim kümeleri bulunmamaktadır. Bütün mümkün sonuçların olasılıkları $\frac{1}{6}$ olup, altı mümkün sonuç için olasılıklar toplamı alındığında 1 değerine ulaşılmaktadır.

Bir tesadüfi deney sonucunda ortaya çıkabilecek mümkün sonuçları $x_1, x_2, x_3, \dots, x_n$ biçiminde gösterelim. Bu sonuçların olasılıklarını da $P(x_1), P(x_2), P(x_3), \dots, P(x_n)$ ile gösterirsek toplam olasılık,

$$\sum_{i=1}^n P(x_i) = 1$$

olacaktır.

Bağdaşmayan olaylar söz konusu olduğunda ortaya çıkan ikinci önemli özellik de, bir tesadüfi deney sonucunda mümkün sonuçlar içinde yer alan A olayının gerçekleşme olasılığı ile gerçekleşmeme olasılıkları toplamının “bir” olmasıdır.

Madeni bir paranın atılması deneyinde sonucun yazı gelmesi olasılığı $\frac{1}{2}$ 'dir, yazı gelmeme olasılığı da (ki bu durum tura gelmesi durumunu ifade eder) $\frac{1}{2}$ 'dir. Dolayısıyla, paranın yazı gelmesi ve yazı gelmemesi olasılıklarının toplamı 1'i vermektedir.

Zar atılması deneyi için de benzer durum söz konusudur. Zar atılması olayının sonuçları da bağdaşmayan olaylardır ve kesişim kümeleri bulunmamaktadır. Bu durumda, zarın altı gelmesi olasılığı $\frac{1}{6}$, altı gelmemesi olasılığı da $\frac{5}{6}$ olmaktadır. Zarın altı gelmemesi, altı dışında yer alan diğer 5 sonuçtan herhangi birinin gelmesi anlamına geldiğinden, zarın altı gelmemesi olasılığı bu beş sonucun olasılıkları toplanarak bulunmakta ve bu değer de $\frac{5}{6}$ olmaktadır.

Zarın altı gelmesi olayı A olayı şeklinde tanımlanırsa, zarın 6 gelmemesi olayı da A' (A-üssü) olarak tanımlanır. Bu durumda,

$$P(A) + P(A')=1$$

$$P(A) = 1 - P(A')$$

$$P(A') = 1 - P(A)$$

eşitlikleri geçerli olmaktadır.

10.3.2.2. Bağdaşan Olaylar

A ve B şeklinde tanımlanan iki olay birlikte meydana gelebiliyorsa bu tür olaylara *bağdaşan olaylar* denir.

Bağdaşan olaylar, A ve B gibi iki basit olayın birlikte meydana gelmesi halinde söz konusu olmaktadır. Dolayısıyla, A ve B olaylarının aynı anda meydana gelebilmesi, bu iki olayın ortak elemanı bulunmasına bağlı olmaktadır. Bağdaşan olaylar söz konusu olduğunda A olayının meydana gelmesi B olayının meydana gelmesini engellemez.

Özetle, A ve B olayları olarak tanımlanan iki basit olay,

- aynı anda meydana gelebiliyorsa ya da,
- birinin meydana gelmesi diğerinin meydana gelmesini engellemiyorsa

A ve B olaylarına **bağdaşan olaylar** denir.

Örneğin, 52 karttan oluşan bir iskambil destesinden tesadüfen bir kart çekildiğini ve çekilen kartın kupa ya da as olma olasılığını hesaplamak istediğimizi düşünelim.

İskambil destesinde, 52 kart bulunmakta ve bu kartların 13 tanesi kupa, 4 tanesi as ve bir tanesi de kupa ası olma özelliği taşımaktadır. Dolayısıyla önceki örneklerden farklı olarak hem kupa hem as olma özelliği taşıyan ortak bir kart söz konusudur.

A olayını çekilen kartın kupa olması şeklinde tanımlarsak, A olayının olasılığını,

$$P(A) = \frac{13}{52}$$

B olayını da çekilen kartın as olması şeklinde tanımlarsak, B olayının olasılığını,

$$P(A) = \frac{4}{52}$$

olarak hesaplarız.

Burada dikkat edilmesi gereken önemli nokta, kupa ası olarak tanımlanan kartın, hem kupa hem as çekme olayları içinde hesaplandığıdır. Bu durumda, çekilen kartın kupa ya da as olması olasılığını hesaplarken adı geçen kartı iki kere işleme dâhil etmiş oluyoruz. Dolayısıyla,

kupa ve as özellik gösteren kartların kesişim kümesinde yer alan kupa ası kartının, birleşik olasılık hesabında iki kere sayılma işlemini ortadan kaldırmamız gerekmektedir. Bunun için, çekilen kartın kupa ya da as olma olasılığı,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$$

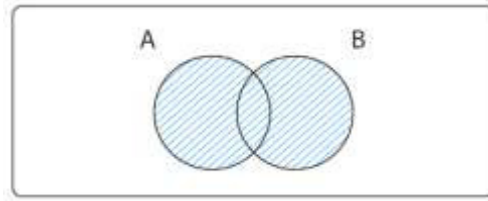
şeklinde hesaplanmaktadır.

Yukarıdaki örnekten de kolayca anlaşılacağı gibi, bağdaşan olaylar söz konusu olduğunda, A ve B gibi adlandırılan iki basit olayın birleşik olasılığı,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

formülüyle hesaplanır.

Bu durumu küme gösterimiyle de izah etmeye çalışalım. Aşağıda verilen Şekil 5.2’de bağdaşan iki olayın küme görüntüsü yer almaktadır:



Şekil 10.2. Bağdaşan Olaylar

Şekil 10.2’den de görüleceği gibi A ve B olaylarının kesişim kümesi vardır ve en az bir ortak elemanları mevcuttur. Çekilen kartın kupa ya da as olması örneğine dönersek, 52 tane kart içinde 13 tane kupa, 4 tane as kart bulunmaktadır. Dolayısıyla kupa ya da as olma özelliği taşıyan $13+4=17$ kart vardır dersek, hem kupa hem as özelliği taşıyan kupa asını kupa ve as olarak ayrı ayrı düşünerek iki kez saymış oluruz. İşte bu sebeple, kupa kartlar içinde değerlendirilen kupa asını, aslar içinde de saydığımız için bir çıkarma işlemiyle bu çift sayıma engel olmamız gerekecektir.

Dolayısıyla, bir birleşik olayın olasılığını hesaplarken dikkat edilmesi gereken nokta, iki olayın ortak elemanının olup olmadığıdır. Ortak elemanın olmaması halinde basit olayların olasılıkları toplamıyla birleşik olasılık hesaplanırken; ortak eleman olması halinde ise, ortak eleman ya da elemanların olasılıklarının çıkartılması gerektiği hatırlanmalıdır.

Örnek:

Bir zar atılıyor. Üste gelen yüzün çift sayı olması A olayı, üçten büyük bir sayı gelmesi ise B olayı olarak tanımlanıyor. A veya B olayının gerçekleşme olasılığını hesaplayınız.

Öncelikle, A ve B olaylarının elverişli sonuçlarına ve kesişim kümesinin olup olmadığına bakalım.

$$A = \{2,4,6\} \text{ ve } P(A) = \frac{3}{6}$$

$$B = \{4,5,6\} \text{ ve } P(B) = \frac{3}{6}$$

$$A \cap B = \{4,6\} \text{ ve } P(A \cap B) = \frac{2}{6}$$

Görüldüğü gibi A ve B olaylarının kesişim kümesi boş küme değildir, ortak elemanları vardır dolayısıyla bağıdaşan olaylar söz konusudur.

A ve B olaylarının bağıdaşan olaylar olması durumunda olasılık hesabı,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

formülüyle yapılmaktadır ve birleşik olasılık,

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}$$

hesaplanmaktadır.

10.4. Koşullu Olasılık

A ve B birbiri ile bağıdaşan ve gerçekleşme olasılığı bulunan iki basit olay olsun. A ve B olaylarının kesişim kümesinin boş küme olmaması yani ortak elemanı bulunması halinde, A olayının kesinlikle meydana gelmesi durumunda B olayının da meydana gelmiş olma olasılığına **koşullu olasılık** denir.

Koşullu olasılıktan söz edebilmek için A ve B olaylarının şu koşulları sağlamaları gerekir: A ve B olayları gerçekleşmesi mümkün olaylar olacak ve kesişim kümelerinde ortak eleman bulunacak.

Başka bir deyişle, koşullu olasılıktan söz edebilmek için,

$$P(A) \neq 0 \text{ ve } P(B) \neq 0 \text{ ve } P(A \cap B) \neq \emptyset$$

koşullarının gerçekleşmiş olması gerekecektir.

Koşullu olasılık A olayının gerçekleşmesi durumunda B olayının da gerçekleşmiş olma olasılığı olarak söylenebilir. Burada A olayı önce gerçekleşmiş olay olması bakımından koşul niteliğindedir ve B olayı için evrensel küme A olayının elemanlarına indirgenmiştir.

A olayının koşul olması durumunda B olayının gerçekleşme olasılığı: $P(B/A)$,

B olayının koşul olması durumunda A olayının gerçekleşme olasılığı ise: $P(A/B)$

şeklinde gösterilmektedir. (İkinci yazılan olayın, koşul durumundaki olay olduğuna dikkat ediniz).

A olayının gerçekleşmesi halinde B olayının da gerçekleşmiş olma olasılığı,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

B olayının gerçekleşmiş olması durumunda A olayının gerçekleşmiş olma olasılığı,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

formülleri yardımıyla hesaplanır.

Formülleri inceleyecek olursak; $P(B/A)$, A olayının koşul olması durumunda B olayı olasılığını göstermektedir ve A ve B olaylarının kesişim kümesinin olasılığı, koşul olan A olayının olasılığına oranlanmaktadır.

Benzer şekilde $P(A/B)$, B olayının koşul olması durumunda A olayının gerçekleşme olasılığıdır ve A ve B olaylarının kesişim kümesinin olasılığı da koşul olan B olayının olasılığına oranlanmaktadır.

Dolayısıyla koşullu olasılık hesabında, A ve B olaylarının kesişim kümesinin olasılığı koşul olan hangi olay ise o olayın olasılığına oranlanmaktadır.

Koşullu olasılık, bir olayı bir başka olaya bağlamak başka bir deyişle, bir olay kesin gerçekleşmişken diğer olayın da gerçekleşme olasılığını araştırmak ve dolayısıyla, evrensel kümeyi koşul olan olaya indirgemek anlamına gelmektedir.

Koşullu olasılığı bir örnek yardımıyla açıklamaya çalışalım:

Örnek:

Bir zarın atılması deneyinde, A olayı zarın üç veya üçten küçük sayı gelmesi, B olayı ise, zarın altıdan küçük çift sayı gelmesi olarak tanımlanıyor. Buna göre, A koşulken B olayının olasılığını, $[P(B/A)]$ ve B olayı koşulken A olayının olasılığını, $[P(A/B)]$ hesaplayınız.

Öncelikle A ve B olaylarının elverişli sonuçlarını göstererek olasılıklarını hesaplayalım:

$$A = \{1,2,3\} \quad ve \quad P(A) = \frac{3}{6}$$

$$B = \{2,4\} \quad ve \quad P(B) = \frac{2}{6}$$

$$A \cap B = \{2\} \quad ve \quad P(A \cap B) = \frac{1}{6}$$

A olayı koşulken B olayının olasılığı,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/6}{3/6} = \frac{1}{3}$$

B olayı koşulken A olayının olasılığı da,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{2/6} = \frac{1}{2}$$

hesaplanmaktadır.

Koşullu olasılıkla ilgili bir örnek çözüm daha yapalım:

Örnek:

52 karttan oluşan bir iskambil destesinden tesadüfen bir kart çekiliyor. Çekilen kartın kupa olması halinde as olması (kupa ası) olasılığını hesaplayınız.

Çekilen kartın kupa olması olayını A, as olması olayını da B olayı olarak tanımlayalım. Bu durumda bizden $P(B/A)$ olasılığını hesaplamamız istenmektedir.

52 iskambil kartı içinde 13 adet kupa, 4 adet as ve aynı anda hem kupa hem as olma özelliği taşıyan 1 adet kupa ası bulunmaktadır. Bu durumda öncelikle; $P(A)$, $P(B)$ ve $P(A \cap B)$ olasılıklarını hesaplayalım:

$$P(A) = \frac{13}{52}$$

$$P(B) = \frac{4}{52}$$

$$P(A \cap B) = \frac{1}{52}$$

Hesapladığımız olasılıklar üzerinden A olayı koşulken B olayının olasılığını yani, çekilen kartın kupa olması durumunda as olması olasılığını,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/52}{13/52} = \frac{1}{13}$$

şeklinde hesaplıyoruz.

Soruda istenmemekle beraber çekilen kartın as olması halinde maça olması olasılığını da hesaplayalım. Bu durumda, $P(A/B)$ olasılığını hesaplamamız gerekecektir ve bu olasılık da,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/52}{4/52} = \frac{1}{4}$$

olmaktadır.

10.5. Bağımsızlık Kavramı ve Bağımsız Olaylar

A ve B gerçekleşmesi mümkün iki olay olsun. Bu durumda,

$$P(A \cap B) = P(A) \cdot P(B)$$

eşitliği sağlanıyorsa A ve B olayları **bağımsız olaylar**dır denir. Başka bir deyişle A olayının gerçekleşme olasılığı B olayının gerçekleşme olasılığından; B olayının gerçekleşme olasılığı da A olayının gerçekleşme olasılığından tamamen bağımsızdır.

Para atılması ya da zar atılması olayları bağımsızlık kavramının anlaşılması için en iyi örneklerdir.

Bir parayı iki kez attığımızı düşünelim. İlk atışta tura gelmesi, ikinci atışta tura gelmesini engellemez. Çünkü ilk atış sonucunda tura gelmesi, paranın yüzünde yer alan yazı tura seçeneklerini azaltmaz ya da değiştirmez. Dolayısıyla, ikinci atışta yine yazı ve tura seçenekleri mevcuttur ve olasılıkları da 1/2 olup değişmemiştir.

Bu konuyu örnek yardımıyla açıklamaya çalışalım:

Örnek:

Bir para ardarda iki kez atılıyor. A olayı ilk paranın yazı olması, B olayı da iki paranın aynı yüzünün gelmesi olarak tanımlanıyor. A ve B olayları bağımsız olaylar mıdır?

İki paranın ardarda atılması durumunda ortaya çıkabilecek mümkün sonuçlar kümesi ya da evrensel küme,

$$E = \{YazıTura, YazıYazı, TuraYazı, TuraTura\}$$

şeklinde ve görüldüğü gibi toplam 4 mümkün sonuç bulunmaktadır.

A ve B olaylarına ilişkin sonuçlar ve olasılıklar ile $A \cap B$ sonuçları ve olasılıkları aşağıdaki gibidir:

$$A = \{YazıTura, YazıYazı\} \quad ve \quad P(A) = \frac{2}{4}$$

$$B = \{YazıYazı, TuraTura\} \quad ve \quad P(B) = \frac{2}{4}$$

$$A \cap B = \{YazıYazı\} \quad ve \quad P(A \cap B) = \frac{1}{4}$$

Şimdi A ve B olaylarının bağımsız olaylar olup olmadığına bakalım. $P(A \cap B) = 1/4$ olduğuna göre, bağımsızlıktan söz edebilmek için aşağıdaki eşitliğin sağlanması gerekir:

$$P(A \cap B) = P(A) \cdot P(B)$$

Dolayısıyla,

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap B) = \frac{2}{4} \cdot \frac{2}{4} = \frac{4}{16} = \frac{1}{4}$$

$$\frac{1}{4} = \frac{1}{4}$$

olmaktadır.

Görüldüğü gibi eşitlik sağlanmıştır ve bu durumda A ve B olayları bağımsız olaylardır, A olayının gerçekleşme olasılığı B olayının gerçekleşme olasılığından tamamen bağımsızdır ve aynı şekilde B olayının gerçekleşme olasılığı da A olayının gerçekleşme olasılığından tamamen bağımsızdır.

Örnek:

Bir zar iki kez atılıyor. Her iki atışta da 6 gelme olasılığı nedir?

İlk zarın 6 gelmesini A, ikinci zarın 6 gelmesini de B olayı olarak tanımlayalım. İki zarın üste gelen yüzleri için $6 \times 6 = 36$ mümkün sonuç vardır ve bunlar aşağıdaki mümkün sonuçlar kümesinde gösterilmiştir:

$$E = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), \dots, (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

İlk atışta altı gelme olasılığı, 36 mümkün sonuç içinde 6 değişik şekilde meydana gelebilmektedir ve olasılığı:

$$A = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\} \quad \text{ve} \quad P(A) = \frac{6}{36} = \frac{1}{6}, \text{dır.}$$

İkinci atışta altı gelme olasılığı ise, yine 36 mümkün sonuç içinde 6 değişik şekilde ortaya çıkabilir ve olasılık değeri de,

$$B = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,6)\} \quad \text{ve} \quad P(B) = \frac{6}{36} = \frac{1}{6}$$

olmaktadır.

Her iki atışta da altı gelme durumu $6 \times 6 = 36$ mümkün sonuç içinde sadece bir tanedir ve o da (6,6)'dır.

Bu durumda, A ve B olaylarının kesişim kümesinde sadece (6,6) durumu vardır ve olasılığı da $1/36$ 'dır.

$$A \cap B = \{6,6\} \quad \text{ve} \quad P(A \cap B) = \frac{1}{36}$$

Şimdi A ve B olaylarının bağımsız olmaları halinde geçerli olacak,

$$P(A \cap B) = P(A) \cdot P(B)$$

eşitliğinin varlığını sınavalım:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$\frac{1}{36} = \frac{1}{36}$$

Görüldüğü gibi bağımsızlık için gerekli koşul sağlanmıştır.

Burada, ilk zarın üste gelen yüzünün ne olduğu, ikinci zarın sonuçlarını değiştirmemektedir. Örneğin, ilk atışta 6 geldiyse zarın üzerindeki yüz sayısı azalmaz yani bu zarın üzerindeki 6 değerinin ortadan kalkması anlamına gelmez. İkinci atışta da, ilk atışta olduğu gibi, zarın üzerinde yine birden altıya kadar altı adet sayı vardır, dolayısıyla mümkün sonuç sayısı ve bu sonuçların olasılıkları değişmemektedir ve bağımsızlık bu sebeple söz konusu olmaktadır.

10.6. Bağımlı Olaylar

Ardarda gerçekleşecek iki olaydan ilkinin alacağı sonuç, ikinci olayın sonuçlarını değiştiriyorsa bu tür olaylara **bağımlı olaylar** denir. Bağımlı olayları iadesiz çekim söz konusu olduğunda görüyoruz. Örneğin, bir kalemlik içinde 5 siyah ve 5 kırmızı kalem olduğunu, ardarda iki kalem çektiğimizi varsayalım ve her iki kalemin de kırmızı olması olasılığını araştıralım.

İlk kalemin kırmızı olması olasılığı, kalemlikte 10 kalem bulunması ve bunların 5 tanesinin kırmızı kalem olması nedeniyle 5/10 olacaktır. İlk çekilen kalemin kırmızı olması halinde ikinci çekiliş için artık kalemlikte 4 kırmızı kalem vardır ve toplam kalem sayısı da 9 tanedir. Bu durumda ikinci kalemin kırmızı olması olasılığı, 4/9 olarak gerçekleşir.

Görüldüğü gibi, ilk kalemin renginin ne olduğu ikinci çekiliş için önem taşımaktadır. Zira, ilk çekilen kalem kalemliğe iade edilmemekte ve kalemlik içindeki kalem bileşimi ile toplam kalem sayısı ve buna bağlı olarak da olasılıklar değişmektedir.

Bu tür olaylara, ikinci olayın olasılığının ilk olayın sonucuna bağlı olması sebebiyle, **bağımlı olaylar** denilmektedir.

Bağımlı olaylar söz konusu ise, yani olaylardan ilkinin meydana gelmesi ikincisinin meydana gelmesini etkiliyorsa, daha önce öğrendiğimiz,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

şeklindeki koşullu olasılık formülünden hareketle bağımlı olaylar için olasılık hesabı,

$$P(A \cap B) = P(A) \cdot P(B/A)$$

formülüyle yapılmaktadır.

Yukarıda bahsettiğimiz 5 siyah, 5 kırmızı kalem bulunan kalemlik örneğini tekrar ele alalım. Bu kalemlikten ardarda iki kalem çekildiğini ve her iki kalemin de siyah olması olasılığını araştıralım.

İlk çekilişte siyah kalem çekmeyi A olayı, ikinci çekilişte siyah kalem çekmeyi B olayı olarak tanımlayalım. Bu durumda, ilk çekilişte kalemlikte 10 kalem vardır ve bunların 5 tanesi siyah olduğuna göre A olayının olasılığı,

$$P(A) = \frac{5}{10}$$

hesaplanmaktadır.

İkinci çekiliş artık, ilk çekilen kalemin siyah olduğu halde, yapılmaktadır. Başka bir deyişle ilk kalem çekilişi, ikinci çekilişin koşulu niteliğindedir. Yani ilk çekilen kalem siyah olduğundan, bir siyah kalem dışardadır ve kalemlikten bir siyah kalem eksilmiştir. Dolayısıyla, artık kalemlikte 9 kalem kalmıştır ve bunların 4 tanesi siyahtır.

İşte böyle bir koşullu çekilişte ardarda çekilecek iki kalemin de siyah olması olasılığı,

$$P(A \cap B) = P(A) \cdot P(B/A)$$
$$P(A \cap B) = \frac{5}{10} \cdot \frac{4}{9} = \frac{20}{90} = \frac{2}{9}$$

olarak hesaplanmaktadır.

Aynı örneği, ardarda 3 kalem çekildiği ve üçünün de siyah olması şeklinde kurgulayalım. İlk kalemin siyah olması A, ikinci kalemin siyah olması B ve üçüncü kalemin siyah olması da C olayı olarak tanımlansın.

İlk çekilişte, toplam 10 kalem vardır ve 5 tanesi siyahtır. Çekilen kalem siyah olduğu varsayıldığından, ikinci çekiliş için, artık kalemlikte 4 siyah kalem ve toplamda da 9 kalem bulunmaktadır. İkinci çekilen kalemin de siyah olduğu varsayıldığından, üçüncü kalem çekilişi için kalemlikte 3 siyah kalem kalmıştır ve toplam kalem sayısı da 8 tanedir.

Bu durumda, çekilen 3 kalemin de siyah olması olasılığı;

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/AB)$$
$$P(A \cap B \cap C) = \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} = \frac{60}{720} = \frac{1}{12}$$

olarak hesaplanmaktadır.

İlk iki kalemin siyah, üçüncü kalemin kırmızı olması olasılığı da benzer şekilde,

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/AB)$$
$$P(A \cap B \cap C) = \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{100}{720} = \frac{5}{36}$$

hesaplanacaktır.

Çekilen ilk iki kalemin siyah olduğu varsayıldığından kalemlikte 8 kalem kalmıştır ve bunların 5 tanesi kırmızı 3 tanesi siyahtır. Üçüncü ve son kalem çekilişinde kırmızı kalem çekme olasılığı bu nedenle 5/8 olarak formülde yerini almıştır.

10.7. Permütasyon ve Kombinasyon

10.7.1. Permütasyon

n , sonsuz olmayan belirli bir sayıyı ifade etmek üzere n sayıda birimin, kaç değişik şekilde sıralanabileceğini gösteren dizilime **permütasyon** denir.

3 birimimiz olduğunu ve bu birimleri x, y, z şeklinde kodladığımızı düşünelim. Bu üç birim, $3!$ kadar değişik şekilde sıralanabilir.

$$3! = 3.2.1 = 6$$

Nitekim bu şekilde, $(xyz), (xzy), (yxz), (yzx), (zxy), (zyx)$ şeklinde 6 değişik diziliş gerçekleştirilebilir.

Birim sayısının çok olması halinde, tüm farklı dizilişleri oluşturmak zordur ve permütasyon yaklaşımı ile kaç değişik diziliş yaratılabileceği kolaylıkla hesaplanabilir.

Söz gelimi, 4 birim varsa, $4!$ kadar farklı diziliş gerçekleştirilebilir ki bu sayı $4! = 4.3.2.1 = 24$ olmaktadır.

Bazı durumlarda ise n adet birim söz konusu olduğunda k adet nesnenin bu n birime kaç değişik şekilde dağıtılabileceği araştırılmaktadır.

Bunun için de,

$$\frac{n!}{(n - k)!}$$

formülünden yararlanılır.

Örnek:

Bir meslek edindirme kursunda ahşap boyama eğitimi alan 10 kursiyer bulunmaktadır. Kurs yönetimi tepsî, fincan ve vazo şeklinde 3 farklı objenin boyama işini kursiyerlere vermek istemektedir. Buna göre, 3 farklı obje kursiyerler arasında kaç değişik şekilde dağıtılabilir?

Burada, kişilerin boyayacağı objenin değişimi grubun farklı şekilde değerlendirilmesini gerekecektir. Şöyle ki, Ayşe'nin fincan, Fatma'nın vazo, Zeynep'in tepsî boyadığı dağılım ile Ayşe'nin vazo, Fatma'nın fincan ve Zeynep'in tepsî boyadığı dağılım farklı dizilişlerdir.

Yukarıdaki permütasyon formülü yardımıyla boyama işleminin,

$$\frac{n!}{(n - k)!} = \frac{10!}{(10 - 3)!} = \frac{10.9.8.7.6.5.4.3.2.1}{7.6.5.4.3.2.1} = 720$$

değişik şekilde dağıtılabileceği anlaşılmaktadır.

10.7.2. Kombinasyon

n, sonsuz olmayan belirli bir sayıyı ifade etmek üzere, n adet birimin sıra farkı gözetilmeksizin (xy ile yx'in aynı sayılması halinde) kaç değişik şekilde dizilebileceğini ya da grup oluşturabileceğini belirleme işlemine **kombinasyon** denir.

Kombinasyon analizine insanların tokalaşması iyi bir örnektir. Zira, tokalaşma iki kişilik bir eylemdir ve X ile Y tokalaşması ile Y ile X tokalaşması aynı olarak değerlendirilir ya da aynı durumu ifade eder.

Kombinasyon analizi,

$$C_k^n = \frac{n!}{k! (n - k)!}$$

formülüyle hesaplanır.

Örnek:

Bir okulda 7 matematik öğretmeni görev yapmaktadır. Matematik eğitiminin daha iyi verilebilmesi için 3 kişilik bir komisyon kurulmasına karar verilmiştir. Bu durumda, 7 öğretmenden 3 kişilik bir komisyon kaç değişik şekilde oluşturulabilir?

Okulda komisyonda görev yapabilecek 7 öğretmen olduğuna göre n=7'dir ve 3 kişilik komisyon kurulması söz konusu olduğundan k=3'tür.

Bu durumda öğretmenler arasından,

$$C_3^7 = \frac{n!}{k! (n - k)!} = \frac{7!}{3! (7 - 3)!} = \frac{7.6.5.4.3.2.1}{(3.2.1) . (4.3.2.1)} = 35$$

değişik komisyon kurmak mümkündür.

10.8. Bağımlı Olaylar Olasılığı Hesabına Kombinasyon Yaklaşımı:

Ardarda gerçekleşecek iki olaydan ilkinin alacağı sonuç, ikinci olayın sonuçlarını değiştiriyorsa bu tür olaylara **bağımlı olaylar** dendiğini daha önce ifade etmiştik. İadesiz çekim söz konusu olduğunda çekilişin yapıldığı grubun içindeki bileşim, her çekim sonrası değişeceğinden, iadesiz çekim durumlarında olayların bağımlı olaylar olabileceğine özellikle dikkat etmek gerekir.

Bağımlı olaylar söz konusu olduğunda koşullu olasılık formülünden türettiğimiz,

$$P(A \cap B) = P(A) . P(B/A)$$

formülünün kullanıldığını biliyoruz. Nitekim, bir kalemlikte 5 siyah, 5 kırmızı kalem bulunması halinde, çekilecek iki kalemin de siyah olması olasılığını yukarıdaki formülü kullanarak hesaplamıştık ($P(A \cap B) = 20/90$).

Kombinasyon yaklaşımı, bağımlı olaylara ilişkin olasılık hesaplarında kullanılabilecek alternatif bir yöntem olarak da karşımıza çıkmaktadır.

Yukarıda klasik çözüm yöntemi ile olasılığını hesapladığımız kalem örneğini kombinasyon yaklaşımı ile de çözelim.

Kalemliğin içinde bulunan 5 siyah kaleminden 2 tanesinin çekilmesi söz konusudur. Aynı zamanda, kalemlik içinde bulunan toplam 10 kalem içinden de yine 2 tanesi seçilecektir.

Bu durumda, beşin ikili kombinasyonu kadar değişik siyah kalem çekmek mümkün olacaktır. Yine aynı şekilde onun ikili kombinasyonu kadar değişik bileşimde kalem çekmek söz konusudur.

İlk çekilen kalemin siyah olması olayını S_1 , ikinci çekilen kalemin siyah olması olayını S_2 ile gösterelim. Bu durumda, çekilen iki kalemin de siyah olması olasılığı,

$$P(S_1 S_2) = \frac{C_2^5}{C_2^{10}} = \frac{\frac{5!}{2!(5-2)!}}{\frac{10!}{2!(10-2)!}} = \frac{10}{45}$$

olarak hesaplanmaktadır ki bu değer, klasik olasılık yaklaşımı ile 20/90 olarak hesapladığımız olasılık değerinin sadeleştirilmiş değeridir.

Örnek:

Bir kutuda, 8 kırmızı, 3 sarı ve 9 mavi misket bulunmaktadır. Bu kutudan tesadüfen üç misket seçilecektir. Çekilen ilk iki misketin kırmızı, üçüncü misketin sarı olması olasılığını hesaplayınız.

Çekim iadesiz yapılacağına göre, her çekiliş kutudaki misketlerin sayısını ve bileşimini değiştirecektir.

Kutudan tesadüfen 3 misket seçilecek olduğuna göre, kutuda bulunan 8 kırmızı misketten 2 tanesinin, 3 sarı misketten de 1 tanesinin çekilme olasılığını yani $P(K_1 K_2 S_3)$ 'ü kombinasyon yaklaşımı ile hesaplayalım:

$$P(K_1 K_2 S_3) = \frac{C_2^8 C_1^3}{C_3^{20}} = \frac{\frac{8!}{2!(8-2)!} \cdot \frac{3!}{1!(3-1)!}}{\frac{20!}{3!(20-3)!}} = \frac{84}{1140}$$

Çekilen ilk iki misketin kırmızı ve üçüncüsünün de sarı olması olasılığı 84/1140 olarak hesaplanmıştır.

Bu Bölümde Neler Öğrendik?

Olasılık kavramını ele aldığımız bu bölümde öncelikle olasılığın objektif ve sübjektif tanımlarını inceledik. Sübjektif olasılığın kişisel değerlendirme ve beklentilere açık ve günlük hayata yönelik bir kavram olduğunu ve bu şekliyle istatistik konusu içinde değerlendirilmediğini belirttik.

İstatistiğin konusunu, objektif olasılık ya da frekans olasılığı denilen olasılık tanımının oluşturduğunu ve bu anlamda olasılığın, elverişli sonuç sayısının toplam sonuç sayısına bölünerek hesaplandığını öğrendik.

Olasılık konusuna yönelik temel kavramlar üzerinde durarak, basit ve birleşik olaylar, bağıdaşan ve bağıdaşmayan olaylar, koşullu olasılık, bağımlı ve bağımsız olaylar, permütasyon ve kombinasyon analizlerini inceledik.

Bölüm Soruları

- 1- Kişilerin beklenti ve isteklerine göre belirledikleri, insandan insana değişen olasılığa denir.
 - a- Olasılık
 - b- Objektif olasılık
 - c- Sübjektif olasılık
 - d- Koşullu olasılık
 - e- Olasılık dağılımı
- 2- Hilesiz bir zar atıldığında üste gelen yüzün çift sayı olması olasılığı nedir?
 - a- $1/6$
 - b- $2/6$
 - c- $1/4$
 - d- $1/2$
 - e- $3/4$
- 3- İki para birlikte atılıyor. En az bir paranın yazı gelmesi olasılığı nedir?
 - a- $3/4$
 - b- $1/4$
 - c- $2/4$
 - d- $1/8$
 - e- 1,00
- 4- Hilesiz bir zarın atılması olayında üste gelen yüzün 6 olmama olasılığı nedir?
 - a- $1/6$
 - b- $5/6$
 - c- $4/6$
 - d- $3/6$
 - e- $2/6$

- 5- Hilesiz bir zar atılıyor, ste gelen yzn tek sayı olması veya 4'den byk sayı gelmesi olasılıęı nedir?
- a- $4/6$
 - b- $3/6$
 - c- $5/6$
 - d- $1/6$
 - e- $2/6$
- 6- 52 karttan oluřan iskambil destesinden tesadfen ekilen bir kartın kupa olması halinde as olması olasılıęı nedir?
- a- $3/13$
 - b- $2/13$
 - c- $4/13$
 - d- $5/13$
 - e- $1/13$
- 7- Bir odada bulunan 4 kiři 4 sandalyeye ka deęiřik řekilde oturtulabilir?
- a- 12
 - b- 24
 - c- 16
 - d- 18
 - e- 14
- 8- Bir serviste grev yapmakta olan 5 hemřireden iki tanesi gece nbetine kalacaktır. Sz konusu iki hemřire ka deęiřik řekilde seilebilir?
- a- 120
 - b- 60
 - c- 10
 - d- 18
 - e- 16

9- 10 kişinin katılacağı bir çekilişte zarfların 5 tanesinde sinema 5 tanesinde de tiyatro bileti bulunmaktadır. İki arkadaş birlikte birer zarf çektiğinde her ikisinde de sinema bileti bulunması olasılığı nedir?

a- 20/90

b- 10/90

c- 30/90

d- 5/10

e- 5/20

10- 5 kişilik bir yönetim kurulundan 3 kişilik bir komisyon kaç değişik şekilde oluşturulabilir?

a- 3

b- 5

c- 4

d- 8

e- 10

Cevaplar

1- c

2- d

3- a

4- b

5- a

6- e

7- b

8- c

9- a

10- e

Yararlanılan Kaynaklar

ALOBA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

11. OLASILIK - TESADÜFİ DEĞİŞKEN, OLASILIK DAĞILIMI VE BEKLENEN DEĞER

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde tesadüfi değişken, olasılık dağılımı ve beklenen değer kavramlarına yönelik bilgi sahibi olacağız.

Bu bağlamda öncelikle bir tesadüfi deney sonucunda ortaya çıkabilecek tüm mümkün sonuçların öngörülerek tesadüfi değişkenin nasıl tanımlandığını öğrenecek ve bununla ilgili örnekler göreceğiz.

Tesadüfi bir değişkenin alabileceği tüm mümkün sonuçlara, olasılıkların nasıl dağıldığını göstererek olasılık dağılımı yaratma sürecini inceleyeceğiz.

Ayrıca, bir tesadüfi değişkenin alabileceği tüm mümkün sonuçlarının olasılıklarına göre ağırlıklı ortalamasını alarak beklenen değer hesabını ve nerede kullanılmakta olduğunu ele alacağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Geçmişini biliyoruz, geleceğimizi ise olasılıkları değerlendirerek planlıyoruz. O halde nedir bu olasılık?
- 2) Olasılıkları doğru değerlendirmek belirsizliği azaltır mı?
- 3) İhtimalleri doğru değerlendirmek başarı olasılığını artırır mı?
- 4) Olasılığın yüksek olması sürpriz sonuçlar doğurabilir mi?
- 5) Olasılığın artması ya da azalması karar süreçlerinde neden etkilidir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Olasılığın objektif ve sübjektif tanımı, bir limit durum olarak olasılık.	Tesadüfi değişken kavramını ve türlerini bilir.	Temel düzeyde istatistik kitaplarının olasılık kısımları.
Beklenen değer kavramı.	Beklenen değer kavramını bilir.	Temel düzeyde istatistik kitaplarının olasılık kısımları.
Olasılık dağılımı.	Olasılık dağılımı yaratma sürecini bilir.	Temel düzeyde istatistik kitaplarının olasılık kısımları..
Olasılık dağılımı türleri.	Olasılık dağılımlarına yönelik ayrımları bilir.	Temel düzeyde istatistik kitaplarının olasılık kısımları.
Olasılık dağılımının koşulları ve olasılık dağılımının nasıl oluşturulduğu.	Olasılık dağılımından söz edebilmek için gerekli koşullar, olasılık dağılımı yaratma sürecini bilir.	Temel düzeyde istatistik kitaplarının olasılık kısımları.

Anahtar Kavramlar

- **Tesadüfi değişken:** Bir tesadüfi deney sonucunda, mümkün sonuçları öngörülebilir değişken.
- **Olasılık dağılımı:** Tesadüfi bir değişkenin alabileceği mümkün sonuçlara, olasılıkların nasıl dağıldığını gösteren ve istatistik analiz ve uygulamalarda kullanılan önemli bir araç.
- **Beklenen Değer:** Bir tesadüfi değişkenin alabileceği tüm mümkün sonuçlarının olasılıklarına göre ağırlıklı ortalaması.

Giriş

Önceki bölümde olasılık konusuna yönelik temel kavram ve çözümlemeler üzerinde durduk. Şimdi de tesadüfî değişken, olasılık dağılımı ve beklenen değer kavramlarını örneklerle açıklayacağız.

11.1. Tesadüfi (Rastlantısal) Değişken

Benzer koşullar altında yinelenildiğinde değişik sonuçlar alınabilen deneylere, **tesadüfi deneyler (rastlantısal deneyler)** denilmektedir.

Tesadüfi deneylerin mümkün sonuçları sayısal değer alabilen nicel sonuçlar olabileceği gibi, sayısal değer alamayan nitel sonuçlar da olabilir. Örneğin bir işyerinde istihdam edilen personel arasından tamamen tesadüfi olarak bir çekiliş yapılsa, personelin cinsiyeti ve eğitimi nitel, yaşı ise nicel özellik gösterecektir.

Benzer şekilde, bir tekstil atölyesinde dikilmekte olan gömleklerin son aşamada kontrol edilerek kusurlu-kusursuz şekilde ayrılması da nitel özellik gösterir.

Bir tesadüfi deneyin tüm mümkün sonuçları göz önüne alındığında, deney sonucunda hangi sonuçların ortaya çıkabileceği bilinmekte ancak, hangi sonucun çıktığı ancak deney sonucunda belirlenmektedir.

İşyeri örneğinde, tesadüfen seçilecek bir personelin cinsiyetinin kadın ya da erkek olabileceği şüphesiz bilinmektedir. Ancak, kadın mı erkek mi olduğu ancak seçim sonrasında anlaşılacaktır. İşyerinde çalışanların % 30'unun kadın % 70'inin erkek olduğunu varsayalım. Bu durumda, seçilecek personel % 30 olasılıkla kadın; % 70 olasılıkla erkek olacaktır.

Bir tesadüfi deney sonucunda ortaya çıkabilecek tüm mümkün sonuçlar, sayısal özellik göstermese bile sayısallaştırılabilirler. Örneğin, tekstil atölyesinde kontrolü yapılan gömlekler, kusurlu ise 1, kusursuz ise 0 ile kodlanabilir. Benzer şekilde, bir hastanede dünyaya gelen bebekler, düşük kilolu ise 1, normal kilolu ise 2 ve aşırı kilolu ise 3 ile kodlanarak sonuçlar sayısallaştırılabilir.

Tesadüfi deneylerin tüm mümkün sonuçlarına sayısal bir değer verilmesi durumunda bir **tesadüfi değişken** tanımlanmış olur.

Madeni bir paranın atılması mümkün sonuçları yazı ve tura olan tesadüfi bir deneydir. Yazı sonucunu 1, tura sonucunu 2 ile kodlarsak bir tesadüfi değişken tanımlamış oluruz.

Tesadüfi değişkenler, X, Y, Z gibi büyük harflerle ifade edilir. Tesadüfi değişkenin alabileceği sonuçlar da x, y, z gibi küçük harflerle gösterilir. Örneğin, $X=x_i$, X tesadüfi değişkeninin x_i sonucunu almasını ifade eder.

Bir zarın atılması da yine tesadüfi bir deneydir ve mümkün sonuçları da 1, 2, 3, 4, 5, 6'dır. Dolayısıyla zar atılması sonucunda ortaya çıkacak sonuçlardan bir tesadüfi değişken tanımlamak mümkündür.

Öte yandan X tesadüfi değişkeni, 1, 2, 3, 4, 5, 6 sonuçları alabilen bir tesadüfi değişken olarak tanımlanabileceği gibi; mümkün sonuçlar arasında yer alan sonuçlar, tek sayılar 0, çift sayılar 1 ile kodlanarak, iki sonuçlu bir tesadüfi değişken de tanımlanabilir.

11.2. Olasılık Dağılımı

Bir tesadüfi deneyin tüm mümkün sonuçlarına birer sayısal değer verilerek tesadüfi bir değişken tanımlanabileceğini öğrendik. Şimdi, bir tesadüfi değişkenin alabileceği tüm mümkün değerlerin olasılıkları ile birlikte ifade edildiğini düşünelim.

X tesadüfi değişkeninin alabileceği değerler,

$$x_1, x_2, x_3, \dots, x_n$$

olsun. X tesadüfi değişkeninin alabileceği $x_1, x_2, x_3, \dots, x_n$ değerlerinin olasılıkları da,

$$P(x_1), P(x_2), P(x_3), \dots, P(x_n)$$

şeklinde gösterilsin.

X tesadüfi değişkeninin x_i değerlerinden yalnızca birini alabildiğini varsayalım. Bu durumda, X tesadüfi değişkeninin x_i değerini alma olasılığı,

$$P(X=x_i) = P(x_i)$$

şeklinde gösterilecektir.

Daha önce ifade ettiğimiz gibi bağdaşmayan, yani kesişim kümesi bulunmayan olaylarda, mümkün tüm sonuçların olasılıkları toplamı tam sistem oluşturur yani toplam olasılık “1” olur:

$$\sum_{i=1}^n P(x_i) = 1$$

Bir tesadüfi deneyin tüm mümkün sonuçlarını olasılıkları ile ilişkilendirdiğimizde, tam sistem olasılığı olan 1 değerini, tüm mümkün sonuçlara olasılıklarına göre dağıtmış ve bir **olasılık dağılımı** yaratmış oluruz.

Zar atılması olayının tesadüfi bir deney olduğunu ifade etmiştik. Burada, tesadüfi değişkenin alabileceği değerler 1, 2, 3, 4, 5, 6 ve her bir sonucun olasılığı da 1/6’dır. Toplam olasılık da tüm mümkün sonuçların olasılıkları toplanarak 1 tam sistem değerini vermektedir. Dolayısıyla, zar atılması sonucunda da bir olasılık dağılımı tanımlanmaktadır.

Benzer şekilde, madeni bir para atılması deneyi de yazı ve tura şeklinde iki mümkün sonucu olan bir tesadüfi deneydir ve yazı ve tura gelmesi olasılığı eşit ve 1/2 olup toplam olasılık 1’dir ve tam sistem değeri vermektedir. Dolayısıyla, madeni para atılması deneyi de bir olasılık dağılımı oluşturur.

X tesadüfi değişkeninin alabileceği bütün mümkün değerleri olasılıkları ile birlikte gösterelim:

$X=x_i$	x_1	x_2	x_3		x_n
$P(X=x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$		$P(x_n)$

Bir olasılık dağılımından söz edebilmek için; dağılımın sağlaması gereken 3 koşul vardır. Bunlar:

- 1- $P(X=x_i) \geq 0$ veya $P(x_i) \geq 0$
- 2- $\sum_{i=1}^n P(X=x_i) = 1$ veya $\sum_{i=1}^n P(x_i) = 1$
- 3- $P(x_i \cup x_j) = P(x_i) + P(x_j)$ ve $i \neq j$

şeklinde sıralanabilir.

Bu koşullardan ilki, tesadüfi deney sonrasında ortaya çıkabilecek tüm mümkün sonuçlar için birer olasılık değeri söz konusu olduğunu ifade eder.

İkincisi, tüm mümkün sonuçların olasılık değerleri toplandığında tam sistem değeri olan 1'e ulaşıldığını gösterir. Böylelikle, hiçbir sonuç dışarda tutulmamış ve olasılık dağılımında yer almış olmaktadır.

Üçüncü koşulda tesadüfi değişkenin, x_i veya x_j sonucunu alma olasılığının, $i \neq j$ olmak koşulu ile iki olasılığın toplamı alınarak oluşturulduğunu göstermektedir. Bu koşul, tesadüfi değişkenin alacağı değerlerin bağdaşmayan sonuçlar olması gerektiğini ifade etmektedir. Hatırlanacağı gibi bağdaşmayan olaylar, kesişim kümesi yani ortak elemanı olmayan olaylardır ve bağdaşmayan olaylarda birleşik olasılık hesabı, basit olayların olasılıkları toplanarak yapılmaktadır.

Şimdi örnek bir olasılık dağılımı yaratalım:

Örnek:

İki zar aynı anda atılarak üstte gelen yüzler toplanmak suretiyle bir tesadüfi değişken tanımlanıyor. Ortaya çıkabilecek tüm mümkün sonuçları ve bu sonuçların olasılıklarını hesaplayarak tesadüfi değişkenin olasılık dağılımını oluşturunuz (GÜLER, 2005:144).

Tablo 11.1. İki zar birlikte atılarak üste gelen yüzlerin toplanması ile elde edilen sonuçlar

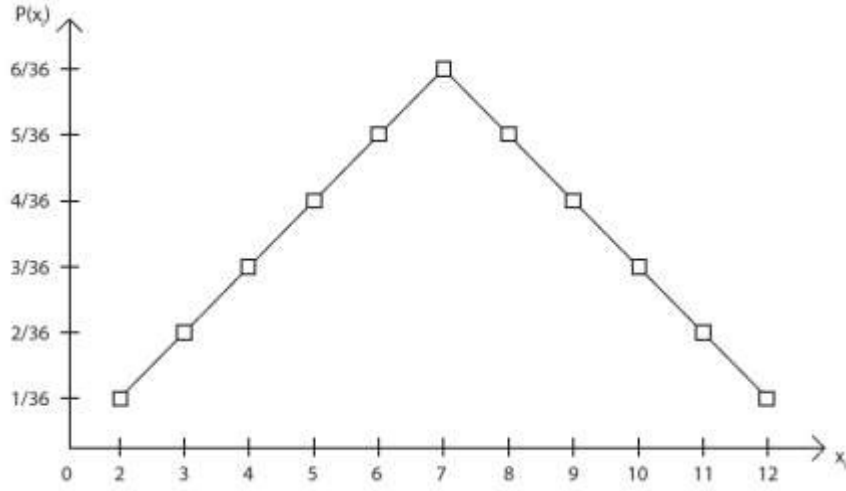
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Tablodan kolaylıkla görüleceği gibi, iki zarın birlikte atılarak üste gelen yüzlerinin toplanması deneyi ile yaratılan tesadüfi değişkenin alacağı değerler 2-12 arasında tam sayılardır.

Tesadüfi değişken için söz konusu olan bütün mümkün sonuçlar ile bu sonuçların hangi olasılıklarla gerçekleşeceğinin gösterildiği tablo aşağıdaki gibi düzenlenmiştir:

Tablo 11.2. İki zarın atılarak üste gelen yüzlerin toplanması şeklinde tanımlanan tesadüfi değişkenin olasılık dağılımı

x_i	Sonuç sayısı	$P(x_i)$	$\Sigma p(x_i)$
2	1	1/36	1/36
3	2	2/36	3/36
4	3	3/36	6/36
5	4	4/36	10/36
6	5	5/36	15/36
7	6	6/36	21/36
8	5	5/36	26/36
9	4	4/36	30/36
10	3	3/36	33/36
11	2	2/36	35/36
12	1	1/36	36/36
	36	1,00	



Grafik 11.1: İki zarın atılarak üste gelen sonuçların toplanması şeklinde tanımlanan tesadüfi değişkenin olasılık dağılımı

Yukarıda, iki zarın birlikte atılarak üste gelen yüzlerin toplanması ile tanımlanan tesadüfi değişkenin olasılık dağılımı görülüyor.

Tesadüfi değişkenin mümkün sonuçlarının olasılıklarına baktığımızda, tamamı sıfırdan farklı ve olası sonuçlardır. Dolayısıyla, olasılık dağılımı koşullarından ilki gerçekleşmiştir.

İkinci olarak, mümkün sonuçların olasılıklarını topladığımızda tam sistem değeri olan 1 değerine ulaşıyoruz. Dolayısıyla dağılıma ilişkin dışarda bırakılmış hiçbir sonuç bulunmamaktadır ve tüm mümkün sonuçlar olasılık dağılımı içinde yer almaktadır.

Üçüncü koşul olarak da mümkün sonuçların her biri bağdaşmayan olaylardır.

Tesadüfi değişkenin kesikli değişken olması durumunda, olasılıkların mümkün sonuçlara dağılımını gösteren dağılıma **kesikli olasılık dağılımı** ya da kısaca **olasılık dağılımı** denmektedir. Yukarıdaki örnek de bir kesikli olasılık dağılımı örneğidir.

Tesadüfi değişkenin sürekli bir değişken olması durumunda ise, sürekli bir olasılık dağılımı söz konusudur ve bu durumda olasılık dağılımına **sürekli olasılık dağılımı** ya da **olasılık yoğunluk fonksiyonu** adı verilmektedir.

11.3. Beklenen Değer (Matematik Ümit)

Bir tesadüfi değişkenin alabileceği tüm mümkün sonuçlarının olasılıklarına göre ağırlıklı ortalamasına **beklenen değer** denir.

Bir X tesadüfi değişkeninin olasılık dağılım fonksiyonunun aşağıdaki şekilde tanımlandığını düşünelim:

$X=x_i$	x_1	x_2	x_3		x_n
$P(X=x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$		$P(x_n)$

Bu durumda, X tesadüfi değişkeninin beklenen değeri $E(X)$,

$$E(X) = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + \dots + x_nP(n)$$

$$E(X) = \sum_{i=1}^n x_iP(x_i)$$

formülü ile hesaplanır.

Beklenen değer, kişisel bir beklenti durumunu değil; X tesadüfi değişkeninin uzun dönem ortalamasını ifade etmektedir. Başka bir deyişle beklenti sözcüğü, tesadüfi deneyin tekrar tekrar yinelenmesi durumunda ortaya çıkacak uzun dönem ortalaması ya da tesadüfi değişkenin hareketleneceği ya da yöneleceği ortalama değeri göstermektedir.

Beklenen değer kavramını bir örnek üzerinden açıklamaya çalışalım:

Bir zarın atılması örneğinde, üste gelen yüzün beklenen değerini hesaplayalım. Bildiğimiz gibi, zar atılması sonucunda 1, 2, 3, 4, 5, 6 sonuçlarından biri ile karşılaşılacaktır ve her bir sonucun olasılığı da eşit olup $1/6$ 'dır.

Bu durumda beklenen değer,

$$E(X) = \sum_{i=1}^n x_iP(x_i)$$

$$E(X) = \left(1 \cdot \frac{1}{6}\right) + \left(2 \cdot \frac{1}{6}\right) + \left(3 \cdot \frac{1}{6}\right) + \left(4 \cdot \frac{1}{6}\right) + \left(5 \cdot \frac{1}{6}\right) + \left(6 \cdot \frac{1}{6}\right)$$

$$E(X) = \frac{21}{6} = 3,5$$

olarak hesaplanmaktadır.

Beklenen değer hesabına yönelik bir örnek daha görelim:

Örnek:

İki arkadaş kendi aralarında bir deney tasarlıyor ve bir madeni parayı atarak yazı gelmesi halinde 10 puan, tura gelmesi halinde de 20 puan olarak kaydediyor. Bu durumda, madeni para atılma deneyinin limit durumda beklenen değerinin ne olacağını hesaplayalım.

Söz konusu tesadüfi deneyin sonuçları için oluşturulan tesadüfi değişkenin olasılık dağılımı aşağıdaki tabloda gösterilmiştir:

$X=x_i$	Yazı 10 puan	Tura 20 puan
$P(X=x_i)$	1/2	1/2

Tablodan görüleceği gibi, paranın yazı gelmesi olasılığı $\frac{1}{2}$ 'dir ve yazı gelmesinin puan karşılığı 10 puan olarak değerlendirilmektedir. Benzer şekilde paranın tura gelmesi olasılığı da yine $\frac{1}{2}$ 'dir ancak paranın tura gelmesinin puan karşılığı 20 puan şeklinde değerlendirilmektedir.

Bu durumda yukarıda tanımlanan tesadüfi değişkenin beklenen değeri,

$$E(X) = \left(10 \cdot \frac{1}{2}\right) + \left(20 \cdot \frac{1}{2}\right)$$

$$E(X) = 15$$

olmaktadır.

Beklenen değerin 15 olması, çok sayıda tekrar edecek yukarıdaki tesadüfi deney sonucunda, atış başına ortalama 15 puan kaydedileceğini ifade eder.

Örnek:

Bir futbol takımının daha önceki sezonda yapmış olduğu maçlar incelenerek, futbol takımının alabileceği puanlar ve bu puanların hangi olasılıklarla ortaya çıkabileceği aşağıdaki tablo ile düzenlenmiştir.

$X=x_i$	Mağlup olduğunda 0 puan	Berabere kaldığında 1 puan	Galip geldiğinde 3 puan
$P(X=x_i)$	0,20	0,20	0,60

Buna göre, takımın yaptığı maçlarda kazandığı puanların beklenen değerini hesaplayınız.

Futbol takımının yaptığı maçlarda aldığı puanlar ve bu puanları elde etme olasılıkları ile bir tesadüfi değişken tanımlarsak, bu tesadüfi değişkenin beklenen değeri,

$$E(X) = (0 \cdot 0,20) + (1 \cdot 0,20) + (3 \cdot 0,60)$$

$$E(X) = 2,00$$

olmaktadır.

Dolayısıyla, bu futbol takımı, oynadığı maçlar sonucunda, maç başına ortalama olarak 2 puan kazanmaktadır.

Konunun daha iyi anlaşılması için bu futbol takımının gelecek lig döneminde toplam 20 maç yapacağını varsayalım. Söz konusu futbol takımı maç başına ortalama 2 puan kazandığına göre, futbol takımımız gelecek lig döneminin sonunda 20 maç yapmış olacak ve ligi $20 \cdot 2 = 40$ puan ile tamamlaması beklenecektir.

Örnek:

Bir sigorta şirketi kendisine başvuran bireylere, prim ödemeleri karşılığında hayat sigortası hizmeti vermektedir. Sigorta şirketinin faaliyet gösterdiği ülkedeki yaşam tablolarına göre 25 yaşında bir bireyin, bir yıl boyunca hayatta kalma olasılığı 0,992, ölüm olasılığı ise 0,008'dir. Sigorta şirketi 25 yaşında olan bir bireyi, 10 TL prim ödemesi karşılığında ölüm halinde 1000 TL ödeme taahhüdü ile sigortalamaktadır.

Buna göre sigorta şirketinin beklenen kazancı nedir?

Söz konusu sigorta şirketi, 25 yaşında hayat sigortası yaptırmak isteyen bir kişiden, sigortalanması karşılığında 10 TL kazanmakta, bu kimsenin o yıl içinde ölmesi halinde ise 1000 TL ölüm tazminatı ödemektedir.

Dolayısıyla, sigortalanan kişiden şirket kasasına +10 TL girmekte, ölüm halinde ise, 1000 liranın 10 lirası sigortalanan kişinin kendi ödemesi olduğundan, ödenen 990 liralık tazminat şirket varlığından karşılanmaktadır. Yani, 10 TL sigortalanan kişinin kendi ödediği prim olduğundan şirket kasasından $1000 - 10 = -990$ TL çıkmaktadır.

Bu duruma yönelik tesadüfi değişkenin sonuçları ve olasılıkları aşağıdaki tablo ile düzenlenmiştir:

$X=x_i$	Yaşama +10 TL	Ölme -990 TL
$P(X=x_i)$	0,992	0,008

Buna göre, sigorta şirketinin beklenen kazancını hesaplayalım.

Sigorta şirketinin kasasına girecek ve çıkacak ödeme miktarları ve bu miktarların olasılıkları ile bir tesadüfi değişken tanımlarsak, bu tesadüfi değişkenin beklenen değeri,

$$E(X) = (+10 \cdot 0,992) + (-990 \cdot 0,008)$$

$$E(X) = 2,00 \text{ TL}$$

hesaplanmaktadır.

Dolayısıyla sigorta şirketi, 25 yaşında hayat sigortası yaptırmak isteyen her bireyden poliçe başına ortalama 2 TL kazanmaktadır.

Bu Bölümde Neler Öğrendik?

Bu bölümde öncelikle tesadüfi değişkenin nasıl tanımlandığını öğrendik. Bir tesadüfi deney sonucunda ortaya çıkabilecek tüm mümkün sonuçların öngörülerek tesadüfi değişkenin nasıl tanımlandığına yönelik örnekler inceledik.

Tesadüfi bir değişkenin alabileceği tüm mümkün sonuçlara, olasılıkların nasıl dağıldığını göstererek olasılık dağılımı yaratma sürecini öğrendik.

Ayrıca, bir tesadüfi değişkenin alabileceği tüm mümkün sonuçlarının olasılıklarına göre ağırlıklı ortalamasını alarak beklenen değer hesabını ve nerede kullanılmakta olduğunu öğrendik.

Bölüm Soruları

- 1- Bir tesadüfi deney sonucunda ortaya çıkabilecek mümkün sonuçlara sayısal bir değer verilerek tanımlanmış olur.
 - a- Tesadüfi deney
 - b- Olasılık dağılımı
 - c- Tesadüfi değişken
 - d- Örnek uzayı
 - e- Mümkün sonuçlar kümesi
- 2- Bir tesadüfi değişkenin alabileceği tüm değerlere olasılıkların nasıl dağıldığı gösterilerek yaratılmış olur.
 - a- Tesadüfi deney
 - b- Tesadüfi değişken
 - c- Örnek uzayı
 - d- Olasılık dağılımı
 - e- Beklenen değer
- 3- Olasılık dağılımından söz edebilmek için, mümkün sonuçlara ait olasılık toplamının olması gerekir.
 - a- Bir
 - b- Sıfır
 - c- Yüz
 - d- Birden küçük
 - e- Birden büyük
- 4- Bir tesadüfi değişkenin alabileceği tüm mümkün sonuçlarının olasılıklarına göre ağırlıklı ortalamasınadenir.
 - a- Tesadüfi deney
 - b- Beklenen değer
 - c- Örnek uzayı
 - d- Olasılık dağılımı
 - e- Tesadüfi değişken

- 5- Bir kesikli tesadüfi değişkenin olasılık dağılımına denir.
- a- Kesikli olasılık dağılımı
 - b- Sürekli olasılık dağılımı
 - c- Olasılık yoğunluğu
 - d- Olasılık yoğunluk fonksiyonu
 - e- Olasılık fonksiyonu
- 6- Sürekli bir tesadüfi değişkenin olasılık dağılımına.....denir.
- a- Kesikli olasılık dağılımı
 - b- Süreksiz olasılık dağılımı
 - c- Olasılık dağılımı
 - d- Olasılık fonksiyonu
 - e- Sürekli olasılık dağılımı
- 7- İki paranın aynı anda atılması deneyi bir tesadüfi değişken olarak tanımlanırsa, bu tesadüfi değişkenin kaç tane mümkün sonucu olabilir?
- a- 2
 - b- 4
 - c- 1
 - d- 0
 - e- 3
- 8- Hilesiz bir zarın atılması deneyi bir tesadüfi değişken olarak tanımlandığında kaç adet mümkün sonuç söz konusudur?
- a- 1
 - b- 2
 - c- 6
 - d- 3
 - e- 2

9- Zar atılması deneyi bir tesadüfî değişken olarak tanımlanırsa, üste gelen yüz kadar puan alınması söz konusu olduğunda beklenen değer nedir?

- a- $21/6$
- b- $1/6$
- c- $2/6$
- d- $15/6$
- e- $13/6$

10- Bir futbol takımının daha önceki sezonda yapmış olduğu maçlar incelenerek, futbol takımının alabileceği puanlar ve bu puanların hangi olasılıklarla ortaya çıkabileceği aşağıdaki tablo ile düzenlenmiştir.

$X=x_i$	Mağlup olduğunda 0 puan	Berabere kaldığında 1 puan	Galip geldiğinde 3 puan
$P(X=x_i)$	0,30	0,20	0,50

Buna göre, takımın yaptığı maçlarda kazandığı puanların beklenen değeri nedir?

- a- 2.00
- b- 1,5
- c- 1,8
- d- 1,6
- e- 1,7

Cevaplar

- 1- c
- 2- d
- 3- a
- 4- b
- 5- a
- 6- e
- 7- b
- 8- c

9- a

10- e

Yararlanılan Kaynaklar

ALOKA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

12. NORMAL DAĞILIM

Bölüm Yazarları:

DOÇ.DR. SEMA ULUTÜRK AKMAN

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde pek çok istatistik çalışmada kullanılan normal olasılık dağılımı üzerinde duracağız.

Normal dağılım, tabiatta gerçekleşen ya da gözlenen çok sayıda olayın olasılık dağılımıdır ve bu sebeple istatistik çalışmalarda çok önemli bir yere sahiptir.

Bu bölümde, normal dağılımın neden normal dağılım olarak adlandırıldığını, normal dağılımı karakterize eden parametreleri, standart normal dağılıma neden ihtiyaç duyulduğunu ve standart normal dağılım kullanılarak olasılık hesabının nasıl yapıldığını öğreneceğiz.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Bağlı değerlendirme sistemi nedir?
- 2) Çan eğrisi ile harf notlarının belirlenmesi nasıl olur?
- 3) Bir olasılık dağılımına neden normal dağılım denir?
- 4) Standart normal dağılım nedir ne amaçla kullanılır?
- 5) Normal dağılım uygulamada neden bu kadar önemli?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Normal dağılım.	Normal dağılımın özelliklerini ve neden normal dağılım olarak adlandırıldığı bilir.	Temel düzeyde istatistik kitaplarının olasılık ve olasılık dağılımı kısımları.
Standart normal dağılım.	Standart normal dağılımın özelliklerini ve normal dağılımdan farkını ve önemini bilir.	Temel düzeyde istatistik kitaplarının olasılık ve olasılık dağılımı kısımları.
Standart değer hesabı.	Standart değer hesaplamasının mantığı ve normal dağılım için önemini bilir.	Temel düzeyde istatistik kitaplarının standart değer ve normal dağılım kısımları.
Standart değer üzerinden normal dağılım olasılıklarını elde etme.	Standart değerlerin standart normal dağılımda neyi ifade ettiğini ve nasıl değerlendirildiğini bilir.	Temel düzeyde istatistik kitaplarının normal dağılım konusu.
Standart normal dağılım olasılık değerleri tablosundan olasılıkları okuma.	Normal eğri üzerinde standart değerlerin yerlerini tespit ederek, standart normal dağılım tablosunun buna göre nasıl kullanılacağını bilir.	Temel düzeyde istatistik kitaplarının normal dağılım konusu.

Anahtar Kavramlar

- Normal dağılım: Tabiatta pek çok tesadüfi değişkenin dağılım şekli.
- Standart normal dağılım: Normal dağılımın standart değer dönüşümü ile standartlaştırılmış hali.
- Bağlı değerlendirme: Bir dağılımda, herhangi bir değer in ortalama ve standart sapması göz önüne alınarak değerlendirilmesi, göreceli değerlendirme.

Giriş

Bir tesadüfi deneyin tüm mümkün sonuçlarına birer sayısal değer verilerek tesadüfi bir değişken tanımlanabileceğini ve bu tesadüfi değişkenin alabileceği mümkün sonuçları olasılıkları ile birlikte göstererek bir olasılık dağılımı yaratılabileceğini “olasılık” bölümünde öğrendik.

Bir tesadüfi değişkenin alabileceği tüm mümkün sonuçların toplamı 1’dir ve tam sistem değeri olan 1 değerinin söz konusu mümkün sonuçlara nasıl dağıldığını gösteren dağılıma **olasılık dağılımı** dedir.

Normal dağılıma hazırlık olması ve konular arasında bağlantı kurulabilmesi açısından, olasılık dağılımı yaratma sürecini kısaca hatırlatmakta fayda görüyoruz.

Bir X tesadüfi değişkeninin alabileceği değerler,

$$x_1, x_2, x_3, \dots, x_n$$

olsun ve bu değerlerinin olasılıkları da,

$$P(x_1), P(x_2), P(x_3), \dots, P(x_n)$$

şeklinde gösterilsin.

X tesadüfi değişkeninin alabileceği bütün mümkün değerleri olasılıkları ile birlikte aşağıdaki gibi gösterebiliriz:

$X=x_i$	x_1	x_2	x_3		x_n
$P(X=x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$		$P(x_n)$

X tesadüfi değişkeninin x_i değerlerinden yalnızca birini alabildiğini varsayalım. Bu durumda, X tesadüfi değişkeninin x_i değerini alma olasılığı,

$$P(X=x_i) = P(x_i)$$

şeklinde gösterilecektir.

Daha önce ifade ettiğimiz gibi bağdaşmayan, yani kesişim kümesi bulunmayan olaylarda, mümkün tüm sonuçların olasılıkları toplamı tam sistem oluşturur yani toplam olasılık “1” olur:

$$\sum_{i=1}^n P(x_i) = 1$$

Bir tesadüfi deneyin tüm mümkün sonuçlarını olasılıkları ile ilişkilendirdiğimizde, tam sistem olasılığı olan “1” değerini, tüm mümkün sonuçlara olasılıklarına göre dağıtmış ve bir **olasılık dağılımı** yaratmış oluruz.

Zar atılması olayının tesadüfi bir deney olduğunu ifade etmiştik. Burada, tesadüfi değişkenin alabileceği değerler 1, 2, 3, 4, 5, 6 ve her bir sonucun olasılığı da $1/6$ 'dır. Toplam olasılık da tüm mümkün sonuçların olasılıkları toplanarak “1”i yani tam sistem değerini vermektedir. Dolayısıyla, zar atılması sonucunda da bir olasılık dağılımı tanımlanmaktadır.

Olasılık dağılımları, ait oldukları tesadüfi değişkenin kesikli ya da sürekli olması durumuna göre, kesikli ve sürekli olasılık dağılımları şeklinde iki ana gruba ayrılır.

Tesadüfi değişkenin kesikli olması halinde, kesikli olasılık dağılımı söz konusu olmaktadır ve binom, hipergeometrik ve poisson dağılımları uygulamada en çok karşılaştığımız kesikli olasılık dağılımlarıdır.

Tesadüfi değişkenin sürekli olması durumunda ise, sürekli olasılık dağılımı ortaya çıkmaktadır ve bu bölümün de konusu olan normal dağılım sürekli olasılık dağılımları içinde çok önemli bir yere sahiptir.

Normal Dağılımı biraz daha yakından tanıyalım:

12.1. Neden “Normal” Dağılım?

Normal dağılım, tabiatı gerçekte gerçekleşen ya da gözlenen pek çok olayın olasılık dağılımının normal dağılıma uygun olması sebebiyle “normal” dağılım olarak adlandırılmaktadır. Söz gelimi, bir toplumda yaşayan belirli sayıda insana bir zekâ testi uygulansa, zekâ testi skorlarının normal dağıldığı gözlenir. Benzer şekilde, bir hastanede dünyaya gelen bebeklerin doğum ağırlıkları bir yıl boyunca gözlenirse, bebeklerin doğum ağırlıklarının da normal dağılıma uygunluk gösterdiği görülür. Yine aynı şekilde, bir margarin fabrikasında üretilen margarinlerin ağırlıkları, belirli bir güzergâhta bir ay boyunca izlenen otobüslerin ulaşım süresi vb. örnekler de normal dağılıma örnek verilebilir.

Gözlem değerlerinin dağılımına kabaca baktığımızda, ortalama değer etrafında toplanma ve ortalamadan uzaklaştıkça gözlem sayısının azaldığını gözlemliyoruz. Zekâ testi örneğini ele alalım. Geniş bir kitleye bu zekâ testini uyguladığımızda, ortalama zekâ testi skoru etrafında gözlem değerlerinin toplandığını ve ortalama değerden uzaklaştıkça insan sayısının yavaş yavaş azaldığını, en düşük seviyede ve en yüksek seviyede zekâ skoruna sahip olanların sayısının ise çok çok azaldığını gözlemliyoruz. Bu durum, olağan, beklenen ya da normal bir durum olarak değerlendirilir.

Normal dağılıma uygunluk gösteren örneklerle baktığımızda, her birinin sürekli bir tesadüfi değişken ile tanımlandığını görürüz. Nitekim normal dağılım da sürekli bir olasılık dağılımıdır.

Öte yandan normal dağılım, pratikte pek çok sürekli değişkenin olasılık dağılımı olarak karşımıza çıkan ve bu sebeple “normal” olarak adlandırılan bir olasılık dağılımıdır. Bilindiği gibi normal sözcüğü, olağan, sıradan ve yaygın olarak görülen vb. anlamlarda kullanılan bir sözcüktür. Normal dağılıma “normal” denmesinin de kökeninde yaygın görülen bir dağılım olması yatar.

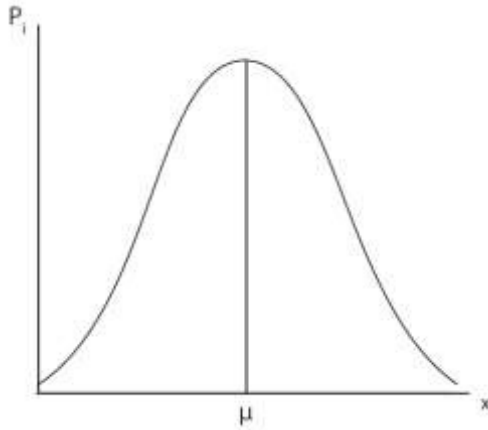
12.2. Normal Dağılım: Kuramsal Çerçeve

Tüm olasılık dağılımlarında, dağılımı karakterize etmek yani nitilemek üzere iki önemli ölçü kullanılır. Bu ölçüler, dağılımın beklenen değeri, diğer deyişle dağılımın ortalaması ve dağılımın standart sapması veya varyansıdır.

Normal dağılım da, ortalaması; μ ve standart sapması; σ (ya da varyansı σ^2) ile ifade edilen sürekli bir dağılımdır.

Normal dağılımın şekli simetriktir ve şeklinin çana benzemesi sebebiyle çan eğrisi olarak da bilinir.

Sürekli bir tesadüfi değişken söz konusu olduğundan ve süreklilik durumu, değişkenin iki değer arasında sonsuz değer alabilme özelliği olarak tanımlandığından, tesadüfi değişkenin alabileceği değerler birbirine çok yaklaşmaktadır. Bu durum, dağılımın şeklini bir eğri formuna dönüştürür. Zira, tesadüfi değişkenin alabileceği değerler birbirine o kadar yakındır ki, tesadüfi değişkenin alacağı değerleri gösteren X eksenindeki noktalar birbirine çok yaklaşmıştır ve normal dağılımın şekli de bu sebeple eğri biçimindedir. Aşağıda verilen Şekil 12.1.’den de görüleceği gibi normal dağılım, dağılımın ortalama değeri (μ) üzerinde maksimum olasılığa sahip, simetrik ve dağılımın sağ ve sol kuyruklarının X eksenine asimptot olduğu yani, dağılımın kuyruklarının sonsuzda X eksenini kestiğinin varsayıldığı, bir dağılımdır.



Şekil 12.1: Normal Dağılım

Normal dağılımın sürekli bir tesadüfi değişkenin olasılık dağılımı olması sebebiyle, X ekseninde yer alan tesadüfi değişkenin alabileceği değerler birbirine o derece yakındır ki, tek bir değer için olasılık değeri sıfır olur.

Normal dağılımda olasılık hesabı, iki değer arasındaki alan hesabıyla yapılır ve eğrinin altındaki toplam alan, toplam olasılık olan ve tam sistem değerini ifade eden “1”e eşittir. Dağılımın orta noktası olan ortalama yani μ değeri, eğrinin altında kalan alanı iki eşit parçaya ayırmakta olup, ortalamanın sağında ve solunda kalan alanlar 0,5 olasılığı ifade eder.

Yani, tesadüfen seçilecek bir birimin ortalamanın altında değer alması olasılığı 0,5’dir, benzer şekilde ortalamanın üstünde değer alma olasılığı da yine 0,5 olmaktadır.

Bir kez daha ifade etmek gerekirse ortalama, dağılımı iki eşit parçaya bölmekte ve dolayısıyla, eğrinin altında kalan ve toplam olasılığı ifade eden “1” değeri de ortalamanın sağında ve solunda kalan kısımlara 0,5 olacak şekilde dağılmaktadır.

Normal dağılımın, sürekli bir olasılık dağılımı olması sebebiyle, olasılık yoğunluk fonksiyonu olarak nitelendiğini daha önce ifade etmiştik.

Normal dağılımın olasılık yoğunluk fonksiyonu,

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2} \quad -\infty \leq X \leq +\infty$$

formülüyle ifade edilmektedir. Ayrıca, yukarıdaki normal dağılım kısaca, $X \sim N(\mu, \sigma)$ şeklinde yazılarak, “X tesadüfi değişkeni μ ortalama ve σ standart sapma ise normal dağılmaktadır” şeklinde okunmaktadır.

Yukarıdaki normal olasılık yoğunluk fonksiyonunun $-\infty$ ile $+\infty$ arasında tanımlı bir fonksiyon olduğunu görüyoruz.

Normal dağılımın formülünde yer alan ve sembollerin neler olduğunu açıklayalım:

X, sürekli bir tesadüfi değişkeni,

μ , tesadüfi değişkenin ortalamasını,

σ , tesadüfi değişkenin standart sapmasını,

π , matematikte sabit pi sayısını, 3.14159...

e, doğal logaritma tabanını, 2,71828...

ifade etmektedir.

Formülden de görüleceği gibi, normal dağılım fonksiyonunu niteleyen iki ölçü bulunmaktadır: Ortalama ve standart sapma. Dolayısıyla, formülde yer alan diğer tüm değerlerin sabit olması sebebiyle, fonksiyonun belirleyicisi ya da niteleyicisi ortalama ve standart sapma olmaktadır. Bu sebeple de ortalaması ve standart sapması farklı sonsuz sayıda normal dağılımdan söz edilebilmektedir.

Normal dağılım sürekli bir dağılım olduğundan olasılıklar, alan hesabı yapılarak bulunur.

Normal dağılım fonksiyonunun altında kalan alanın toplam olasılık değeri olan “1” e eşit olduğunu ifade etmiştik. Normal dağılım fonksiyonunun $-\infty$ ile $+\infty$ aralığında tanımlı olması ve toplam alanın da bir olması sebebiyle, dağılım fonksiyonunun söz konusu tanım aralığında integrali alındığında,

$$\int_{-\infty}^{+\infty} f(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2} = 1$$

eşitliği gerçekleşir.

Öte yandan, olasılık hesabının, alan hesabı ile yapılması sebebiyle $a < b$ olmak koşuluyla a ve b gibi iki değer arasındaki olasılık, yani a ile b arasındaki alan hesabının da,

$$\int_a^b f(X) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2}$$

biçimindeki integralin hesaplanmasıyla yapılması gerekecektir.

12.3. Standart Normal Dağılım

Daha önce de ifade ettiğimiz gibi, ortalaması ve standart sapması farklı olan her normal dağılım için olasılık hesabı, integral çözümlemesi ile yapılacaktır ve yukarıdaki denklemin integrali de ortalaması ve standart sapması farklı her normal dağılım için tekrar tekrar hesaplanacaktır.

Bu oldukça zahmetli integral hesaplarını yaparak olasılıkları hesaplamak yerine, z dönüşümü ile herhangi bir normal dağılımı standart normal dağılıma dönüştürmek ve standart normal dağılım için hazırlanmış olasılık değerleri tablosunu kullanarak ilgili alanı yani olasılığı kolaylıkla hesaplamak mümkündür.

$X \sim N(\mu, \sigma)$ gösterimi gereği μ ortalama ve σ standart sapma ise normal dağılan X tesadüfi değişkenine ilişkin her X_i değerine karşılık gelecek bir z_i değeri hesaplanmakta, başka bir deyişle X tesadüfi değişkeni z tesadüfi değişkenine dönüştürülmektedir. Standart z değişkeni de $z \sim N(0,1)$ sıfır ortalama ve bir standart sapma ile normal dağılmaktadır.

Normal dağılım fonksiyonunda yer alan formülün,

$$((X_i - \mu)/\sigma)^2$$

kısımının, dağılım ölçüleri bölümü içinde öğrendiğimiz z dönüşümü yani, standart değer dönüşümü formülü olduğunu hatırlayalım. Adı geçen bölümde açıkladığımız gibi z dönüşümü,

$$z = \frac{X_i - \bar{X}}{\sigma}$$

formülüyle yapılmaktadır ve bir değerin kendi ortalamasından kaç standart sapma uzağa düştüğünü gösteren bir ölçüdür.

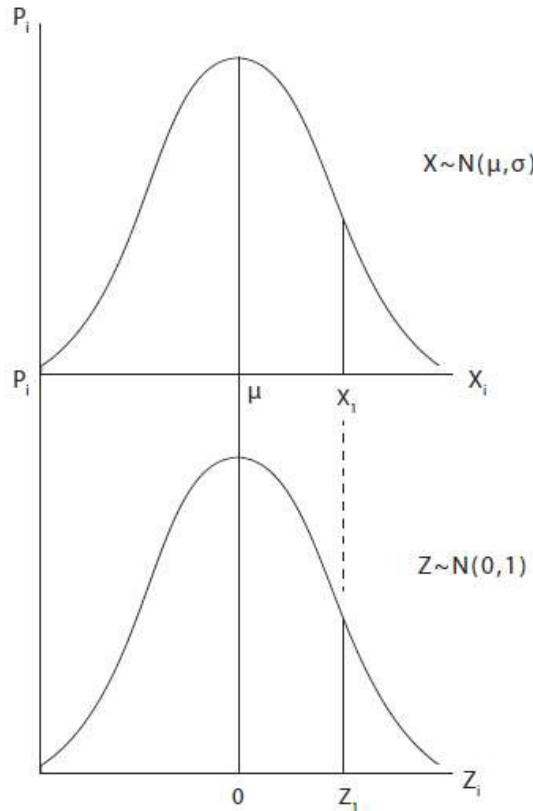
$\bar{X}$ yerine μ yazarsak, z dönüşüm formülü,

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

olmaktadır. Yukarıdaki her iki z dönüşüm formülü de ortalamayı ve standart sapmayı kullanarak standartlaştırma işlemini gerçekleştirir. Formüllerde yer alan $\bar{X}$ ve μ değeri dağılımın ya da tesadüfi değişkenin ortalamasını ifade etmektedir. İki sembolün arasındaki farkı örnekleme konusu içinde ayrıntılı olarak ele alacağız. Bu aşamada her iki sembolün de ortalamayı ifade ettiğini bilmek yeterlidir.

Standart değer (z) dönüşümü vasıtasıyla, μ ortalama ve σ standart sapma ise normal dağılan X tesadüfi değişkenine ilişkin her X_i değerine karşılık 0 ortalama ve 1 standart sapma ile normal dağılan bir z_i değeri hesaplanmakta ve bu şekilde oluşacak her z_i değeri için önceden hazırlanmış olan normal olasılık dağılımı tablosu üzerinden olasılıklar kolaylıkla okunacaktır.

Burada sözüne ettiğimiz z dönüşümünü Şekil 12.2. yardımıyla açıklamaya çalışalım:



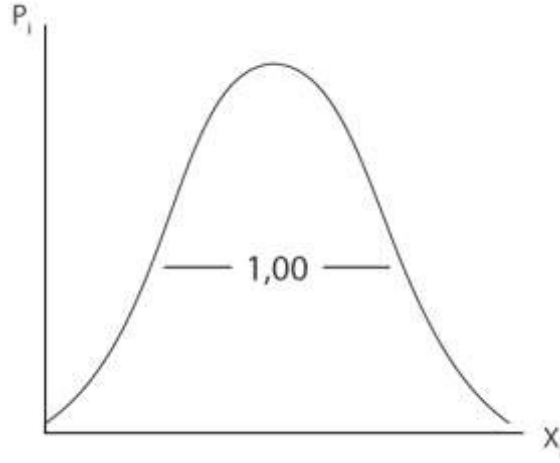
Şekil 12.2: Normal Dağılım ve Standart Normal Dağılım

12.4. Standart Normal Dağılım Tablosu Kullanılarak Olasılıkların Belirlenmesi

Normal dağılıma sahip bir tesadüfi değişkenin belirli değerleri alabilmesine yönelik olasılıkların, standart normal dağılım dönüşümü ve normal olasılık değerleri tablosu kullanılarak nasıl belirlendiğini öğrenelim. Standart normal dağılım olasılıkları tablosuna (normal eğri alanları tablosuna) bölüm sonunda yer verilmiştir.

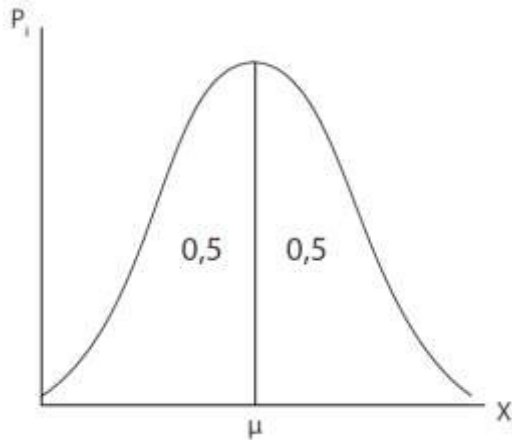
Standart değer dönüşümü ve standart normal dağılım için oluşturulan normal eğri alanları tablosu kullanılarak olasılıkların nasıl belirlendiğine geçmeden önce, normal dağılım konusuyla ilgili olarak mutlaka bilinmesi ve her zaman hatırlanması gereken dört temel kural üzerinde durmak istiyoruz.

- Normal eğrinin altında kalan toplam alan, daima “1”dir.



Şekil 12.3: Normal Eğri Altındaki Alan

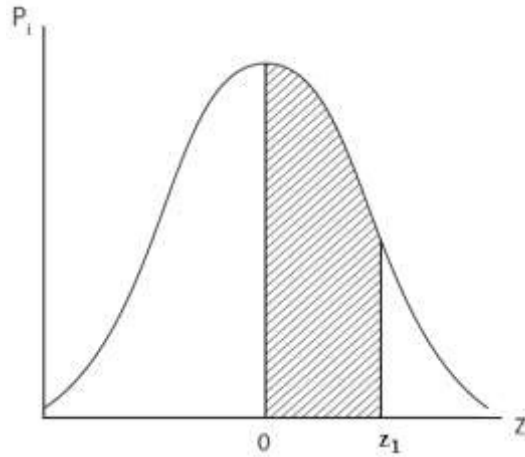
- Normal dağılım söz konusu olduğunda aritmetik ortalama, eğri altındaki alanı iki eşit kısma ayırır. Yani, ortalamanın sağında 0,5 ve solunda 0,5 olasılıkla gerçekleşecek iki alan oluşur.



Şekil 12.4: Normal Eğri Altındaki Alan ve Ortalama

- Standart normal dağılım tablosu ortalama ile ilgili z değeri arasındaki alanı yani olasılığı vermektedir. Dolayısıyla, bunun dışında olasılık hesabı yapmak gerektiğinde yukarıda bahsedilen eğri altındaki toplam alanın “1” olduğu ve ortalamanın eğri altındaki alanı 0,5 olasılıklı iki eşit kısma ayırdığı unutulmamalıdır.
- Temelinde sürekli bir değişkenin bulunması sebebiyle, normal dağılımda tesadüfi değişkenin alabileceği değerler birbirine o derece yaklaşmıştır ki tek bir değer için olasılık sıfırdır. Dolayısıyla, $P(X < 40)$ ya da $p(X \leq 40)$ arasında bir fark oluşmamaktadır.
- Standart normal eğri alanları tablosu sadece pozitif z değerleri için hazırlanmıştır. Normal dağılımın simetrik olması ve aritmetik ortalamanın eğri alanını iki eşit kısma ayırması sebebiyle, z değerinin pozitif olması üstünde bir değer söz konusu olduğunu, z değerinin negatif olması ise ortalamanın altında bir değer söz konusu olduğunu gösterir. Bir z değerinin pozitif ya da negatif olması sadece ortalamanın altında ya da üstünde bir değer olduğunu ifade eder, + ya da – bölgede olsa da ortalamaya uzaklık aynı olacağı için olasılık değerleri de aynı olacaktır.

Standart normal dağılım tablosunda yer alan olasılıkların hangi bölgeye ilişkin olduğunu gösteren şekil aşağıda verilmektedir. Standart normal dağılım olasılıklarına ilişkin hazırlanmış farklı tablolar bulunmakla beraber, her tabloda olasılık değerleri ve tablo hazırlanırken hangi bölgenin dikkate alındığı mutlaka taranarak belirtilmektedir. Aşağıda, İstatistik dersi kapsamında kullanacağımız standart normal dağılım tablosunda geçerli olan yaklaşım Şekil 12.5. üzerinde gösterilmektedir.



Şekil 12.5: Standart Normal Dağılım Tablosunda Okunan Olasılıkların Hangi Bölgeye Ait olduğu

Normal dağılımla ilgili yapılacak tüm uygulamalarda bu dört temel özelliği mutlaka hatırlamak ve dikkate almak gerekir.

Normal dağılımla ilgili olarak yapılacak uygulamalarda karşılaşılabilecek tüm mümkün durumlar için birer çözüm örneği vermenin konunun daha kolay anlaşılmasını sağlayacağını düşünüyoruz.

Örnek:

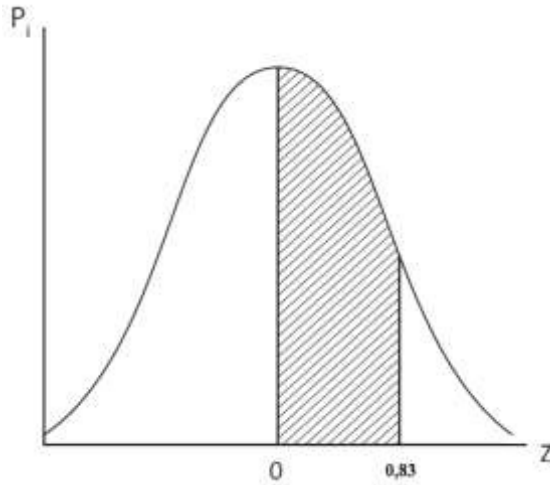
Bir kamu bankası personel alımı için ilana çıkmış ve iş için başvuran çok sayıda adaya, ekonomi, maliye, finans, işletme, hukuk alanlarında hazırlanan bir test uygulamıştır. Adayların uygulanan testte aldıkları puanların ortalaması 60 ve standart sapması 12 hesaplanmıştır. Adayların aldıkları puanların normal dağıldığı bilindiğine göre, aşağıdaki şıklarda istenen olasılıkları hesaplayınız.

1- Tesadüfen seçilecek bir adayın 60-70 arasında puan almış olma olasılığı nedir?

Bunun için öncelikle 70 değerine karşılık gelen z standart değerini hesaplayalım:

$$z_{70} = \frac{X_i - \mu}{\sigma} = \frac{70 - 60}{12} = + 0,83$$

Şekil 12.6. dan da görüldüğü gibi, istenen olasılık herhangi bir adayın 60-70 aralığında, yani standart z değeri olarak da 0-0,83 aralığında puan almış olma olasılığının hesaplanmasıdır.



Şekil 12.6: Ortalama ile ortalamadan büyük bir değer arasındaki alan için olasılık hesabı

Standart normal dağılım tablosunda +0,83 z değerine karşılık gelen olasılık değerini belirleyelim. Bunun için, tablonun 0,0 ile başlayıp aşağıya doğru devam eden ilk sütununda 0,8 değerinin yer aldığı satırı buluyoruz. z değerimiz 0,83 olduğu için 0,83 - 0,80 = 0,03 değerini de 0,00 ile başlayan ilk satırda belirliyor ve sonra bu iki sütun ve satırın kesişiminde yer alan değeri buluyor ve okuyoruz.

Tablodan da görüleceği gibi, ilk sütundaki 0,8 değerinin bulunduğu satır ile ilk satırda yer alan 0,03 değerinin bulunduğu sütunun kesişim noktasında 0,2967 değeri yer almaktadır. Dolayısıyla, tesadüfen seçilecek bir öğrencinin 60-70 arasında puan almış olma olasılığı %

29,67'dir. Başka bir deyişle sınava giren adayların % 29,67'sinin 60-70 arasında puan alması beklenir.

2- Tesadüfen seçilecek bir adayın 45-60 arasında puan almış olması olasılığı nedir?

Yine öncelikle 45'in z standart değerini hesaplamamız gerekecektir. Dağılımın ortalamasının 60 ve standart sapmasının da 12 olduğunu hatırlarsak 45'e karşılık gelen z değeri,

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = - 1,25$$

hesaplanmış olur.

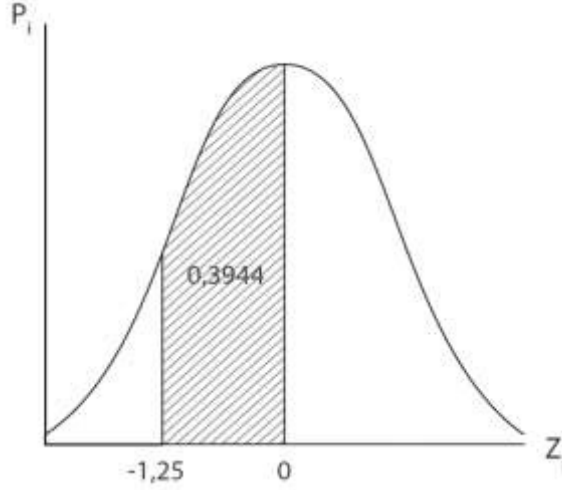
z değerinin negatif çıkmasının, söz konusu puanın ortalamanın altında bir puan olduğunu göstermek dışında bir anlamı yoktur. Zira, normal dağılım simetrik bir dağılım olduğundan ve ortalama bu dağılımı tam ortadan iki eşit kısma ayırdığından, - 1,25 değeri ile + 1,25 değeri ortalamaya eşit uzaklıktadır.

Standart normal dağılım tablosu sadece pozitif z değerleri için düzenlenmiştir. Ancak, dağılımın simetrik olması sebebiyle negatif z değerleri için de aynı olasılıklar geçerlidir. Bu sebeple, z değeri negatif olduğunda da, işaret ihmal edilerek, z değerine karşılık gelen alan standart normal dağılım tablosundan aynı şekilde bulunur. z değerinin negatif olması sebebiyle, tablodan okunan alanın, ortalama ile ortalamanın altında yer alan söz konusu z değeri arasındaki alan olduğu uygulamacı tarafından bilinir.

Şimdi, sınava giren adaylar arasından tesadüfen seçilecek bir adayın, 45 ile 60 arasında değer alma olasılığını yani $P(45 \leq X \leq 60)$ 'ı belirleyelim (standart değer olarak da $P(-1,25 \leq z \leq +1,25)$).

Bunun için, standart normal dağılım tablosunun 0,0 ile başlayan ilk sütununda 1,2 değerinin bulunduğu satırı belirliyoruz. Daha sonra $1,25 - 1,2 = 0,05$ değerini de 0,0 ile başlayan ilk satırda arıyoruz. Daha sonra, ilk sütunda yer alan 1,2 satırı ile ilk satırda yer alan 0,05 sütununun kesiştiği hücredeki değeri okuyoruz. Bu değer, 0,3944 olarak tespit ediliyor.

Dolayısıyla, sınava giren adaylar arasından tesadüfen seçilecek birinin 45-60 arasında puan almış olma olasılığı % 39,44'dür ya da diğer deyişle, sınava giren adayların % 39,44'ü 45-60 arasında puan almıştır.



Şekil 12.7: Ortalama ile ortalamadan küçük bir değer arasındaki olasılığın belirlenmesi

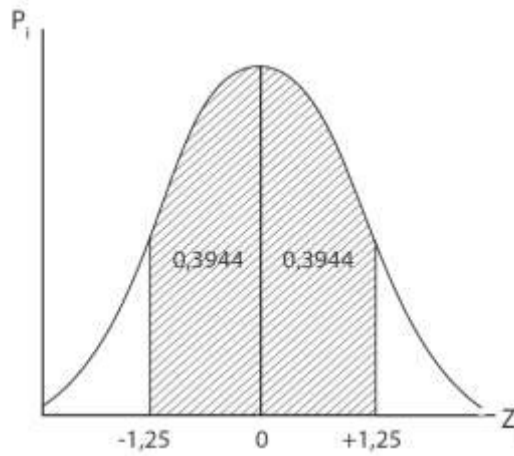
3- Tesadüfen seçilecek bir adayın 45-75 arasında puan almış olma olasılığı nedir?

Bu örnekte, 45 değeri ortalama değer olan 60'ın altında, 75 değeri ise üzerinde yer almaktadır. Her zaman olduğu gibi öncelikle 45 ve 75 puanlarına karşılık gelen z değerlerini hesaplamamız gerekecektir.

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1,25$$

$$z_{75} = \frac{X_i - \mu}{\sigma} = \frac{75 - 60}{12} = +1,25$$

Görüldüğü gibi, 45 ve 75 puanlarına karşılık gelen z değerleri sırasıyla $-1,25$ ve $+1,25$ hesaplanmıştır. Şimdi bu değerleri normal dağılım üzerinde göstererek, hesaplamamız gereken alanı tarayalım.



Şekil 12.8: $P(-1,25 \leq z \leq 1,25)$

Standart normal olasılık değerleri tablosundan + 1,25 için olasılık değeri, ilk sütunda 1,2; ilk satırda ise (1,25 - 1,2=0,05) 0,05 değerini bularak ilgili satır ve sütunların kesişim noktasındaki olasılık 0,3944 olarak tespit ediliyor. Bu değerin 0 ile +1,25 arasındaki alanın olasılığını vermekte olduğuna dikkat ediniz.

Öte yandan, -1,25 için de tablodan olasılık değerinin 0,3944 olduğu anlaşılıyor bu değer de ortalamanın solunda kalan alan -1,25 ile 0 arasındaki alanın olasılığını veriyor.

Şekil 12.8'den de görüleceği gibi, ortalamanın sağında kalan +1,25 değeri ile ortalama arasındaki alan 0,3944; ortalamanın solunda kalan yani ortalama ile -1,25 değeri arasındaki alanda 0,3944 olduğundan $-1,25 \leq 0 \leq +1,25$ arasındaki toplam taralı alan iki alanın toplamı alınarak bulunacaktır.

Dolayısıyla $-1,25 \leq 0 \leq +1,25$ arasındaki toplam alan,

$$0,3944 + 0,3944 = 0,7888$$

% 78,88 hesaplanmaktadır.

Dolayısıyla, adaylar arasından tesadüfen seçilecek bir adayın 45 ile 75 arasında puan alması olasılığı % 78,88'dir. Başka bir deyişle sınava giren adayların % 78,88'i 45-75 aralığında puan almıştır.

4- Tesadüfen seçilecek bir adayın 80'den fazla puan almış olma olasılığı nedir?

80 puana karşılık gelen z değeri,

$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = + 1,67$$

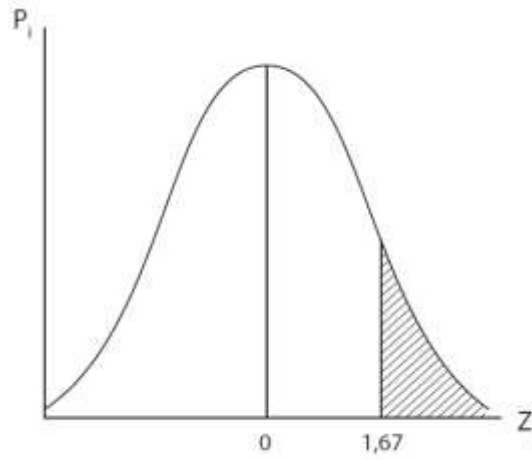
olarak hesaplanıyor. Adayın 80 puandan fazla alma olasılığı istendiğine göre, $P(X > 80)$ olasılığının hesaplanması gerekiyor. Dolayısıyla, $z > 1,67$ bölgesinin olasılığını bilmek gerekiyor. $z=1,67$ değerine karşılık gelen olasılığı standart normal dağılım tablosundan, 0,4525 olarak belirliyoruz (İlk sütunda yer alan 1,6 değeri ile ilk satırda yer alan (1,67-1,6=0,07) 0,07 değerlerinin satır ve sütunlarını kesiştirdiğimizde 0,4525 değerine ulaşıyoruz).

0,4525 değeri ortalama ile 80 arasındaki alana ilişkin olasılık değeridir. Standart değer olarak da (0 - 1,67) arasındaki alanı ifade eder. Oysa bizden istenen 80'den fazla puan almış olma olasılığıdır ki bu $z > 1,67$ bölgesidir.

Ortalamanın sağında kalan alanın 0,5 olduğunu hatırlayalım. Bu durumda, $z > 1,67$ bölgesinin alanını bulmak için iki alanın farkını alarak istenen olasılığı,

$$0,5 - 0,4525 = 0,0475$$

hesaplıyoruz. Dolayısıyla, tesadüfen seçilecek bir adayın 80'den fazla puan almış olma olasılığı, % 4,75 olmaktadır. Başka bir deyişle, bu sınava giren adayların % 4,75'i 80 puanın üzerinde puan almıştır.



Şekil 12.9: $P(z > 1,67)$

5- Tesadüfen seçilecek bir adayın 45'den az puan alma olasılığı nedir?

Standart değer dönüşümü ile, 45'e karşılık gelen z değerini,

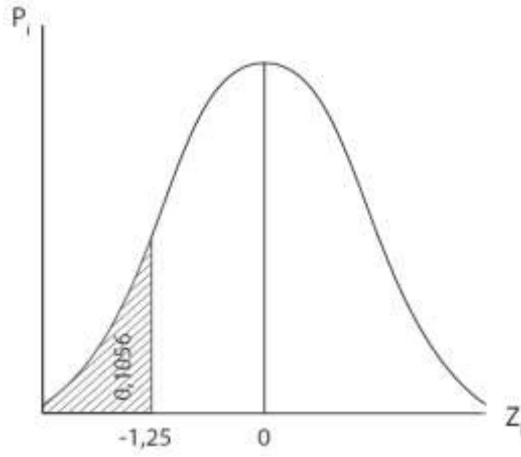
$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1,25$$

hesaplıyoruz. Bu değerın standart normal dağılım olasılıkları tablosundaki olasılık değeri, ilk sütunda yer alan 1,2 değeri ile ilk satırda yer alan 0,05 değerlerinin satır ve sütunlarının kesiştirilmesi sonucunda 0,3944 olarak belirleniyor. Bu değer, ortalama (60) ile 45 arasındaki alanın olasılığıdır. Dolayısıyla, 45'in sol tarafındaki alanı bulmak için, ortalamanın solunda kalan tüm alanın 0,5 olduğu hatırlanarak ve iki alanın farkının alınması gerekir.

Ortalamanın solunda kalan tüm alandan yani 0,5'den , ortalama ile -1,25 değeri arasındaki alan çıkarılarak olasılık,

$$0,5 - 0,3944 = 0,1056$$

olarak hesaplanıyor. Bu durumda, tesadüfen seçilecek bir adayın 45'den az puan alma olasılığı % 10,56'dır, Başka bir deyişle sınava giren adayların % 10,56'sı 45'den az puan almıştır.



Şekil 12.10: $P(-1,25 < z)$

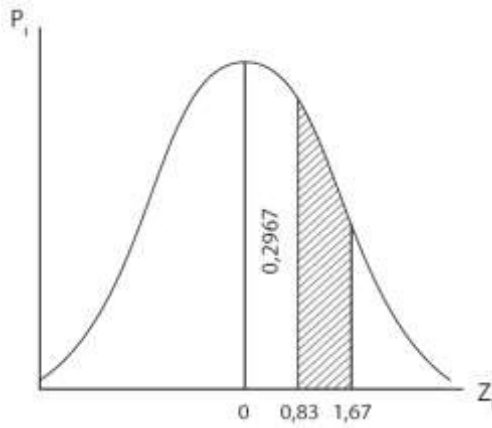
6- Tesadüfen seçilecek bir adayın 70-80 arasında puan alması olasılığı nedir?

Öncelikle 70 ve 80 puanlarının standart değer karşılıklarını bulalım:

$$z_{70} = \frac{X_i - \mu}{\sigma} = \frac{70 - 60}{12} = + 0,83$$

$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = + 1,67$$

Her iki z değeri de pozitif ve ortalamanın sağında kalan bölgede yer almaktadır. Şekil üzerinde, z değerlerini göstererek, bizden istenen alanı belirleyelim. Şekil 12.11'den 70 ve 80'e karşılık gelen z değerleri ve istenen taralı alan takip edilebilir.



Şekil 12.11: $P(0,83 \leq z \leq 1,67)$

Bu örnekte olasılığı istenen alan, $70 \leq X \leq 80$ arasında kalan alandır ve bu alan z değeri olarak da $0,83 \leq z \leq 1,67$ aralığına karşılık gelmektedir.

$0 \leq z \leq 1,67$ alanı standart normal dağılım tablosundan 0,4525;

$0 \leq z \leq 0,83$ alanı ise yine standart normal dağılım tablosundan 0,2967 olarak belirleniyor.

Bu durumda şekil 6.11'den de izlenebileceği gibi $0,83 \leq z \leq 1,67$ arasındaki alan, büyük alandan küçük alanın çıkartılması suretiyle,

$$0,4525 - 0,2967 = 0,1558$$

hesaplanıyor. Dolayısıyla, iş başvurusu yapan adaylar arasından seçilecek herhangi birinin 70 ile 80 arasında puan alma olasılığı % 15,58 olarak hesaplanmakta, başka bir deyişle adayların % 15,58'i 70-80 arasında puan almış olmaktadır.

7- Tesadüfen çekilecek bir adayın 35-45 arasında puan almış olması olasılığı nedir?

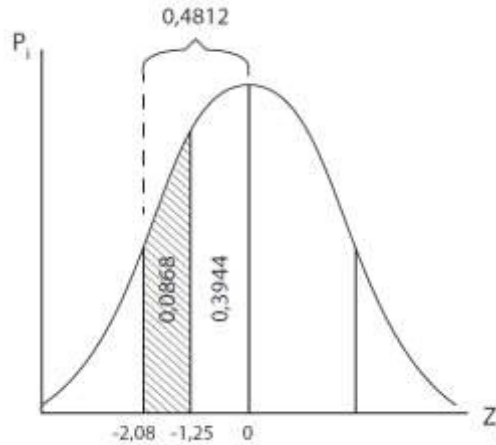
Öncelikle 35 ve 45 puanlarına karşılık gelen z değerlerini hesaplayalım:

$$z_{35} = \frac{X_i - \mu}{\sigma} = \frac{35 - 60}{12} = -2,08$$

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1,25$$

35 ve 45 puanlarına karşılık gelen her iki z değeri de negatif işaretlidir zira, her iki değer de ortalamanın altındadır.

Şekil 12.12'den de görüleceği gibi, 35-45 aralığında puan almış alma olasılığını hesaplayabilmek için iki z değerine karşılık gelen alanları bulmak ve sonra farkını almak gerekecektir.



Şekil.12.12: $P(-2,08 \leq z \leq -1,25)$

45'e karşılık gelen z değeri olan 1,25 değerinin standart normal dağılım tablosundaki karşılığı 0,3944'dür ve bu olasılık herhangi bir adayın 45-60 arasında puan alma olasılığıdır.

35'e karşılık gelen z değeri ise 2,08 olup bu değerin standart normal dağılım tablosundaki olasılık değeri de 0,4812'dir ve bu olasılık 35-60 arasında puan alma olasılığını vermektedir.

O halde, Şekil 12.12'den de görüleceği gibi, 35-45 arasında puan alanların olasılığını belirlemek için 35-60 puan alanından 45-60 puan alanını çıkarmak gerekecektir. Bu durumda, tesadüfen seçilecek bir adayın 35-45 arasında puan almış olma olasılığı,

$$0,4812 - 0,3944 = 0,0868$$

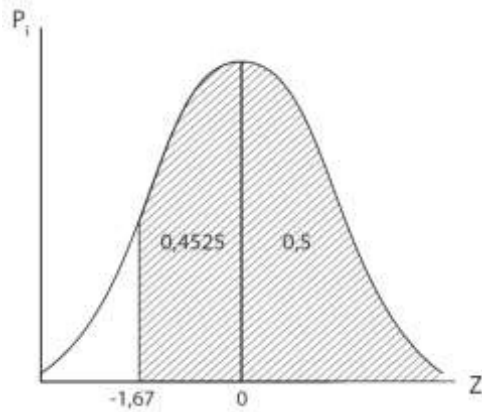
biçiminde hesaplanmaktadır. Bu değer aynı zamanda, test sınavına giren personel adaylarının % 8,68'inin 35-45 puan aralığında puan aldığının da göstergesidir.

8- Tesadüfen seçilecek bir adayın 40'dan fazla puan almış olma olasılığı nedir?

Önce, 40 puana karşılık gelen standart değeri hesaplayalım:

$$z_{40} = \frac{X_i - \mu}{\sigma} = \frac{40 - 60}{12} = -1,67$$

Z=1,67 değerine karşılık gelen standart normal dağılım tablosundaki olasılık değeri 0,4525'dir ve bu değer, ilgili z değeri ile ortalamayı ifade eden sıfır değeri arasındaki alanı; yani puan ölçeğinde 40-60 arasındaki alanı vermektedir. Bu durum Şekil 12.13'den de anlaşılmaktadır.



Şekil 12.13: P(z > -1,67)

Standart normal dağılım olasılıkları tablosunda, z=1,67 değerinin olasılığı, tablonun ilk sütununda 1,6; ilk satırında da 0,07 değerlerinin bulunması ve bu iki değerlerin bulunduğu satır ve sütunların birbiriyle kesiştiği noktada yer alan değeri tespit etmek suretiyle bulunur. Nitekim, 1,6 ile 0,07 değerlerinin kesiştiği noktada 0,4525 değeri bulunmaktadır.

0,4525 olasılığı, ilgili z değeri ile ortalama arasındaki alanı verir. Yani, 40-60 arasında kalan alan 0,4525 olmaktadır. Oysa bizden istenen 40 puanın üstünde puan alanların olasılığıdır. Bu nedenle, ortalamanın eğri altındaki toplam alanı ikiye böldüğü ve her bir alanın 0,5 olduğu kuralını hatırlayarak 40 puanın üstünde alma olasılığını,

$$0,5 + 0,4525 = 0,9525$$

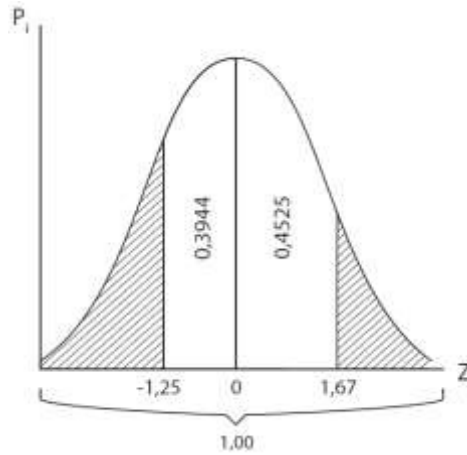
hesaplıyoruz. Dolayısıyla, tesadüfen seçilecek bir adayın 40 puanın üstünde puan almış olma olasılığı % 95,25’dir veya başka bir deyişle, sınava girenlerin % 95,25’i 40 puandan fazla puan almıştır.

9- Tesadüfen seçilecek bir adayın 45 puandan az ya da 80 puandan fazla puan almış olma olasılığı nedir?

45 ve 60 puanlara karşılık gelen z değerlerini bularak, şeklimizi çizelim ve istenen olasılık bölgesini tarayalım.

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1,25$$

$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = +1,67$$



Şekil 12.14: $P(-1,25 < z ; z > 1,67)$

Şekil 12.14’ten de görüleceği gibi, bizden istenen tesadüfen seçilecek bir adayın 45 puandan az ya da 80 puandan fazla almış olma olasılığıdır. Dolayısıyla, taralı bölgelerden de anlaşılacağı gibi eğrinin -1,25 değerinin solunda ve +1,67 değerinin sağında kalan alanlarının toplamı istenmektedir.

-1,25 standart değeri için standart normal olasılık değerleri tablosundan olasılığı 0,3944 ve +1,67 standart değeri için de olasılığı 0,4525 olarak belirliyoruz. Bu olasılık değerleri ilgili z değeri ile ortalamayı ifade eden sıfır değeri arasındaki alanları göstermektedir.

Bizden istenen kuyruklarda yer alan taralı bölgeler olduğuna göre, bu aşamada çözüm iki değişik yaklaşımla bulunabilir. Şöyle ki,

- Önce z değerlerine karşılık gelen olasılıklar toplanarak, bulunan değer toplam olasılık olan “1”den çıkarılır. Bu noktada eğrinin altında kalan toplam alanın “1” olduğunu hatırlayalım.

Bu çözüm yöntemi benimsenirse istenen olasılık,

$$P(X < 45 \text{ veya } X > 80) = 1 - (0,3944 + 0,4525)$$

$$P(X < 45 \text{ veya } X > 80) = 1 - 0,8469$$

$$P(X < 45 \text{ veya } X > 80) = 0,1531$$

olarak hesaplanır.

- İkinci bir yaklaşımda ise istenen olasılık, z değerlerine göre standart normal dağılım tablosundan elde edilen olasılıkları, ortalamanın sağında ve solunda kalan alanların 0,5 olasılığa sahip olduğunu dikkate alıp ayrı ayrı 0,5’den çıkararak kuyruk bölgelerinin alanını ayrı ayrı bulmak ve daha sonra çıkan sonuçları toplamak suretiyle de hesaplanabilir.

Şimdi önce, tesadüfen seçilecek bir adayın 40 puandan az almış olma olasılığını hesaplayalım. Ortalamanın solunda kalan alan 0,5 olduğuna göre, istenen kuyruk bölgesinin olasılığı,

$$0,5 - 0,3944 = 0,1056$$

olarak hesaplanır.

Tesadüfen seçilecek bir adayın 80 puandan yüksek puan almış olma olasılığı ise, benzer mantıkla,

$$0,5 - 0,4525 = 0,0475$$

olarak hesaplanmaktadır.

Son aşamada da, tesadüfen seçilecek bir adayın 45’den az ya da 80’den çok puan almış olma olasılığını bulmak için yukarıda hesapladığımız iki kuyruk bölgesi olasılığını toplamamız gerekecektir. Dolayısıyla,

$$0,1056 + 0,0475 = 0,1531$$

sonucuna ulaşılmaktadır. Bu sonuca göre, sınava giren adaylar arasından tesadüfen seçilecek herhangi birinin 45 puandan düşük ya da 80 puandan yüksek puan almış olma olasılığı % 15,31’dir. Başka bir deyişle, sınava giren adayların % 15,31’i 45 puandan az ya da 80 puandan yüksek puan almıştır.

Yukarıdaki örnekte, normal dağılımla ilgili olarak karşılaşılabilecek tüm olasılık hesabı versiyonlarına yönelik çözümlere yer verilmiştir.

Örnek:

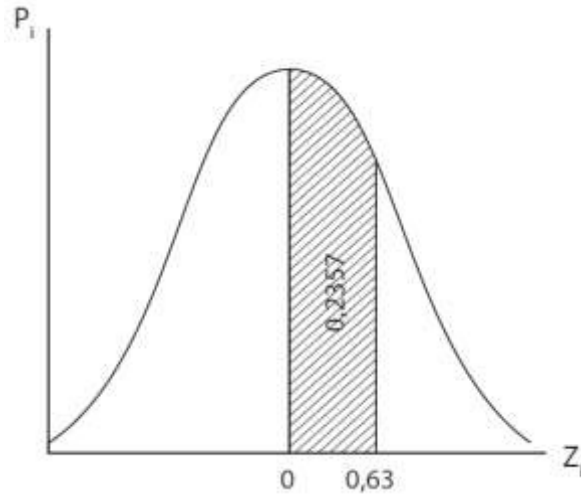
Bir margarin fabrikasında üretilmekte olan margarinlerin ağırlıklarının ortalaması 250 gr ($\mu = 250$ gr) ve standart sapması 8 gr ($\sigma = 8$ gr) olarak normal dağıldığı bilinmektedir. Günlük üretim içinden tesadüfen seçilecek bir margarin paketinin,

- a- 250-255 gr arasında bulunma olasılığını hesaplayınız.
- b- 255 gr'dan fazla olması olasılığını hesaplayınız.
- c- 255 gr'dan az olması olasılığını hesaplayınız.

a-Tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını hesaplayabilmek için öncelikle, 255 değerine karşılık gelen standart z değerini hesaplamalıyız.

$$z_{255} = \frac{X_i - \mu}{\sigma} = \frac{255 - 250}{8} = +0,625 \cong 0,63$$

Standart normal eğri alanları tablosundan, $z = 0,63$ değerinin alanını (ilk sütunda 0,6 ve ilk satırda 0,03 değerlerini bulup, satır ve sütunlarının kesişim noktasına bakarak) 0,2357 olarak belirliyoruz. Bu alan ortalama ile ilgili değer arasındaki alanı verdiğinden, tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını % 23,57 olarak hesaplamış oluyoruz.

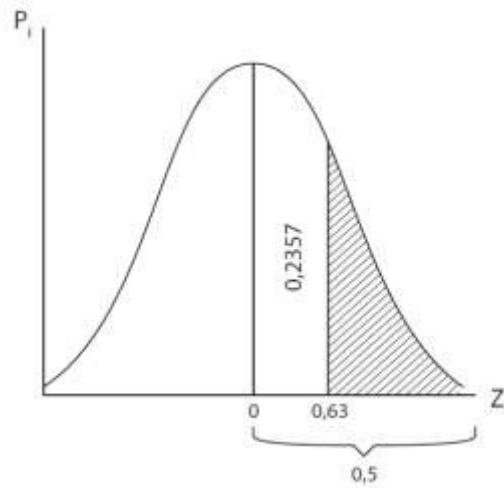


Şekil 12.15: $P(0 \leq z \leq 0,63)$

b-Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olma olasılığını bulmak için de, ortalamanın sağında kalan bütün alan 0,5 olduğundan istenen olasılığı,

$$0,5 - 0,2357 = 0,2643$$

hesaplıyoruz. Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olması olasılığı % 26,43'tür. Başka bir deyişle, bu fabrikada üretilmekte olan margarinlerin % 26,43'ü 255 gramın üzerinde ağırlığa sahiptir.

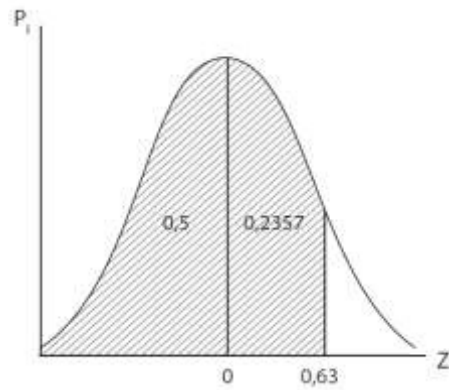


Şekil 12.16: $P(z > 0,63)$

c-Tesadüfen seçilecek bir margarinin 255 gramdan az olması olasılığını hesaplamak için de ortalamanın solunda kalan alanın 0,5 olduğunu hatırlayarak, bu değeri 250-255 gr aralığının olasılığı olan 0,2357 değeri ile toplamamız gerekecektir:

$$0,5 + 0,2357 = 0,7357$$

Dolayısıyla, tesadüfen seçilecek bir margarinin 255 gramdan az ağırlığa sahip olması olasılığı % 73,57 olmaktadır. Başka bir deyişle bu fabrikada üretilen margarinlerin % 73,57'si 255 gramdan az ağırlığa sahiptir.



Şekil 12.17: $P(z < 0,63)$

Şekil 6.15, Şekil 12.16 ve Şekil 12.17'den yukarıdaki örneğin çözümü için oluşturulmuş z değerlerini ve olasılıkları görebiliriz.

12.5. Normal Dağılım Olasılıklarını Kullanarak Beklenen Değer Hesabı

Normal dağılıma sahip tesadüfi değişkenlerin belirli bir aralıkta değer alma olasılıklarını hesaplamayı öğrendik. Bunun yanında, kitlenin içinde kaç birimin ilgili aralıkta değer alacağını

da belirleyebiliriz. Başka bir deyişle, hesapladığımız olasılık değeri üzerinden, ilgili aralıkta yer alacak birimlerin frekansının ne olacağını bulabiliriz.

Olasılık bölümü içinde öğrendiğimiz beklenen değer kavramını hatırlayalım. Beklenen değer, gözlem sayısı ile olasılığın çarpımıdır:

$$E(X) = n.p$$

Dolayısıyla, normal dağılım çözümlemelerinde de hesaplanan olasılık toplam frekans ile çarpılarak ilgili olasılık alanına düşen gözlem sayısı ya da birim sayısı kolaylıkla hesaplanabilir

Şimdi yukarıda çözümünü yaptığımız margarin fabrikası örneğinde bir günde toplam 1000 adet margarin üretimi yapıldığını varsayalım.

Tesadüfen seçilecek bir margarinin 250-255 gram arasında olması olasılığını 0,2357 olarak hesaplamıştık. Bu durumda, günlük üretim içinde 250-255 gram aralığında ağırlığa sahip margarin sayısı da,

$$E(X) = n.p$$

$$E(X) = 1000 \cdot 0,2357 = 235,7 \cong 236$$

olarak hesaplanacaktır.

Benzer şekilde, kaç adet margarinin 255 gramdan fazla olacağını da belirleyebiliriz. Bunun için, tesadüfen seçilecek bir margarinin 255 gramdan fazla olması olasılığı olan 0,2643 değerini toplam frekans ile çarpmamız yeterli olacaktır.

$$E(X) = n.p$$

$$E(X) = 1000 \cdot 0,2643 = 264,3 \cong 264$$

Dolayısıyla bir gün içinde üretilen 1000 adet margarinin 264 tanesinin 255 gramdan daha fazla ağırlığa sahip olmasını bekleyebiliriz.

Örnek:

Bir konserve fabrikasında üretilmekte olan konserveelerin ağırlığı normal dağılıyor. Ortalama ağırlık 500 gr ve standart sapma 10 gr olduğuna göre tesadüfen seçilecek bir konserve kutusunun,

a- 525 gramdan fazla olması olasılığı nedir?

b- Günde 5000 adet konserve üretilmesi durumunda kaç konserveenin 525 gramdan fazla olması beklenir?

a- Tesadüfen seçilecek bir konserve kutusunun 525 gramdan fazla olması olasılığını hesaplayabilmek için öncelikle 525 değerine karşılık gelen z değerini hesaplıyoruz.

$$z_{525} = \frac{X_i - \mu}{\sigma} = \frac{525 - 500}{10} = +2,5$$

Standart normal eğri alanları tablosunda $z=2,5$ standart değerine karşılık gelen alan, tablonun ilk sütununda bulunan 2,5 değeri ile ilk satırdaki 0,00 değerlerinin kesişim noktasında yer alan 0,4938 olasılığını vermektedir.=

Bu durumda, tesadüfen seçilecek bir kutunun 525 gramdan fazla olması olasılığı,

$$0,5 - 0,4938 = 0,0062$$

olarak hesaplanmaktadır. Tesadüfen seçilecek bir konserve kutusunun 525 gramdan fazla olması olasılığı % 0,62 olmaktadır.

b-Günde 5000 konserve üretilmesi halinde,

$$E(X) = n.p = 5000 \cdot 0,0062$$

$$E(X) = 31 \text{ adet}$$

konserve kutusunun 525 gramdan fazla olması beklenebilir.

Bu Bölümde Neler Öğrendik?

Bu bölümde sürekli olasılık dağılımları içinde çok önemli bir dağılım olan ve istatistik yöntem ve uygulamalarında çokça kullanılan normal dağılım konusunu öğrendik.

Normal dağılım, tabiatta gözlenen pek çok değişkenin gösterdiği olasılık dağılımı olması sebebiyle, normal dağılım olarak nitelendirilen bir dağılımdır.

Normal dağılım, simetrik bir dağılımdır, aritmetik ortalama ve standart sapma olmak üzere iki niteleyici değeri bulunmaktadır. Aritmetik ortalaması ve standart sapması farklı sonsuz sayıda normal dağılım örneği söz konusudur.

Normal dağılımın sürekli bir dağılım olması ve olasılık hesabının eğri altında kalan alan ile ölçülmesi sebebiyle, olasılık hesabı normal dağılım fonksiyonunda integral alma işlemi ile yapılır.

Ortalaması ve standart sapması farklı olan her normal dağılım için uzun uzun integral hesabı yapmak yerine, normal dağılım standart normal dağılıma dönüştürülmekte ve standart normal dağılım için hazırlanmış olasılık tablosu kullanılarak aranan olasılık değerlerine kolaylıkla ulaşılabilmektedir.

Normal dağılım, kısaca z dönüşümü olarak da tanımlanan standart değer dönüşümü ile standart normal hale dönüştürülebilmekte ve böylelikle, standart normal dağılım için hazırlanmış olasılık dağılım tablosu kullanılarak hesaplanmak istenen olasılık değerleri tablodan kolaylıkla bulunabilmektedir.

Sonuç olarak bu bölümde, normal dağılımı, özelliklerini ve normal dağılımda olasılık hesabının nasıl yapıldığını öğrendik.

Bölüm Soruları

- 1- Ortalaması 75 ve standart sapması 7 olan bir dağılımda 80'e karşılık gelen z değeri nedir?
 - a- 0,85
 - b- 0,71
 - c- -0,71
 - d- 0,65
 - e- -0,65
- 2- Ortalaması 40 standart sapması 6 olan bir dağılımda 38'e karşılık gelen z değeri nedir?
 - a- 0,33
 - b- -0,38
 - c- -0,33
 - d- -0,45
 - e- 0,38
- 3- Normal eğrinin altında kalan toplam alan dir.
 - a- 1'dir.
 - b- -0,5'dir.
 - c- 0,5'dir.
 - d- 0,75'dir.
 - e- Değişmektedir.
- 4- Normal eğri alanları tablosunda $z=1,35$ 'e karşılık gelen olasılık değeri aşağıdakilerden hangisidir?
 - a- 0,3515
 - b- 0,4225
 - c- 0,4125
 - d- 0,4115
 - e- 0,4200

- 5- Normal eğri alanları tablosunda, $z= 2,40$ 'a karşılık gelen olasılık değeri nedir?
- a- 0,4918
 - b- 0,4885
 - c- 0,4225
 - d- 0,4865
 - e- 0,4985
- 6- Normal eğri alanları tablosuna göre, $P(-1,25 \leq z \leq 1,20)$ olasılığı nedir?
- a- 0,7445
 - b- 0,7793
 - c- 0,7558
 - d- 0,7642
 - e- 0,7456
- 7- Normal eğri alanları tablosuna göre $P(-2,00 \leq z \leq 0)$ alanının olasılığı nedir?
- a- 0,4552
 - b- 0,4435
 - c- 0,4772
 - d- 0,4886
 - e- 0,4663
- 8- Bir okulda öğrencilere bir deneme testi uygulanıyor. Ortalama puan 78 ve standart sapma 20 olduğuna göre, tesadüfen seçilecek bir öğrencinin 95'den fazla puan alması olasılığı nedir?
- a- 0,1977
 - b- 0,2125
 - c- 0,85
 - d- 0,3023
 - e- 0,8023

- 9- İstanbul – Konya arasındaki uçuş süresinin normal dağıldığı bilinmektedir. Daha önce yapılan gözlemler sonucunda ortalama uçuş süresi 2700 saniye ve standart sapmada 300 saniye hesaplanmıştır. Buna göre, herhangi bir uçuş süresinin 3300 saniyeden fazla olması olasılığı nedir?
- a- 0,0554
 - b- 0,4772
 - c- 0,4335
 - d- 0,0228
 - e- 0,0115
- 10- Çimento üreten bir firmanın ürettiği çimento torbalarının ağırlıkları normal dağılıyor. Ortalama torba ağırlığı 50 kg ve standart sapma 0,5 kg hesaplanıyor. Torba ağırlığının 49 kg'dan az olması halinde çimento paketi ağırlık açısından kalitesiz olarak ayrılmaktadır. Gün içinde 1000 paket çimento üretildiğinde kaç paket çimentonun kalitesiz olduğu söylenebilir?
- a- 22,8
 - b- 47,72
 - c- 15,6
 - d- 18,7
 - e- 21,4

Cevaplar

- 1- b
- 2- c
- 3- a
- 4- d
- 5- a
- 6- b
- 7- c
- 8- a
- 9- d
- 10- a

Yararlanılan Kaynaklar

ALOPA, K. B. (2003), *İstatistiksel Analiz Metodları*, 6. Baskı, İstanbul: Çağlayan Kitabevi.

BÜLBÜL, Ş. (2006), *Tanımlayıcı İstatistik*, İstanbul: Der Yayınları.

ÇUBUKÇU, K. M. (2015), *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*, Ankara: Nobel Yayınevi.

GÜLER, F. (2012), *İstatistik Metotları ve Uygulamaları*, Üçüncü baskı, İstanbul: Beta Yayıncılık.

GÜRSAKAL, N. (2015), *Betimsel İstatistik*, Sekizinci Baskı, Bursa: Dora Yayıncılık.

GÜRTAN, K. (1979), *İstatistik ve Araştırma Metodları*, Gözden geçirilmiş ve genişletilmiş yeni baskı, İstanbul, İstanbul Üniversitesi Yayınları No:96, 1979.

HAIR, J. F., BLACK C. W., BARRY J. B., ANDERSON R. E. (2010) *Multivariate Data Analysis*, Seventh Edition, USA: Pearson Education Inc.

İPEK, M. (1988), *Betimsel İstatistik*, İstanbul: Beta Yayıncılık.

NEWBOLD, P., CARLSON W. L., THORNE B. M., (2012), *Statistics for Business and Economics*, Eighth Edition, USA: Pearson Education Inc.

ORHUNBİLGE, N., (2014), *Tanımsal İstatistik Olasılık ve Olasılık Dağılımları*, Gözden geçirilmiş ikinci baskı, Ankara: Nobel Akademik Yayıncılık.

ÖZSOY, O., (2010), *İktisatçılar ve İşletmeciler için İstatistik*, Ankara, Siyasal Kitabevi.

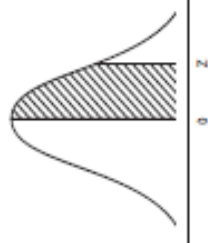
SERPER, Ö. (2014), *Uygulamalı İstatistik*, Yedinci Baskı, Bursa: Ezgi Kitabevi.

SPIEGEL, M. R. (1995), *İstatistik*, Çeviren: Ayaydın, A. vd., İstanbul: Bilim Teknik Yayınevi.

STEVENS, S.S. (1946), “On the Theory of Scales of Measurement” , *Sciences*, New Series, Vol. 103, No. 2684, ss.677-680.

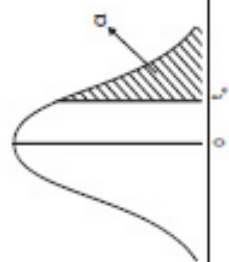
TURANLI, M., GÜRİŞ S., (2005), *Temel İstatistik*, Genişletilmiş İkinci Baskı, İstanbul: Der Yayınları.

Standart Normal Dağılım Tablosu



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2133	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

T Dağılımı Tablosu



v	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
∞	1.282	1.645	1.960	2.326	2.576

13. BASİT İNDEKSLER

Bölüm Yazarları:

DR. ÖĞR.ÜYESİ HAKAN BEKTAŞ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde oldukça geniş bir uygulama alanına sahip olan indeksler ele alınacaktır. Ekonomi alanında öne çıkan birçok gösterge indeks olarak sunulmaktadır. Bunlara örnek olarak tüketici fiyat indeksi, sanayi üretim indeksi, konut satış indeksi verilebilmektedir. Ancak indekslerin sadece ekonomi alanında kullanılmadığını belirtmek isteriz. Bu bölümde indeks kavramını tanımladıktan sonra, hesaplama yöntemi üzerinde duracağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) İndeks nedir?
- 2) Gündelik hayatta takip ettiğimiz indeksler nelerdir?
- 3) İndekslerin sınıflandırılması hangi kriterlere göre yapılmaktadır?
- 4) Basit indeks nedir?
- 5) İndeks hesabında esas alınan döneme göre indeks türleri nelerdir?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İndeks kavramı	Bir değişkenin ya da değişkenlerin değerlerinde zaman ortaya çıkan değişmelerin oransal ölçüsü olan indeks kavramını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
İndeks türleri	İndekslerin hangi kriterlere göre sınıflandırıldığını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Basit indeks	Basit indeks kavramı ve nasıl hesaplandığını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Esas alınan dönem açısından indeks türleri	Esas alınan dönem açısından indeks türlerini ve nasıl hesaplandığını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Sabit esaslı indeksin esas (baz) yılının değiştirilmesi	Sabit esaslı indekste esas yılın nasıl değiştirildiğini bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü

Anahtar Kavramlar

- İndeks: Bir değişkenin ya da değişkenlerin değerlerinde zaman ortaya çıkan değişmelerin oransal ölçüsü
- Esas (baz) yılı: Karşılaştırmada esas (temel, referans) alınan yılı göstermektedir.
- Basit indeks: İndeks hesabında tek bir maddenin ele alınması durumunda hesaplanan indeks türüdür.

Giriş

İndeksler, geniş bir uygulama alanına sahip bir kavram olarak karşımıza çıkmaktadır. Ekonomi alanında öne çıkan göstergelerin birçoğu, indeks olarak raporlanmaktadır. Bunlara; tüketici fiyat indeksi, sanayi üretim indeksi, perakende satış indeksleri, işgücü maliyet indeksi ve inşaat üretim indeksi örnek olarak verilebilmektedir. Yazılı ve görsel basını takip edenler, bahsedilen indeksler ve daha fazlasına ilişkin haberleri belirli periyotlarla görmektedir. Fakat indekslerin kullanımı ekonomi alanıyla sınırlı değildir. Mutluluk indeksi, yaşam kalitesi indeksi, insani yoksulluk indeksi, insani gelişme indeksi ve sağlıklı yeme indeksi gibi sosyal alanlarda da indeksler hesaplanmaktadır. Dolayısıyla yoksulluk ve mutluluk gibi karmaşık kavramlar sayısallaştırılarak, mekânsal veya zamansal değişimler takip edilebilmektedir.

İndeks, en basit şekliyle değişimin ölçüsü olarak tanımlanabilmektedir. Bu tanıımı somut hale getirebilmek için Tablo 13.1’de yer alan veriyi inceleyelim. Tablo 13,1’de 2012 ile 2014 yılları arasında İstanbul için kişi başına düşen Gayrisafi Yurtiçi Hâsıla (GSYİH) değerleri verilmiştir. Bu tablonun ikinci sütununda Türk Lirası cinsinden değerler, üçüncü sütununda ise 2012 yılının değerine göre ölçeklendirilmiş değerler bulunmaktadır. Üçüncü sütunda yer alan değerler aracılığıyla, GSYİH’nin zamana göre değişimi gözlenebilmektedir. Bu aşamada öncelikle karşılaştırmada baz alınacak yılı seçmemiz gerekmektedir. Bu örnekte baz yılı olarak 2012 alınmış ve 100’e eşitlenmiştir. Böylelikle, hesaplanan değerler indeks olarak adlandırılmaktadır (Ralph vd. , 2015).

Tablo 13.1: Kişi Başına Düşen Gayrisafi Yurtiçi Hâsıla (İstanbul)		
	Kişi Başına Gayrisafi Yurtiçi Hâsıla (Türk Lirası)	2012 yılındaki Gayrisafi Yurtiçi Hâsıla Değerine Göre (2012=100)
2012	34637	100,00
2013	39468	113,95
2014	43645	126,00

Kaynak: Türkiye İstatistik Kurumu (<http://www.tuik.gov.tr>)

Tablo 13.1 incelendiğinde; 2012 yılı değerinin baz yılı olarak kabul edildiği anlaşılmaktadır. **Baz yılı**, karşılaştırmada esas (temel) alınan yılı göstermektedir. Değişimini görmek istediğimiz değer ise, **cari değer** olarak adlandırılmaktadır. Dolayısıyla indekste; biri cari değer diğeri ise baz yılı ait değer olmak üzere iki değer bulunmaktadır. Bu değerler oranlanıp 100 ile çarpılmaktadır. Bu tablodan hareketle; 2013 yılında kişi başına GSYİH 2012 yılına göre %13,95 oranında, 2014 yılında kişi başına GSYİH 2012 yılına göre %26,00 oranında artış kaydedildiği anlaşılmaktadır.

Bir değişkenin ya da değişkenlerin değerlerinde zaman içinde veya mekânlar arasında ortaya çıkan değişmelerin oransal ölçüsü olarak tanımlanabilen indeks, iki değer birbirine oranlanmasıyla elde edilmektedir. Böylelikle; verinin taşıdığı bilgi okuyucular için

basitleştirilmekte ve karşılaştırmalara olanak tanınmaktadır. Buna örnek olarak, her ayın ilk haftasında Türkiye İstatistik Kurumu tarafından açıklanan tüketici fiyat indeksi verilebilmektedir. Hane halklarının tüketimine yönelik mal ve hizmet fiyatlarının zaman içindeki değişimi ölçülmektedir. Yani, piyasada tüketime konu olan mal ve hizmetlerin fiyatlarında değişim ölçülerek enflasyon oranı hesaplanmaktadır. Dolayısıyla kamuoyu için mal sepetinin zaman içindeki fiyat değişimi anlaşılabilir ve karşılaştırılabilir şekilde sunulmaktadır.

Bu bölümde sadece zaman içindeki fiyat değişimlerine ilişkin indeks hesabına yer verilecektir.

13.1. İndeks Türleri

Zaman serilerine dayanan indekslere **zaman indeksi** adı verilmektedir. Bu indekslerin hesaplanmasında **kapsanan madde (mal ve hizmet) sayısı**, **esas alınan dönem** ve **maddelerin önem (ağırlık) derecesi** dikkate alınarak sınıflandırma yapılmakta olup; Tablo 13.2’de sunulmuştur.

Tablo 13.2: İndekslerin Sınıflandırılması	
Madde Sayısına Göre İndeksler	Basit İndeksler
	Bileşik İndeksler
Esas Alınan Döneme Göre İndeksler	Sabit Esaslı İndeksler
	Değişik Esaslı İndeksler
Maddelerin Öneme Göre İndeksler	Tartısız İndeksler
	Tartılı İndeksler

Tablo 13.2 incelendiğinde; indekslerin üç açıdan sınıflandırıldığı görülmektedir. İndeks hesabında tek bir madde ele alınıyorsa **basit indeks**, birden fazla madde için hesaplanıyorsa **bileşik indeks** olarak adlandırılmaktadır. Zaman indeksleri hesaplanırken esas alınan dönem sabit tutulabilir ya da değiştirilebilir. Eğer indeksin hesaplamasında bir dönem baz olarak alınıyorsa **sabit esaslı indeks**, sabit bir dönem yerine hesaplanacak dönem değerinin bir önceki dönem değerine oranlanmasıyla elde ediliyorsa **değişik esaslı indeks (zincirleme indeks)** olarak isimlendirilmektedir. Son olarak ise, her bir maddenin eşit öneme sahip olması durumunda **tartısız indeks** ve maddelerin farklı önem derecelerine sahip olması durumunda ise **tartılı indeks** olmak üzere iki sınıfa ayrılmaktadır. Tartısız ve tartılı indeks, birden fazla maddenin bulunduğu bileşik indeksler için hesaplanabilir. Bileşik indeksler, 14.bölümde ele alınacaktır.

13.2. Basit İndeks

Tek bir madde için indeks hesaplanması halinde basit indeks söz konusu olmaktadır. Örneğin, et fiyatı indeksi, tiyatro bileti fiyatı indeksi, benzin litre fiyatı indeksi vb. indeksler basit indekslerdendir.

13.2.1. Sabit Esaslı Basit İndeks

Zaman serisindeki belirli bir dönemin sabit tutulup; serinin değerlerinin sabit dönem (baz) değerine oranlanmasıyla **sabit esaslı indeks** değerleri hesaplanmaktadır. Eğer bu hesap, tek bir madde için yapılırsa; **sabit esaslı basit indeks** elde edilmektedir.

Sabit esaslı basit indeksin (SEİ) formülüne Eşitlik 13.1’de yer verilmiştir.

$$SEİ_i = \frac{F_i}{F_0} \times 100 \quad (13.1)$$

Eşitlik 13.1’deki F_i , indeks değeri hesaplanacak dönemdeki fiyatı; F_0 ise esas kabul edilen dönemdeki fiyatı göstermektedir. Yani, bu iki değer birbirine oranlandıktan sonra 100 ile çarpıldığında indeks değerine ulaşılmaktadır. Bu indeks ile ilgili bir örnek çözelim.

Örnek: Aşağıdaki tabloda 2014-2016 yılları arasında İstanbul’daki ekmek fiyatları bulunmaktadır. 2014 yılını esas dönem kabul etmek suretiyle sabit esaslı basit indeks değerlerini hesaplayalım.

Yıl	Ekmek Fiyatı (lira)
2014	0,85
2015	1,00
2016	1,25

Tabloda yer alan fiyat bilgilerinden hareketle; sabit esaslı basit indeks formülünü uygulayalım.

Yıl	Ekmek fiyatları için sabit esaslı basit indeks (2014=100)
2014	$(0,85 / 0,85) \times 100 = 100$
2015	$(1,00 / 0,85) \times 100 = 117,65$
2016	$(1,25 / 0,85) \times 100 = 147,06$

Ekmek fiyatları için hesapladığımız sabit esaslı indeks serisinde; her yıl gerçekleşen ekmek fiyatını, esas dönem olarak kabul ettiğimiz 2014 yılı ekmek fiyatına bölüp, 100 ile çarptık. Böylelikle sırasıyla 117,65 ve 147,06 değerlerini elde ettik. Peki, bu değerleri nasıl yorumlayacağız.

2015 yılında ekmek fiyatının 2014 yılına göre %17,65 oranında,

2016 yılında ekmek fiyatının 2014 yılına göre %47,06 oranında artış gösterdiğini söyleyebiliriz. Bu durumu şu şekilde de yorumlayabiliriz.

2014 yılında 100 lira ödenerek satın alınan bir adet ekmek 2015 yılında 117,65 lira ödenerek,

2014 yılında 100 lira ödenerek satın alınan bir adet ekmek 2016 yılında 147,06 lira ödenerek alınabilmektedir.

Sabit esaslı indeksin hesaplanmasında esas (baz) dönemin seçimi önem arz etmektedir. Esas dönem, indeksin hesaplanma amacına uygun olarak belirlenmesi gerekmektedir. Bu süreçte, fiyatların çok yüksek ya da çok düşük seyrettiği kriz ya da refah dönemlerinin seçilmemesi önerilmektedir. Ayrıca esas kabul edilecek dönemin, çok eski bir yıl olmamasına da dikkat edilmesi gerekmektedir.

13.2.2. Değişik Esaslı Basit İndeks

Esas döneme göre değil de sürekli bir önceki döneme göre indeks değerleri hesaplanabilmektedir. Bu durumda, değişik esaslı indeks değerleri hesaplanmış olmaktadır.

Değişik esaslı basit indeksin (DEİ) formülüne Eşitlik 5.2’de yer verilmiştir.

$$DEİ_i = \frac{F_i}{F_{i-1}} \times 100 \quad (13.2)$$

Eşitlik 13.2’deki F_i , indeks değeri hesaplanacak dönemdeki fiyatı; F_{i-1} ise bir dönem fiyatı göstermektedir. Yani, bu iki değer birbirine oranlandıktan sonra 100 ile çarpıldığında indeks değerine ulaşılmaktadır. Bu indeks hesabına örnek olmasın adına yine ekmek fiyatlarını ele alalım.

Örnek: Aşağıdaki tabloda 2014-2016 yılları arasında İstanbul’daki ekmek fiyatları bulunmaktadır. Değişik esaslı basit indeks değerlerini hesaplayalım.

Yıl	Ekmek Fiyatı (lira)
2014	0,85
2015	1,00
2016	1,25

Tabloda yer alan fiyat bilgilerinden hareketle; değişik esaslı basit indeks formülünü uygulayalım.

Yıl	Ekmek fiyatları için sabit esaslı basit indeks (2014=100)
2014	$(0,85 / 0,85) \times 100 = 100$
2015	$(1,00 / 0,85) \times 100 = 117,65$
2016	$(1,25 / 1,00) \times 100 = 125,00$

Ekmek fiyatları için hesapladığımız değişik esaslı indeks serisinde; her yıl gerçekleşen ekmek fiyatını, bir önceki dönem fiyatına bölüp, 100 ile çarptık. Böylelikle sırasıyla 117,65 ve 125,00 değerlerini elde ettik. Peki, bu değerleri nasıl yorumlayacağız.

2015 yılında ekmek fiyatının 2014 yılına göre %17,65 oranında,

2016 yılında ekmek fiyatının 2015 yılına göre %25 oranında artış gösterdiğini söyleyebiliriz.

Değişik esaslı indeks serisinde yer alan değerleri şu şekilde de yorumlayabiliriz.

2014 yılında 100 lira ödenerek alınan bir adet ekmek 2015 yılında 117,65 lira ödenerek,

2015 yılında 100 lira ödenerek alınan bir adet ekmek 2016 yılında 125 liraya alınabilmektedir.

Sabit esaslı indeksten değişik esaslı indekse, değişik esaslı indeksten sabit esaslı indekse geçiş mümkündür. Bu işlemin nasıl gerçekleştirileceğine ilişkin yöntem ise bir sonraki alt başlıkta irdelenmiştir.

13.2.3. Sabit Esaslı İndeksten Değişik Esaslı İndekse Geçiş

Bir zaman serisine ilişkin sabit esaslı indeks değerleri kullanılarak; değişik esaslı indeks değerlerine geçilebilmektedir. Bu aşamada herhangi bir dönemin sabit esaslı indeks değeri bir önceki dönemin sabit esaslı indeks değerine bölünüp; 100 ile çarpılmaktadır. Değişik esaslı indeks değerlerine ulaşabilmek için Eşitlik 13.3'teki formül kullanılmaktadır.

$$DEİ_i = \frac{SEİ_i}{SEİ_{i-1}} \times 100 \quad (13.3)$$

Eşitlik 13.3'te yer alan formüldeki $SEİ_i$ i. dönemdeki sabit esaslı indeks değerini, $SEİ_{i-1}$ ise i-1. dönemdeki sabit esaslı indeks değerini göstermektedir.

2014-2016 yılları arasındaki ekmek fiyatlarından hareketle hesapladığımız sabit esaslı indeks değerlerini kullanarak değişik esaslı indeks değerine geçmeye çalışalım. Hatırlanacağı üzere; sabit esaslı indeks değerleri sırasıyla 117,65 ve 147,06'dır.

$$DEİ_{2015} = \frac{SEİ_{2015}}{SEİ_{2014}} \times 100 = \frac{117,65}{100} \times 100 = 117,65$$

$$DEİ_{2016} = \frac{SEİ_{2016}}{SEİ_{2015}} \times 100 = \frac{147,06}{117,65} \times 100 = 125,00$$

Eşitlik 13.3'te yer alan formül kullanılarak; ekmek fiyatları için hesaplanan sabit esaslı indeks (2014=100) değerlerinden değişik esaslı indeks değerlerine geçilmiştir.

13.2.4. Değişik Esaslı İndeksten Sabit Esaslı İndekse Geçiş

Herhangi bir dönemin değişik esaslı indeks değerlerinden başlanarak geriye (geçmişe) doğru birbirini izleyen indeks değerleri çarpılmaktadır. Ardından ise bulunan değer, 100 ile çarpılır.

2014-2016 yılları arasındaki ekmek fiyatlarından hareketle hesapladığımız değişik esaslı indeks değerlerini kullanarak sabit esaslı indeks (2014=100) değerine geçmeye çalışalım. Hatırlanacağı üzere; sabit esaslı indeks değerleri sırasıyla 117,65 ve 125,00'dir.

$$SEİ_{2015} = \frac{117,65}{100} \times \frac{100}{100} \times 100 = 117,65$$

$$SEİ_{2016} = \frac{125,00}{100} \times \frac{117,65}{100} \times \frac{100}{100} \times 100 = 147,06$$

Değişik esaslı indeks değerlerinden sabit esaslı indeks değerlerine geçilmiş olup; bulunan değerlerin önceki kısımda bulduğumuz sabit esaslı indeks değerleriyle aynı olduğu görülmektedir.

13.2.5. Sabit Esaslı İndeks Serisinde Esas Dönemin Değiştirilmesi

Herhangi bir dönemin esas alınmasıyla hesaplanan sabit esaslı indekste temel (referans, baz) yıl değiştirilebilmektedir. Esas dönem eskidikçe sabit esaslı indeks değerleri büyümekte ve kıyaslama yapma imkânı zorlaşmaktadır. Bunun yanı sıra farklı esas dönemlere sahip indeksleri karşılaştırmak için temel (referans) yılın değiştirilmesine ihtiyaç duyulmaktadır. Bu aşamada kullanılacak formül, Eşitlik 13.4'te verilmiştir.

$$SEİ_{i,y} = \frac{SEİ_i}{SEİ_{0,y}} \times 100 \quad (5.4)$$

Eşitlik 13.4'te $SEİ_{i,y}$: yeni esas döneme göre hesaplanacak indeksi, $SEİ_i$: sabit esaslı indeksi değerini, $SEİ_{0,y}$: yeni esas dönemin indeks değerini göstermektedir.

2014-2016 yılları için 2014=100 esas dönemli indeks değerlerinden hareketle; esas dönem 2016 yılına kaydırılmıştır. Bu aşamada, Eşitlik 13.4'teki formül kullanılmıştır.

Yıl	Sabit esaslı basit indeks (2014=100)	Sabit esaslı basit indeks (2016=100)
2014	100	$(100,00 / 147,06) \times 100 = \mathbf{68}$
2015	117,65	$(117,65 / 147,06) \times 100 = \mathbf{80}$
2016	147,06	$(147,06 / 147,06) \times 100 = \mathbf{100}$

2014=100 esas dönemli indeks değerlerinden hareketle 2016 yılını baz alan indeks değerleri hesaplanmış ancak sağlamasını yapmak adına ekmek fiyatlarının üzerinden tekrardan 2016=100 için indeks değerleri bulunmuştur.

Yıl	Ekmek Fiyatı (adet/lira)	Sabit esaslı basit indeks (2016=100)
2014	0,85	$(0,85 / 1,25) \times 100 = 68$
2015	1,00	$(1,00 / 1,25) \times 100 = 80$
2016	1,25	$(1,25 / 1,25) \times 100 = 100$

Gerek 2014 yılını baz alan indeks değerlerinden hareketle hesaplanan gerek ekmek fiyatları üzerinden hesaplanan 2016=100 indeks değerlerinin aynı olduğu gözlenmiştir.

Bu Bölümde Neler Öğrendik?

İndeksleri incelediğimiz bu bölümde, öncelikle indeks kavramı üzerinde durarak, indekslerin kullanım amacını ele aldık ve kullanıldığı alanları öğrendik.

En basit şekliyle değişimin ölçüsü olarak tanımlanabilen indeks kavramı anlatıldıktan sonra indeks türleri üzerinde durduk. Bu bağlamda zaman serilerine dayanan zaman indeksine ait türlerden bahsettik. Kapsanan madde (mal ve hizmet) sayısına göre ve esas alınan döneme göre indeks türlerini öğrendik.

Bölüm Soruları

- 1-, bir değişkenin ya da değişkenlerin değerlerinde zaman içinde veya mekânlar arasında ortaya çıkan değişmelerin oransal ölçüsü olarak tanımlanabilmektedir.
 - a- Baz yıl
 - b- Cari değer
 - c- İndeks
 - d- Reel değer
 - e- Madde sepeti
- 2-, karşılaştırmada esas (temel) alınan yılı göstermektedir.
 - a- Baz yıl
 - b- Cari değer
 - c- İndeks
 - d- Beklenen değer
 - e- Cari yıl
- 3- Aşağıdakilerden hangisi esas alınan döneme göre indeks türlerinden biridir?
 - a- Basit indeksler
 - b- Sabit esaslı indeksler
 - c- Bileşik indeksler
 - d- Tartılı indeksler
 - e- Tartısız indeksler

4- Aşağıdakilerden hangisi madde sayısına göre indeks türlerinden değildir?

- I- Basit İndeksler
- II- Bileşik İndeksler
- III- Değişik Esaslı İndeksler
- a- Yalnızca I
- b- Yalnızca II
- c- Yalnızca III
- d- I - II
- e- I – II – III

5- Aşağıdakilerden hangisi esas alınan döneme göre indeks türlerindendir?

- I- Basit İndeksler
- II- Bileşik İndeksler
- III- Değişik Esaslı İndeksler
- a- Yalnızca I
- b- Yalnızca II
- c- Yalnızca III
- d- I - II
- e- I – II – III

6- Zaman serisindeki belirli bir dönemin sabit tutulup; serinin değerlerinin sabit dönem (baz) değerine oranlanmasıyla indeks değerleri hesaplanmaktadır.

- a- Tartılı
- b- Tartısız
- c- Sabit esaslı
- d- Değişik esaslı
- e- Basit

- 7- Aşağıdaki tabloda 2014-2016 yılları için İstanbul'daki su fiyatı (litre / lira) verilmiştir. Bu tablodan hareketle; 2015 yılı için sabit esaslı basit indeks değerini hesaplayınız. (Baz yılı, 2014'tür)

Yıl	Su Fiyatı (litre / lira)
2014	0,50
2015	0,75
2016	1,00

- a- 75
- b- 100
- c- 125
- d- 150
- e- 200

- 8- Aşağıdaki tabloda 2014-2016 yılları için X marka cep telefonun Türkiye satış fiyatları (adet / lira) verilmiştir. Bu tablodan hareketle; 2015 yılı için sabit esaslı basit indeks değerini hesaplayınız. (Baz yılı, 2015'tir)

Yıl	Su Fiyatı (litre / lira)
2014	1999
2015	2499
2016	2999

- a- 79,99
- b- 100
- c- 125,01
- d- 150
- e- 200

- 9- Aşağıdaki tabloda 2014-2016 yılları için İstanbul'daki su fiyatı (litre / lira) verilmiştir. Bu tablodan hareketle; 2015 yılı için değişik esaslı basit indeks değerini hesaplayınız.

Yıl	Su Fiyatı (litre / lira)
2014	0,50
2015	0,75
2016	1,00

- a- 75
- b- 100
- c- 125
- d- 150
- e- 200

- 10- Aşağıdaki tabloda 2014-2016 yılları için Ankara'daki ekmek fiyatlarından hareketle hesaplanmış sabit esaslı indeks değerleri yer almaktadır. Bu tabloyu kullanarak, 2014 yılı için değişik esaslı basit indeks değerini hesaplayınız. (Sabit esaslı basit indeks için baz yılı 2014'tür)

Yıl	Sabit Esaslı İndeks Değeri
2014	100
2015	120
2016	140

- a- 50
- b- 100
- c- 120
- d- 140
- e- 150

Cevaplar

1- c

2- a

3- b

4- c

5- c

6- c

7- d

8- b

9- d

10- b

Yararlanılan Kaynaklar

AKMAN ULUTÜRK, S. (2008), “2003 Temel Yılı Tüketici Fiyat İndeksine İlişkin Düşünceler”, *İstanbul Üniversitesi İktisat Fakültesi Maliye Araştırma Merkezi Konferansları*, 51. Seri.

CİLLOV, H., (1993), *İktisadi Olaylara Uygulanan İstatistik Metodları*, İstanbul: İÜ Basımevi ve Film Merkezi, 1993.

GÜLER, F., (2012), *İstatistik Metotları ve Uygulamaları*, 3.Basım, İstanbul: Beta Yayınları.

GÜRSAKAL, N., *Betimsel İstatistik*, 7. Baskı, Bursa: Dora Yayınları.

RALPH, J., O’NEILL, R., WINTON J., (2015), *A Practical Introduction to Index Numbers*, John Wiley & Sons Ltd.

SERPER, Ö., (2014), *Uygulamalı İstatistik*, 7.Baskı, Bursa: Ezgi Kitabevi.

TURANLI, M., GÜRİŞ, S., (2005), *Temel İstatistik*, Genişletilmiş 2.Basım, İstanbul: Der Yayınları.

14. BİLEŞİK İNDEKSLER

Bölüm Yazarları:

DR. ÖĞR.ÜYESİ HAKAN BEKTAŞ

Bu Bölümde Neler Öğreneceğiz?

Bu bölümde oldukça geniş bir uygulama alanına sahip olan bileşik indeksler ele alınacaktır. Ekonomi alanında öne çıkan birçok gösterge indeks olarak sunulmaktadır. Bunlara örnek olarak tüketici fiyat indeksi, sanayi üretim indeksi, konut satış indeksi verilebilmektedir. Ancak indekslerin sadece ekonomi alanında kullanılmadığını belirtmek isteriz. Bu bölümde birden fazla mal ve hizmet ele alındığında indeksin nasıl hesaplanması gerektiği üzerinde duracağız.

Bölüm Hakkında İlgi Oluşturan Sorular

- 1) Birden fazla mal ve hizmet söz konusu olduğunda indeks hesabı nasıl yapılmaktadır?
- 2) Bileşik indeks nedir?
- 3) Bileşik indekslerin türleri nelerdir?
- 4) Tartılı ve tartısız bileşik indekslere neden ihtiyaç duyulmaktadır?
- 5) Reel fiyat serisi nasıl hesaplanmaktadır?

Bu Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Bileşik indeks kavramı	Bileşik indeks kavramını ve gündelik hayatta bileşik indekslerin kullanım alanlarını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Bileşik indekslerin türleri	Bileşik indekslerin hangi kritere göre sınıflandırıldığını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Bileşik indekslerde tartı	Tartılı bileşik indekslere neden ihtiyaç duyulduğunu bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Tartılı bileşik indeksler	Tartılı bileşik indekslerin türlerini ve nasıl hesaplandığını bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü
Reelleştirme	Cari fiyat serinin deflate edilmesini bilir.	İstatistik kitaplarının içinde yer alan indeksler bölümü

Anahtar Kavramlar

- Bileşik İndeksler: Birden fazla mal ve hizmet için hesaplanan indeks türüdür.
- Tartı: Bileşik indeks hesabında ele alınan mal ve hizmetlerin farklı önem derecelerine sahip olduğunda kullanılan ağırlıklardır.
- Reel Fiyat Serisi: Cari fiyatlarla hesaplanmış serilerin enflasyon etkisinden arındırılmasıyla elde edilen seridir.

Giriş

On üçüncü bölümde sabit ve değişik esaslı basit indeks üzerinde durulmuştur. Bilindiği üzere, basit indeks hesabında tek bir madde bulunmaktadır. Ancak birden fazla mal ve hizmet söz konusuysa bu durumda basit indeksler yerine bileşik indekslerin hesaplanması gerekmektedir.

14.1. Bileşik İndeksler

Birden fazla madde söz konusu olduğunda ise **bileşik indeks** hesabı yapılmaktadır. Günlük hayatta birden çok mal ve hizmet tüketmekte olduğumuzdan, bir maddenin fiyatı yerine çok sayıda maddenin fiyatlarını takip ederiz. Dolayısıyla basit indekse göre bileşik indeksten daha çok yararlanırız. Bu maddelerin önem derecesi ya da ağırlığı bileşik indeks hesabında önem arz etmektedir. Eğer bütün maddeler eşit öneme sahip ise **tartısız bileşik indeks**, maddelerin önem dereceleri birbirinden farklıysa **tartılı bileşik indeks** hesaplanmaktadır. Tartısız bileşik indekste, takip edilen maddelerin fiyatlarının ortalamalarından yararlanılarak indeks değerine ulaşılmaktadır. Bu kısımda aritmetik ortalama kullanılarak bileşik indeks hesabı gerçekleştirilecektir.

On üçüncü bölümde basit indeks başlığı altında incelenen “sabit esaslı basit indeks” ve “değişik esaslı basit indeks” hesaplamalarından yararlanılarak, bu kısımda “sabit esaslı bileşik indeks” ve “değişik esaslı bileşik indeks” ele alınacaktır.

Tartısız bileşik indeks hesaplarında kullanmak üzere örnek bir madde sepeti oluşturalım.

Yıl	Ekmek Fiyatı (adet/lira)	Süt Fiyatı (litre/lira)	Su Fiyatı (litre/lira)
2014	0,85	2,00	0,75
2015	1,00	2,50	1,00
2016	1,25	2,90	1,25

Bileşik indeks hesabında üç maddemiz olduğunu varsayarak; öncelikle her bir madde için sabit esaslı basit indeks değerlerini hesaplayalım.

Yıl	Sabit esaslı basit indeks (2014=100) (Ekmek)	Sabit esaslı basit indeks (2014=100) (Süt)	Sabit esaslı basit indeks (2014=100) (Su)
2014	100,00	100,00	100,00
2015	117,65	125,00	133,33
2016	147,06	145,00	166,67

Her bir madde için sabit esaslı basit indeks değerini hesapladıktan sonra; değişik esaslı basit indeks değerlerini hesaplayalım.

Yıl	Değişik esaslı basit indeks (Ekmek)	Değişik esaslı basit indeks (Süt)	Değişik esaslı basit indeks (Su)
2014	100,00	100,00	100,00
2015	117,65	125,00	133,33
2016	125,00	116,00	125,00

14.1.1. Tartısız Bileşik İndeksler

Bu kısımda aritmetik ortalama temelli tartısız **sabit esaslı bileşik indeks** ve **değişik esaslı bileşik indeks** hesabına yer verilecektir.

Tartısız sabit esaslı bileşik indeks hesabında; sepette yer alan her bir madde için sabit esaslı basit indeks değerleri hesaplanır. Ardından her bir yıl için hesaplanan indeks değerlerinin aritmetik ortalaması bulunur. Böylelikle yıllar itibarıyla indeks değerlerine ulaşılmış olur. Ekmek, süt ve su bulunan sepetimizden hareketle hesaplanan sabit esaslı basit indeks değerleri kullanarak tartısız bileşik indeks değerlerini hesaplayalım.

Yıl	(2014=100) (Ekmek)	(2014=100) (Süt)	(2014=100) (Su)	Tartısız Sabit Esaslı Bileşik İndeks (2014=100)
2014	100,00	100,00	100,00	$(100,00+100,00+100,00)/3=100$
2015	117,65	125,00	133,33	$(117,65+125,00+133,33)/3=125,33$
2016	147,06	145,00	166,67	$(147,06+145,00+166,67)/3=152,91$

Yukarıda yer alan tablonun **Tartısız Sabit Esaslı Bileşik İndeks** sütununda 2014=100 için değerler hesaplanmıştır. Her bir madde için hesaplanan sabit esaslı basit indeks değerlerinin yıl bazında aritmetik ortalaması alınmıştır. Elde edilen değerler şu şekilde yorumlanmaktadır.

2014 yılında 100 lira ödeyerek aldığımız bu üç maddeyi 2015 yılında 125,33 lira ödeyerek almışız. Diğer bir ifadeyle, 2015 yılında bu üç maddenin fiyatlarının ortalaması baz yıl olan 2014'e göre %25,33 artış kaydetmiştir. 2014 yılında 100 lira ödeyerek aldığımız bu üç maddeyi 2016 yılında 152,91 lira ödeyerek almışız.

Aynı madde sepeti için **Tartısız Değişik Esaslı Bileşik İndeks** değerlerini hesaplayalım. Bu aşamada, her bir madde için değişik esaslı basit indeks değerleri hesaplanmış, ardından ise yıl bazında aritmetik ortalamaları bulunmuştur.

Yıl	Değişik esaslı basit indeks (Ekmek)	Değişik esaslı basit indeks (Süt)	Değişik esaslı basit indeks (Su)	Tartısız Değişik Esaslı Bileşik İndeks
2014	100,00	100,00	100,00	$(100,00+100,00+100,00)/3=100,00$
2015	117,65	125,00	133,33	$(117,65+125,00+133,33)/3=125,33$
2016	125,00	116,00	125,00	$(125,00+116,00+125,00)/3=122$

Tartısız deęişik esaslı bileşik indeks deęerleri sırasıyla 100, 125,33 ve 122,00 olarak hesaplanmıştır. Görüldüğü üzere; ilk iki yıla ait indeks deęerinin, sabit esaslı tartısız bileşik indeks deęeriyle aynıdır. Bundan dolayı sadece 2016 yılına ait aritmetik ortalama temelli tartısız deęişik esaslı bileşik indeks deęerini yorumlayacağız. 2015 yılında 100 lira ödeyerek aldığımız bu üç ürün, 2016 yılında 122 lira ödeyerek almışız.

Aritmetik ortalama temelli tartısız bileşik indeks hesabında sabit ve deęişik esaslı indeksler arasında geçiş yapılamamaktadır. Ayrıca aritmetik ortalamanın özellikleri göz önünde alındığında, seride aykırı deęerlerin bulunması durumunda geometrik ortalama temelli tartısız bileşik indeks hesabının yapılması önerilmektedir. Ancak geometrik ortalama temelli tartısız bileşik indeksin hesaplanması, aritmetik temelliye göre daha karmaşık olduğundan bu kitabın kapsamı dışında bırakılmıştır.

14.1.2. Tartılı Bileşik İndeksler

Tartısız bileşik indeks hesabında dikkate alınan maddeler, eşit önem derecesine sahiptir. Ancak bu maddelerin önem dereceleri birbirinden farklı ise, tartısız bileşik indeks deęerinin hesaplanması neticesinde yanıltıcı sonuçlara ulaşılabacağı bilinmektedir. Bu problemin üstesinden gelebilmek için, bileşik indeks hesabında maddelerin önem derecesini gösteren tartılardan yararlanılmaktadır. Yani, tartılı bileşik indeks hesaplanmaktadır. Bu indeks türüne ihtiyaç duymamızın nedenini anlatabilmek için bir örnek vermek gerekirse; sadece ekmek ve patates cipsinden oluşan bir madde sepetiniz olduğunu düşünün. Bu maddelerin bütçeniz içindeki payını düşündüğünüzde; ekmek fiyatının %10 pahalalanması ile patates cipsinin %10 zamlanmasının bütçenizde aynı etkiye meydana getirip getirmediğini kendinize sormamız gerekmektedir. Eğer ekmekteki fiyat artışı bütçenizi daha fazla etkiliyorsa, bileşik indeks hesabında ekmek ile patates cipsinin eşit öneme sahip olduğunu söyleyemeyiz. Peki maddelerin önem derecelerinin farklı olması durumunda, bunu indeks hesabına nasıl dâhil edeceğiz.

Hatırlanacağı üzere; birden fazla madde söz konusu olduğunda, madde fiyatlarında meydana gelen oransal deęişmelerin ortalamasının belirlenmesi için **bileşik indeksler** kullanılmaktadır. Eğer maddeler eşit öneme sahip ise, madde fiyatlarının ortalamalarından yararlanılarak indeks deęerine ulaşılmaktadır. Tartılı bileşik indekslerde ise, tartılı aritmetik ortalama hesabındakine benzer olarak tartılar kullanılmaktadır. Tartılı aritmetik ortalamayı hatırlamak için bir öğrencinin sadece üç ders aldığını varsayalım.

Ders	Not (X_i)	Dersin Kredisi (t_i)	$X_i \times t_i$
İktisat	3	5	15
İstatistik	4	3	12
Demografi	3	3	9

Bu öğrencinin üç farklı ders aldığı gözükmemektedir. Bunlar; iktisat, istatistik ve demografi dersleridir. Öğrencisinin değerlendirilmesi 4 üzerinden yapılmıştır. Derslerin kredisine bakıldığında; haftalık ders saatlerinin farklı olduğu, dolayısıyla öğrencinin iktisat dersinden aldığı 3 ile demografi dersinden aldığı 3 notlarının başarı notu üzerinde aynı etkiye sahip olmadığı anlaşılmaktadır. Öğrencinin notu ile dersin kredisi çarpılarak bulunan üçüncü sütunda, her bir dersin öğrencinin başarı puanına katkısı görülmektedir. Peki, öğrencinin not ortalaması nasıl elde edilecektir. Bu aşamada dersin kredisi, tartı olarak kullanılacaktır.

$$\bar{X} = \frac{\sum X_i \times t_i}{\sum t_i} = \frac{15 + 12 + 9}{11} = \frac{36}{11} = 3,27$$

Öğrencinin tartılı not ortalamasının 3,27 olduğu hesaplanmıştır. Bu örnekten hareketle, tartılı bileşik indekslerdeki tartıyı belirleyelim. Ekmek ve patates cipsi olmak üzere; tüketime konu olan sadece iki maddenin olduğu örneği hatırlayalım. Bu örnekte, madde fiyatlarının %10 zamlanması bütçemizde aynı etkiyi meydana getirip getirmediğini tartışmıştık. Bu noktadan hareketle, tükettiğimiz maddenin fiyatı ve miktarının çarparsak; bütçemizden o maddeyi satın almak için ödediğimiz tutara ulaşırız. Böylelikle, o maddenin bütçemiz içindeki payını da ortaya çıkarmış oluruz. Yani, maddenin fiyat ve miktar çarpımı tartı olarak ele alınmaktadır.

$$Tartı (t) = Fiyat \times Miktar = F \times M$$

Zaman serilerinden fiyatların oransal değişimlerini ortaya koyabilmek için karşılaştırma başlangıç dönemine göre yapılmaktadır. Dolayısıyla tartının hesaplanmasında başlangıç dönemine ait fiyat kullanılmaktadır. Tartının oluşturulmasında kullanılan ikinci unsur olan miktar içinse; başlangıç dönemindeki değer ya da hesaplamanın yapıldığı dönemdeki değer ele alınabilir. Bahsettiğimiz bilgilerden hareketle, tartılı bileşik indeks formülünü oluşturmaya çalışalım.

$$I_L = \frac{\sum \frac{F_i}{F_0} \times t}{\sum F_0 \times M_0} \times 100 = \frac{\sum \frac{F_i}{F_0} \times F_0 \times M_0}{\sum F_0 \times M_0} \times 100 = \frac{\sum F_i \times M_0}{\sum F_0 \times M_0} \times 100$$

Formülde görüldüğü üzere; fiyat ve miktar çarpımından oluşan tartı kullanılmıştır. Bu formülde; F_i : maddenin i. dönem fiyatını, F_0 : maddenin başlangıç dönemindeki fiyatını, M_0 : maddenin başlangıç dönemindeki miktarını göstermektedir. Formül incelendiğinde; tüketicinin başlangıç dönemindeki tüketim kalıbına (madde sepetine) ne kadar ödeme yaptığı payda kısmında görülmektedir. Formülün pay kısmı ise, başlangıç dönemindeki tüketim kalıbına i.

dönemde ne kadarlık bir ödeme yapıldığını göstermektedir. Özetlemek gerekirse; başlangıç dönemindeki 100 lira ödenen tüketim kalıbına, bugün ne kadarlık bir ödeme yapılması gerekmektedir. Bu indeksin tartılandırılmasında kullanılan miktara (M_0) bakıldığında, başlangıç dönemine ait olduğu görülmektedir. Dolayısıyla başlangıç dönem miktarlarının kullanıldığı bu indekse, **Laspeyres Fiyat İndeksi** adı verilmektedir.

Tartılı bileşik indeks hesabında, tartının oluşturulmasında kullanılan ikinci unsur olan miktar için başlangıç dönem değerini değil de hesaplanan dönemin değeri dikkate alınır; elde edilecek formül şu şekildedir:

$$I_P = \frac{\sum \frac{F_i}{F_0} \times t}{\sum F_0 \times M_i} \times 100 = \frac{\sum \frac{F_i}{F_0} \times F_0 \times M_i}{\sum F_0 \times M_i} \times 100 = \frac{\sum F_i \times M_i}{\sum F_0 \times M_i} \times 100$$

Formülde görüldüğü üzere; fiyat ve miktar çarpımından oluşan tartı kullanılmıştır. Bu formülde; F_i : maddenin i. dönem fiyatını, F_0 : maddenin başlangıç dönemindeki fiyatını, M_i : maddenin i. dönemdeki miktarını göstermektedir. Formülün pay kısmı, tüketicinin madde sepeti için ne kadarlık bir harcama yaptığını göstermektedir. Formülün payda kısmı ise, i. dönemde satın alınan madde sepeti, başlangıç döneminde tercih edilseydi yapılması gereken harcamayı göstermektedir. Formül incelendiğinde, her dönem tüketim kalıbının değişebileceği görülmektedir. Bu indeks ise **Paasche Fiyat İndeksi** olarak adlandırılmaktadır.

Ayrıca Laspeyres ve Paasche Fiyat İndekslerinin çarpımlarının karekökü alınarak elde edilen **Fisher'in İdeal İndeksi** de bulunmaktadır.

$$I_F = \sqrt{I_L \times I_P}$$

Bu kitap kapsamında tartılı bileşik indeksler başlığı altında bu üç indeks anlatılacaktır.

14.1.2.1. Laspeyres Fiyat İndeksi

Tartılı bileşik indekslerden biri olan Laspeyres Fiyat İndeksi, başlangıç (temel) dönemindeki fiyat ve miktarı tartı olarak kullanılmaktadır. Dolayısıyla başlangıç dönemindeki tüketim kalıbının sürdürülebilmesi için cari dönemde ne kadar ödenmesi gerektiği hususunda bilgi sunmaktadır. Bilindiği üzere, Türkiye İstatistik Kurumu tarafından hane halklarının tüketimine yönelik mal ve hizmet fiyatlarının zaman içindeki değişimi ölçülmektedir. Bu amaç doğrultusunda; Laspeyres Fiyat İndeksi kullanılmaktadır. Burada belirli bir hayat standardını koruyabilmek için harcanması gereken para miktarındaki değişimler ortaya konmaktadır. Diğer bir ifadeyle, hayat pahalılığında zaman içinde meydana gelen değişimler incelenmektedir. Bu aşamada mal ve hizmet sepetindeki maddelerin belirlenmesi ve ağırlıklandırılması önem arz etmektedir. Çünkü Laspeyres Fiyat İndeksinde, başlangıç döneminde tespit edilen maddeler ve ağırlıklar dikkate alınarak değişimler incelenmektedir. Ancak zaman içinde tüketicilerin tüketim kalıplarında yer alan madde kompozisyonu değişebileceğinden, madde sepetinde yer alacak mallar ile hizmetlerin ve ağırlıklarının güncellenmesi gerekmektedir. Aksi durumda, indeks değeri gerçeği yansıtmaktan uzaklaşacaktır. Laspeyres Fiyat İndeksinin hesaplanmasında bahsedildiği üzere; aşağıda yer alan formül kullanılmaktadır.

$$I_L = \frac{\sum F_i \times M_0}{\sum F_0 \times M_0} \times 100$$

Şimdi bu formülü kullanarak örnek uygulama yapalım. Aşağıdaki tabloda yer alan fiyat ve miktar bilgilerinden hareketle, esas dönemin 2014 yılı olduğu Laspeyres Fiyat İndeksi değerlerini hesaplayınız.

Yıl	Ekmek (adet)		Süt (lt)		Su (lt)	
	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
2014	0,85	10	2,00	7	0,75	15
2015	1,00	10	2,50	6	1,00	14
2016	1,25	8	2,90	3	1,25	12

$$I_L^{2014} = \frac{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)}{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)} \times 100 = \frac{33,75}{33,75} \times 100 = 100,00$$

$$I_L^{2015} = \frac{(1,00 \times 10) + (2,50 \times 7) + (1,00 \times 15)}{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)} \times 100 = \frac{42,50}{33,75} \times 100 = 125,92$$

$$I_L^{2016} = \frac{(1,25 \times 10) + (2,90 \times 7) + (1,25 \times 15)}{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)} \times 100 = \frac{51,55}{33,75} \times 100 = 152,74$$

Her bir dönem için Laspeyres Fiyat İndeksi değerleri hesaplanmıştır. Dikkat edilirse; payda kısmı her bir hesaplama için sabittir. Diğer bir ifadeyle, başlangıç dönemindeki tüketim kalıbını göstermektedir. Elde edilen indeks değerleri şu şekilde yorumlanmaktadır:

2014 yılında 100 lira ödenerek satın alınan madde sepeti 2015 yılında 125,92 lira; 2016 yılında ise 152,74 lira ödenerek alınmaktadır.

14.1.2.2. Paasche Fiyat İndeksi

Bu indeks, Laspeyres Fiyat İndeksine benzemekle beraber tartısı farklıdır. Bu indeksin hesaplanmasında; sabit bir tüketim kalıbı yerine hesaplanan dönemin tüketim kalıbı dikkate alınmaktadır.

Paasche Fiyat İndeksinin hesaplanmasında bahsedildiği üzere; aşağıda yer alan formül kullanılmaktadır.

$$I_P = \frac{\sum F_i \times M_i}{\sum F_0 \times M_i} \times 100$$

Şimdi bu formülü kullanarak örnek uygulama yapalım. Aşağıdaki tabloda yer alan fiyat ve miktar bilgilerinden hareketle, esas dönemin 2014 yılı olduğu Paasche Fiyat İndeksi değerlerini hesaplayınız.

Yıl	Ekmek (adet)		Süt (lt)		Su (lt)	
	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
2014	0,85	10	2,00	7	0,75	15
2015	1,00	10	2,50	6	1,00	14
2016	1,25	8	2,90	3	1,25	12

$$I_P^{2014} = \frac{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)}{(0,85 \times 10) + (2,00 \times 7) + (0,75 \times 15)} \times 100 = \frac{33,75}{33,75} \times 100 = 100,00$$

$$I_P^{2015} = \frac{(1,00 \times 10) + (2,50 \times 6) + (1,00 \times 14)}{(0,85 \times 10) + (2,00 \times 6) + (0,75 \times 14)} \times 100 = \frac{39,00}{31,00} \times 100 = 125,81$$

$$I_P^{2016} = \frac{(1,25 \times 8) + (2,90 \times 3) + (1,25 \times 12)}{(0,85 \times 8) + (2,00 \times 3) + (0,75 \times 12)} \times 100 = \frac{33,70}{21,80} \times 100 = 154,59$$

Elde edilen değerler, şu şekilde yorumlanmaktadır:

2014 yılında 100 lira ödenerek alınan madde sepeti 2015 yılında 125,81 liraya; 2016 yılında ise 154,59 liraya alınabilmektedir.

14.1.2.3. Fisher'in İdeal İndeksi

Zaman içinde tüketim kalıbının değişmesi, esas kabul edilen dönemin güncelliğini yitirmesi, hesaplanan dönemin tüketim kalıbının dikkate alınması nedeniyle her dönem için farklı tartı kullanılması gibi sakıncalardan dolayı Fisher'in İdeal İndeksi kullanılmaktadır. Laspeyres ve Paasche Fiyat İndekslerinin formüllerinin çarpılıp; karekökünün alınmasıyla elde edilmektedir. Diğer bir ifadeyle, bahsedilen indekslerin geometrik ortalamasıdır.

$$I_F = \sqrt{I_L \times I_P}$$

Laspeyres ve Paasche Fiyat İndeksleri için verilen örneklerden hareketle; Fisher'in İdeal İndeksi hesaplanmıştır.

Yıl	Laspeyres Fiyat İndeksi	Paasche Fiyat İndeksi	Fisher'in İdeal İndeksi
2014	100,00	100,00	100,00
2015	125,92	125,81	125,86
2016	152,74	154,59	153,66

14.2. Cari Fiyat Serisinin Deflate Edilmesi

Fiyat indeksleri, parasal miktarlarla açıklanan zaman serisi değerlerinin reel hale dönüştürülmesinde de kullanılmaktadır. Cari fiyatlarla hesaplanmış seriler, enflasyon etkisini içinde barındırdıklarından, farklı dönemlerin verilerinin karşılaştırılmasında gerçek artış veya azalışları göstermemektedir. Gerçek değişimin ortaya koyulabilmesi için serinin, enflasyon etkisinden arındırılması gerekmektedir. Bu aşamada, her yılın cari fiyat değerini o yıla karşılık gelen sabit esaslı indeks değerine bölüp, 100 ile çarpmak gerekmektedir. Böylelikle cari fiyat serisinden, reel fiyat serisine geçilmiş olur. Bu süreçte kullanılan formül, aşağıda gösterilmiştir.

$$Reel Fiyat_i = \frac{Cari Fiyat_i}{Sabit Esaslı İndeks_i} \times 100$$

Bu konuyla örnek yardımıyla açıklayabilmek için aylık gelirimizin 100 lira olduğunu ve bir adet tükenmez kalemin de 10 lira olduğu varsayalım. Bu durumda, aylık gelirimizle 10 adet tükenmez kalem alabiliriz. Bir sonraki sene ise, aylık gelirimiz 120 liraya yükseldi ancak bir adet tükenmez kalemin fiyatı da 15 lira oldu. Bu durumda aylık gelirimiz ile sadece 8 kalem alabiliriz. Dolayısıyla aylık gelirimiz 100 liradan 120 liraya yükselmiş olmasına rağmen, satın alabileceğimiz kalem miktarı azalmıştır. Halbuki gelirimiz arttığı için daha fazla kalem almayı beklerdik. İşte bu örnekteki 100 lira ve 120 lira değerleri, cari değerleri göstermektedir. Peki, reel değerler yani satın alma gücü kaç lira olacaktır?

Tekstil sektöründe çalışan bir kişinin yıllar itibariyle ortalama aylık geliri ve 2014=100 bazlı tüketici fiyat indeksi değerleri aşağıdaki tabloda yer almaktadır. Bu bilgilerden hareketle, çalışanın reel ortalama aylık gelirini hesaplayalım.

Yıl	Ortalama Aylık Gelir (lira)	Tüketici Fiyat İndeksi (2014=100)
2014	1250	100
2015	1580	125
2016	1750	138

Tablo incelendiğinde; 2014 – 2016 yıllarında çalışan bir kişinin ortalama aylık gelirin ve sabit esaslı indeks olarak tüketici fiyat indeksine ait değerlerin yer aldığı görülmektedir.

$$Reel\ Fiyat_{2014} = \frac{1250}{100} \times 100 = 1250\ lira$$

$$Reel\ Fiyat_{2015} = \frac{1580}{125} \times 100 = 1264\ lira$$

$$Reel\ Fiyat_{2016} = \frac{1750}{140} \times 100 = 1268\ lira$$

Tekstil sektöründe çalışan kişinin reel ortalama aylık gelirin neredeyse hiç değişmediği gözlenmiştir. Dolayısıyla cari gelirinde yaşanan yüksek oranlı artış, reel gelirine pek yansımamıştır.

Bu Bölümde Neler Öğrendik?

Bileşik indeksleri incelediğimiz bu bölümde, öncelikle bileşik indeks kavramı üzerinde durarak, bileşik indekslerin kullanım amacını ele aldık ve türlerini öğrendik.

Parasal miktarlarla açıklanan zaman serisi değerlerinin reel hale dönüştürülmesinde fiyat indekslerinden nasıl yararlanılacağı konusunu ele aldık. Reel değer ve cari değer kavramları üzerinde durduk. Bir örnek üzerinden reel fiyat serisinin nasıl elde edildiğini öğrendik.

Bölüm Soruları

- 1- Birden fazla madde söz konusu olduğunda, madde fiyatlarında meydana gelen oransal değişmelerin ortalamasının belirlenmesi için indeksler kullanılmaktadır.
 - a- Basit
 - b- Bileşik
 - c- Sabit esaslı
 - d- Değişik esaslı
 - e- Sabit esaslı basit
- 2- Başlangıç dönemindeki birden fazla mal ve hizmeti içeren tüketim kalıbının sürdürülebilmesi için cari dönemde ne kadar ödeme yapılması gerektiği hususunda indeksi kullanılmaktadır.
 - a- Sabit esaslı basit
 - b- Değişik esaslı basit
 - c- Paasche fiyat
 - d- Laspeyres fiyat
 - e- Fisher'in ideal
- 3- Cari fiyatlarla hesaplanmış seriler, enflasyon etkisini içinde barındırdıklarından karşılaştırmalarda gerçek artış veya azalışları görebilmek için sabit esaslı fiyat indeksleri kullanılarak elde edilmektedir.
 - a- Reel fiyat serisi
 - b- Cari fiyat serisi
 - c- Nominal fiyat serisi
 - d- Basit seri
 - e- Bileşik seri

4- Aşağıdaki tabloyu kullanarak altının 2015 yılı reel fiyatını hesaplayınız.

Yıl	Altın fiyatı (gram/lira)	Tüketici Fiyat İndeksi (2014=100)
2014	140	100
2015	150	125
2016	175	138

- a- 71
- b- 120
- c- 140
- d- 172
- e- 199

5- Aşağıdaki tabloyu kullanarak X marka cep telefonunun 2016 yılı reel fiyatını hesaplayınız.

Yıl	X marka cep telefonu (adet/lira)	Tüketici Fiyat İndeksi (2014=100)
2014	1699	100
2015	2499	140
2016	3099	160

- a- 1699,00
- b- 1785,00
- c- 1915,40
- d- 1936,88
- e- 2099,00

- 6- Aşağıdaki tablodan hareketle; 2016 yılına ait Laspeyres fiyat indeksi değerini hesaplayınız.

Yıl	Margarin (adet)		Süt (lt)		Su (lt)	
	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
2014	1,5	10	2,15	7	0,50	20
2015	2,00	10	2,50	6	1,00	17
2016	2,25	10	2,90	3	1,50	15

- a- 100,00
b- 143,57
c- 160,74
d- 181,77
e- 193,79

- 7- Aşağıdaki tablodan hareketle; 2015 yılına ait Laspeyres fiyat indeksi değerini hesaplayınız.

Yıl	Margarin (adet)		Süt (lt)		Su (lt)	
	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
2014	1,5	10	2,15	7	0,50	20
2015	2,00	10	2,50	6	1,00	17
2016	2,25	10	2,90	3	1,50	15

- a- 100,00
b- 143,57
c- 160,74
d- 181,77
e- 193,79

- 8- Aşağıdaki tablodan hareketle; 2014 yılına ait Laspeyres fiyat indeksi değerini hesaplayınız.

Yıl	Margarin (adet)		Süt (lt)		Su (lt)	
	Fiyat	Miktar	Fiyat	Miktar	Fiyat	Miktar
2014	1,5	10	2,15	7	0,50	20
2015	2,00	10	2,50	6	1,00	17
2016	2,25	10	2,90	3	1,50	15

- a- 100,00
- b- 143,57
- c- 160,74
- d- 181,77
- e- 193,79
- 9- Aşağıda birden fazla mal ve hizmetten oluşan madde sepetindeki miktar ve fiyat bilgilerinden hareketle Laspeyres Fiyat İndeksi ve Paasche Fiyat İndeksi değerleri elde edilmiştir. 2015 yılına ait Fisher'in İdeal İndeks değerini hesaplayınız.

Yıl	Laspeyres Fiyat İndeksi	Paasche Fiyat İndeksi
2014	100,00	100,00
2015	126,35	124,99
2016	153,41	153,67

- a- 100,00
- b- 124,99
- c- 125,67
- d- 126,35
- e- 128,50

10- Aşağıda birden fazla mal ve hizmetten oluşan madde sepetindeki miktar ve fiyat bilgilerinden hareketle Laspeyres Fiyat İndeksi ve Paasche Fiyat İndeksi değerleri elde edilmiştir. 2016 yılına ait Fisher'in İdeal İndeks değerini hesaplayınız.

Yıl	Laspeyres Fiyat İndeksi	Paasche Fiyat İndeksi
2014	100,00	100,00
2015	126,35	124,99
2016	153,41	153,67

- a- 100,00
- b- 153,41
- c- 153,54
- d- 153,67
- e- 155,41

Cevaplar

- 1- b
- 2- d
- 3- a
- 4- b
- 5- d
- 6- d
- 7- b
- 8- a
- 9- c
- 10- c

Yararlanılan Kaynaklar

AKMAN ULUTÜRK, S. (2008), “2003 Temel Yılı Tüketici Fiyat İndeksine İlişkin Düşünceler”, *İstanbul Üniversitesi İktisat Fakültesi Maliye Araştırma Merkezi Konferansları*, 51. Seri.

CİLLOV, H., (1993), *İktisadi Olaylara Uygulanan İstatistik Metodları*, İstanbul: İÜ Basımevi ve Film Merkezi, 1993.

GÜLER, F., (2012), *İstatistik Metotları ve Uygulamaları*, 3.Basım, İstanbul: Beta Yayınları.

GÜRSAKAL, N., *Betimsel İstatistik*, 7. Baskı, Bursa: Dora Yayınları.

RALPH, J., O’NEILL, R., WINTON J., (2015), *A Practical Introduction to Index Numbers*, John Wiley & Sons Ltd.

SERPER, Ö., (2014), *Uygulamalı İstatistik*, 7.Baskı, Bursa: Ezgi Kitabevi.

TURANLI, M., GÜRİŞ, S., (2005), *Temel İstatistik*, Genişletilmiş 2.Basım, İstanbul: Der Yayınları.