ÖNSÖZ

Atatürk Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü'nün "Olasılık ve İstatistik" dersinde okutulmak üzere değişik zamanlarda okuttuğumuz notları birleştirerek bu ders notunu hazırladık.

Bu kitap sahasında iddialı olmayıp, sadece konulara giriş ve tanıtma niteliğindedir. Bu yüzden anlaşılır şekilde yazılmaya özen gösterilmiştir. Konuların anlaşılması için çok sayıda çözülmüş örneğe yer verilmiştir. Dolayısıyla bu kitaptan faydalanmak için, Kümeler Teorisi ve Analiz derslerini bilmek yeterlidir.

Kitap 10 bölümden meydana gelmektedir. Birinci bölüm, sayma metodlarına ayrılmıştır. Burada verilenler sonraki her bölümde bir araç olarak kullanılmıştır.

İkinci bölüm ise, olasılık teorisine ilişkin klasik ve aksiyomatik tanımlar, koşullu olasılık ve Bayes teoremine ayrılmıştır.

Üçüncü bölümde, rastlantı değişkenleri, olasılık fonksiyonu ve dağılımları, ayrıca beklenen değer ve özellikleri hakkında bilgi verilmiştir.

Dördüncü ve beşinci, üçüncü bölümde tanıtılan kesikli ve sürekli rastlantı değişkenlerin dağılımlarına ayrılmıştır.

Altıncı bölüm ve diğerleri istatistik ile ilgili verilerin düzenlenmesi ve analiz edilmesine ayrılmıştır.

Kitabı hazırlarken büyük bir özen göstermemize rağmen yine de bazı hataların olabileceğini kabul ediyoruz. Bunların giderilmesi için okuyucuların ilgisini bekliyoruz.

Kitabın hazırlanmasında yardımlarını gördüğümüz Matematik Bölümü elemanlarına, basım için yardımlarını esirgemeyen Fen-Edebiyat Fakültesi Dekanlığına teşekkür ederiz.

Yazarlar

Contents

Chapter 1

TEMEL KAVRAMLAR

1.1 Sayma Metodları

Verilen bir küme çeşitli yönleri ile incelenebilir. Bu bölümde biz, sonlu bir kümenin eleman sayısı ile ilgileneceğiz. Eleman sayısının tesbiti için bazı metodlar vardır. Bunların en iyi bilineni, verilen kümenin elemanlarını tek tek saymaktır. Ancak, bu her zaman için pratik değildir. Bu nedenle, saymayı daha pratik yapmamızı sağlayan metodları vereceğiz.

1.1.1 Saymanın Temel Prensibi

Aşağıdaki örnekleri inceleyerek anlatılmak isteneni biraz daha gözler önüne serelim.

i. Trabzon'dan Giresun'a gitmek için 10 otobüs, 15 minibüs ve 3 tane de gemi vardır. Dolayısıyla Trabzon'dan Giresuna gitmek isteyen birisi otobüsü veya minibüsü veya gemiyi tercih edecektir. Bu durumda otobüs veya minibüs veya gemi ile gitme tercihi 10+15+3=28 şekilde olur.

Bunu daha toplu olarak şöyle söyleyebiliriz: A, "otobüs ile gitme"; B "minibüs ile gitme" ve C de "gemi ile gitme" olaylarını gösterirse

$$A \text{ veya } B \text{ veya } C \text{ olayı}, \quad n(A) + n(B) + n(C) = 10 + 15 + 3 = 28$$

farklı şekilde olur. Burada $n(X)$ ile X kümesinin eleman sayısı gösterilmiştir.

Genel Kural: $A_1, A_2, \ldots, A_k$ olaylarının oluş sayısı sırasıyla $n_1, n_2, \ldots, n_k$ olsun. Bu takdirde

$$A_1 \text{ veya } A_2 \text{ veya } \cdots \text{ veya } A_k \text{ olayı, } \; n_1 + n_2 + \cdots + n_k$$

farklı şekilde olur.

ii. X şehrinden Y şehrine gitmek için 3 yol; Y şehrinden Z şehrine gitmek için 5 yol vardır. Y den geçmek şartıyla X den Z şehrine gitmek isteyen birisi (3)(5)=15 farklı yol tercihi yapabilir.

A ile "X den Y ye gitmeyi"; B ile "Y den Z ye gitmeyi" gösterelim. Y den geçmek şartıyla X den Z ye gitmek için A ve B olaylarının olması gerekir. Dolayısıyla

$$A \text{ ve } B \text{ olayı, } \; (3)(5)=15$$

farklı şekilde olur.

Genel Kural: $A_1, A_2, \ldots, A_k$ olaylarının oluş sayısı sırasıyla $n_1, n_2, \ldots, n_k$ olsun. Bu takdirde

$$A_1 \text{ ve } A_2 \text{ ve } \cdots \text{ ve } A_k \text{ olayı, } \; n_1 n_2 \cdots n_k$$

farklı şekilde olur.

Yukarıdaki genel kurallardaki "ve", "veya" kelimelerinin kullanışlarına dikkat ediniz.

Örnek 1.1 $1, 2, 3, 4, 5$ *sayıları ile hiç bir sayı tekrarlanmadığına göre, üç basamaklı kaç sayı yazılabilir.*

Çözüm 1.1 *Her basamağı bir kutu olarak düşünürsek* $1, 2, 3, 4, 5$ *sayıları ile doldurulacak üç yer vardır. İlk yer beş sayının herhangi biri ile doldurulabilir. İkinci yer geriye kalan dört sayının herhangi biri ile, üçüncü yer de* *üç sayının herhangi biri ile doldurulabilir. Saymanın temel ilkesine göre,* $1, 2, 3, 4, 5$ *sayıları ile hiç bir sayı tekrarlanmadığına göre* $(5)(4)(3) = 60$ *tane üç basamaklı sayı yazılabilir.*

1.1.2 Faktöriyel Sembolü

1 sayısından n sayısına kadar bütün pozitif tam sayıların çarpımı $n!$ faktöriyel ile gösterilir. Yani,

$$n! = (1)(2)(3)\cdots(n-2)(n-1)n$$

olur. Özel olarak $0! = 1$ olarak tanımlanır. Ayrıca $1! = 1$ dir.

n büyüdükçe $n!$ hesaplamak oldukça güçtür. Bunun için büyük sayıların faktöriyeli **Stirling Formülü** denilen

$$n! \approx \sqrt{(2\pi n)}\left(\frac{n}{e}\right)^n \tag{1.1}$$

ile yaklaşık olarak hesaplanır.

1.1.3 Permütasyonlar

Birbirinden farklı n tane elemanın (hepsini alarak) düzenlenmesine **permütasyon** denir.

Teorem 1.1 *n tane farklı elemanın hepsi sıralandığı zaman elde edilecek değişik düzenlerin sayısı $n!$ kadardır.*

İspat 1.1 *Doldurulacak n tane yer vardır. İlk yer n tane nesnenin herhangi biri ile n yolla ve ikinci yer, geriye kalan $(n-1)$ nesne ile $(n-1)$ yolla doldurulabilir. İşlem benzer şekilde sürdürülürse saymanın temel ilkesine göre farklı düzenlerin sayısı,*

$$n(n-1)\cdots(2)(1) = n!$$

olarak bulunur.

n farklı nesnenin permütasyonlarının (sıralanmasından elde edilen düzenlerin sayısı) $P(n,n)$ ile gösterilir. Buna göre,

$$P(n,n) = n! \tag{1.2}$$

dir.

Örnek 1.2 *7 kişi bir gişe önünde kaç farklı şekilde sıralanabilir.*

Çözüm 1.2 *7 kişinin farklı düzende sıralanma sayısı*

$$P(7,7) = 7! = 5040$$

olur.

n nesnenin bazıları aynı olması durumunda, permütasyonları bulmak için kullanılan genel formül aşağıdaki teoremde verilmiştir.

Teorem 1.2 *n tane nesnenin n_1 tanesi bir türden, n_2 tanesi ikinci türden, ... , n_k tanesi $k-$ıncı türden olsun. Bu n nesnenin tümü sıralanırsa, elde edilecek farklı düzenlerin sayısı,*

$$\frac{n!}{n_1!n_2!\cdots n_{k-1}!n_k!}, \quad n = n_1 + n_2 + \cdots + n_k \tag{1.3}$$

dır.

İspat 1.2 *İspatı n_1 nesnenin aynı, diğerlerinin farklı olduğunu kabul ederek yapacağız. Aynı nesneler $x_1, x_2, \cdots, x_{n_1}$ olsun. Geriye kalan nesnelerden oluşan bir tek düzeni gözönüne alalım ve onu k ile gösterelim. Bu durumda*

$$x_1x_2\cdots x_{n_1}k, \;\; x_{n_1}x_1x_2\cdots x_{n_1-1}k, \;\; \cdots$$

yazılır. Bunların sayısı $n_1!$ dir. Ancak bunlar bizim istediğimiz şekilde düzen olarak 1 tanedir. Bu her durum için aynıdır. Eğer bütün nesneler farklı olsaydı $n!$ tane düzen elde edilecekti. $x_1x_2\cdots x_{n_1}$ aynı olduğundan $n_1!$ çarpanı kadar düzen aynı olacağından, istenen farklı düzenlerin sayısı

$$\frac{n!}{n_1!} \tag{1.4}$$

dir. Benzer muhakeme uygulanarak $n_1, n_2, \cdots, n_k$ *tane benzer nesne için elde edilecek farklı düzenlerin sayısı*

$$\frac{n!}{n_1!n_2!\cdots n_k!} \tag{1.5}$$

olur.

Örnek 1.3 *İSTATİSTİK kelimesinin harflerini her düzende kullanma şartıyla, kaç farklı kelime elde edilebilir?*

Çözüm 1.3 *İSTATİSTİK kelimesinde, beş farklı harf vardır: İ,S,T,A,K. Burada* $n_1 = 3$, $n_2 = 2$, $n_3 = 3$, $n_4 = 1$ *ve* $n_5 = 1$ *dir. Elde edilecek değişik kelime sayısı*

$$\frac{10!}{3!2!3!1!1!} = 50400$$

olur. Burada şu hususa dikkat etmek gerekir: Birinci İ harfini $\dot{I}_1$, *ikinci İ harfini* $\dot{I}_2$, *üçüncü İ harfini* $\dot{I}_3$ *ile gösterelim. Dolayısı ile,*

$$\dot{I}_1STAT\dot{I}_2ST\dot{I}_3K, \quad \dot{I}_2STAT\dot{I}_1ST\dot{I}_3K, \quad \dot{I}_3STAT\dot{I}_1ST\dot{I}_2K,$$

$$\dot{I}_1STAT\dot{I}_3ST\dot{I}_2K, \quad \dot{I}_2STAT\dot{I}_3ST\dot{I}_1K, \quad \dot{I}_3STAT\dot{I}_2ST\dot{I}_1K$$

kelimeleri aynıdır. Permütasyon olarak bunların sayısı 1 *dir.*

Örnek 1.4 *4 kırmızı, 3 beyaz ve 1 mavi bayraktan oluşan bir kümeden her biri 8 bayrak ihtiva eden kaç farklı düzen elde edilir?*

Çözüm 1.4

$$\frac{8!}{4!3!1!} = 280$$

bulunur.

Bir başka permütasyon çeşidi de dairesel permütasyondur. n tane farklı nesneyi bir daire çevresinde sıralarsak, elde edeceğimiz değişik düzenlerin sayısı

$$P(n,n) = (n-1)! \tag{1.6}$$

olur.

Örnek 1.5 *Bir masada yemek yiyen 4 arkadaşın farklı permütasyonlarının sayısı nedir?*

Çözüm 1.5 *Masa etrafında oturan 4 kişiyi A, B, C ve D ile gösterelim. Aşağıdaki şekildeki dört düzenleme aynıdır. Her durumda, A nın solunda B, sağında ise D vardır.*

Şekil 1.1:

Farklı düzenlemelerin sayısını bulmak için bir kişiyi (mesela A yı) sabit tutacağız ve sonra B, C ve D ile yapılacak düzenlemelerin sayısını hesaplayacağız. O halde

$$(n-1)! = (4-1)! = 3! = 6$$

olur.

n tane farklı nesneyi bir halka çevresinde sıralarsak elde edeceğimiz değişik düzenlerin sayısı

$$\frac{(n-1)!}{2} \tag{1.7}$$

dir.

Eğer A, B, C, D şeklindeki dört farklı boncuk bir halkaya geçirilmiş ise aşağıdaki iki düzen aynıdır. Çünkü A ve C sabit tutularak halka 180 derece döndürülmekle bir şekilden diğeri elde edilir.

.

Şekil 1.2:

Örnek 1.6 *1 beyaz, 1 mavi, 1 kırmızı ve 2 sarı boncuk bir halkaya geçiriliyor.*

i. Kaç farklı düzen elde edilir?

ii. Kırmızı ve beyaz top yan yana gelmek şartıyla kaç farklı düzen elde edilir?

Çözüm 1.6 *i. Eğer boncuklar farklı olsaydı, 5 boncuk ile yapılacak düzenlemelerin sayısı* $\frac{4!}{2}$ *dir.*

Ancak boncukların ikisi sarı olduğundan düzenlemelerin sayısı

$$\frac{4!}{2(2!)} = 6$$

olur.

ii.

$$\frac{2!3!}{2(2!)} = 3$$

olur. Paydaki 2! kırmızı ve beyaz topların kendi aralarında oluşturacakları düzenlerin sayısını; 3! ise bir halkadaki 4 nesnenin oluşturacağı düzenleri göstermektedir. Paydadaki 2! iki sarı *topun varlığından dolayı yazılmış; 2 ise halka üzerinde yapılacak düzenlemelerin sayısını veren formülden gelmektedir. Bu sonuçları şekil olarak aşağıdaki gibi gösteririz.*

Şekil 1.3:

Verilen n nesnenin hepsini kullanarak değilde, k tanesini kullanarak elde edilecek permütasyonlarla da ilgileneceğiz. Bu tip permütasyonlar iki sınıfa ayrılır.

Birincisi, k elemanın hepsinin farklı olması yani tekrarsız olması durumudur. Örneğin, a, b, c harflerinden oluşan 2-li tekrarsız permütasyonlar,

$$ab, ac, bc, ba, ca, cb$$

dır.

İkincisi, k elemanın hepsinin farklı olmaması yani tekrarlı olması durumudur. Örneğin, a, b, c harflerinden oluşan 2-li tekrarlı permütasyonlar

$$aa, ab, ac, bb, bc, ba, cc, ca, cb$$

dır.

Teorem 1.3 *n tane farklı nesnenin k tanesi $(k \leq n)$ tekrarsız olarak sıralanırsa, elde edilecek değişik düzenlerin sayısı $P(n, k)$ ile gösterilir ve*

$$P(n,k) = \frac{n!}{(n-k)!} \tag{1.8}$$

yazılır. Eğer bu k nesne tekrarlı olarak sıralanırsa, elde edilecek değişik düzenlerin sayısı

$$n^k$$

dır.

İspat 1.3 *Doldurulacak k tane yer vardır. İlk yer, n nesneden herhangi biri ile ve ikinci yer, geriye kalan $(n-1)$ tane nesneden herhangi biri ile doldurulur. Benzer işlemler ile k.yer $n-(k-1)$ yolla doldurulabilir. Saymanın temel ilkesine göre*

$$P(n,k) = n(n-1)\cdots[n-(k-1)] \tag{1.9}$$

yazabiliriz. Bu eşitlik $(n-k)!/(n-k)!$ ile çarpılırsa,

$$P(n,k) = \frac{n!}{(n-k)!} \tag{1.10}$$

elde edilir.

Diğer hal, saymanın temel prensibinden görülür.

Örnek 1.7 *7 haneli telefon numarası*

i. Hiç bir şart olmaksızın,

ii. Her rakam farklı olmak üzere

kaç şekilde yazılabilir.

Çözüm 1.7 *i. 10 tane rakamın herhangi biri her haneye gelebilir. O halde bu nesneler tekrarlı olarak sıralanmaktadır. Dolayısıyla*

$$10^7 = 10\,000\,000$$

tane telefon numarası yazılabilir.

ii. Bu durumda nesneler tekrarsız olarak sıralandığından

$$P(10,7) = \frac{10!}{(10-7)!} = (10)(9)(8)(7)(6)(5)(4) = 604800$$

tane numara yazılabilir.

1.1.4 Kombinasyonlar

Birbirinden farklı n tane nesne verilsin. Bu nesnelerden herhangi r tanesi rasgele alınır ve bu r nesne herhangi bir sırada yazılırsa, buna r mertebeli kombinasyon denir.

Kombinasyonlar iki sınıfa ayrılır: Birincisi r nesnenin de farklı yani tekrarsız olması, ikincisi ise r nesnenin farklı olmaması yani bazılarının tekrarlı olmasıdır.

Permütasyonda sıra önemlidir. Kombinasyonda ise sıranın önemi yoktur. Örneğin, abc, acb, bca, bac, cab, cba permütasyon olarak birbirinden farklıdır, ama kombinasyon olarak hepsi aynıdır. Böylece şu benzetmeyi yapabiliriz: Değişme özelliği permütasyonun nesneleri arasında yok, kombinasyonun nesneleri arasında vardır. Bunu daha iyi görmek için Örnek 1.8'e bakınız.

Teorem 1.4 *n farklı nesnenin r mertebeli tekrarsız kombinasyonlarının sayısı*

$$C(n,r) = \binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{1.11}$$

dir. Tekrarlı kombinasyonların sayısı ise

$$C(n,r) = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!((n+r-1)-r)!} \tag{1.12}$$

dır.

İspat 1.4 *Bilindiği gibi n farklı nesnenin r tanesi alınarak elde edilen tekrarsız permütasyonların sayısı $P(n,r) = \frac{n!}{(n-r)!}$ dır. Burada her bir kombinasyon $r!$ kadar tekrar edilmiştir. Teorem 1.2 ye göre her bir kombinasyonun 1 tane olduğu düzenlemelerin sayısı*

$$C(n,r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

dir. İkinci kısmın ispatı tümevarım ile yapılır.

Örnek 1.8 *a, b, c, d gibi dört harften üç harf alarak yapılan kombinasyon sayısını bulunuz.*

Çözüm 1.8 *Burada* $n = 4$ *ve* $r = 3$ *olduğundan*

$$C(n,r) = \frac{n!}{r!(n-r)!} = \frac{4!}{3!(4-3)!} = 4$$

dür.

Kombinasyon ve permütasyon kavramlarını mukayese etme bakımından aşağıdaki tabloyu verelim:

Kombinasyonlar	*Permütasyonlar*
abc	$abc, acb, bac, bca, cab, cba$
abd	$abd, adb, bad, bda, dab, dba$
acd	$acd, adc, cad, cda, dac, dca$
bcd	$bcd, bdc, cbd, cdb, dbc, dcb$

Kombinasyonlar ve permütasyonlar arasında

$$P(n,r) = r!C(n,r) \Longrightarrow C(n,r) = \frac{P(n,r)}{r!}$$

bağıntısı vardır.

Örnek 1.9 *8 kişi arasından 4 kişi kaç değişik şekilde seçilebilir?*

Çözüm 1.9 *Seçimde sıralama önemli değildir.Burada* $n = 8$ *ve* $r = 4$ *dür. Dolayısıyla farklı seçim sayısı,*

$$C(8,4) = \binom{8}{4} = \frac{8!}{4!(8-4)!} = \frac{8!}{4!4!} = 70$$

olarak bulunur.

Örnek 1.10 *4 kız ve 7 erkek öğrenciden 5 kişilik bir kurul seçilecektir. Kurul, 2 kız ve 3 erkek öğrenciden kurulmak istendiğine göre, kaç farklı seçim yapılabilir?*

Çözüm 1.10 *Kurula seçilecek 2 kız öğrenci için* $\binom{4}{2}$, *3 erkek öğrenci için de* $\binom{7}{3}$ *farklı seçim yapılabilir. Saymanın temel ilkesine göre 5 kişilik değişik kurul sayısı*

$$\binom{4}{2}\binom{7}{3} = 210$$

olarak bulunur.

Örnek 1.11 *a, b, c elemanlarının üçlü tekrarlı kombinasyonlarını ve ikili tekrarlı kombinasyonlarını yazınız.*

Çözüm 1.11 *Üçlü tekrarlı kombinasyonlar: aaa, aab, aac, abb, abc, acc, bbb, bbc, bcc, ccc.*

İkili tekrarlı kombinasyonlar: aa, ab, ac, bb, bc, cc.

Kombinasyonla ilgili şu özellik vardır:

Teorem 1.5

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}$$

eşitliği vardır.

İspat 1.5

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad ve \quad \binom{n}{r+1} = \frac{n!}{(r+1)![n-(r+1)]!}$$

olduğunu biliyoruz. Buradan,

$$\begin{aligned}
\binom{n}{r} + \binom{n}{r+1} &= \frac{n!}{r!(n-r)!} + \frac{n!}{(r+1)![n-(r+1)]!} \\
&= \frac{n!}{r!(n-r)!}(1 + \frac{n-r}{r+1}) \\
&= \frac{n!}{r!(n-r)!}(\frac{r+1+n-r}{r+1})
\end{aligned}$$

$$\begin{aligned} &= \frac{n!(n+1)}{(r+1)!(n-r)!} \\ &= \frac{(n+1)!}{(r+1)!(n-r)!} \\ &= \frac{(n+1)!}{(r+1)![(n+1)-(r+1)]!} \\ &= \binom{n+1}{r+1} \end{aligned}$$

elde edilir.

Ayrıca şu iki özelliği de verebiliriz:

i. $\binom{n}{0} + \binom{n+1}{1} + \cdots + \binom{n+k}{k} = \binom{n+k+1}{k}$

ii. $\binom{n}{n-k} = \binom{n}{k}$

Tanım 1.1 *$n_1, n_2, \cdots, n_r$ pozitif tamsayılar ve $n_1 + n_2 + \cdots + n_r = n$ olsun.*

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

şeklinde tanımlanır.

Örnek 1.12 $\binom{7}{4,3,0}$ *ifadesinin değerini bulunuz.*

Çözüm 1.12

$$\binom{7}{4,3,0} = \frac{7!}{4!3!0!} = 35$$

olarak hesaplanır.

Teorem 1.2 de bazı nesneleri aynı olan bir kümenin permütasyonlarının sayısının nasıl bulunacağını gördük. Şimdi bu teoremi biraz daha değişik bir şekilde ifade edeceğiz. Teorem 1.2 de bazı nesnelerin aynı olduğu, burada ise nesnelerin farklı olduğu durum gözönüne alınmıştır.

Teorem 1.6 *n elemanlı bir A kümesi, A_1 kümesinde n_1, A_2 kümesinde $n_2, \cdots, A_k$ kümesinde n_k eleman olmak üzere $\{A_1, A_2, \cdots, A_k\}$ şeklinde parçalanmak isteniyor. Bu şekilde A kümesi*

$$\frac{n!}{n_1!n_2!\cdots n_k!}, \quad (n_1 + n_2 + \cdots + n_k = n) \tag{1.13}$$

kadar farklı parçalara ayrılır.

İspat 1.6 *A_1 kümesine n nesneden n_1 tane farklı nesneyi $\binom{n}{n_1}$ farklı şekilde ve A_2 kümesine $n - n_1$ nesneden n_2 tane farklı nesneyi $\binom{n-n_1}{n_2}$ farklı şekilde ve nihayet A_k kümesine $n - n_1 - \cdots - n_{k-1}$ nesneden n_k tane farklı nesneyi $\binom{n-n_1-\cdots-n_{k-1}}{n_k}$ farklı şekilde seçebiliriz. O halde saymanın temel prensibine göre parçalanmaların sayısı*

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{k-1}}{n_k}$$

dır. Gerekli işlemleri yaparsak

$$\frac{n!}{n_1!n_2!\cdots n_k!}, \quad n_1 + n_2 + \cdots + n_k = n$$

elde edilir.

Örnek 1.13 *9 oyuncak, 4 kardeş arasında en küçüğü 3 tanesini diğerleri eşit sayıda olmak üzere kaç farklı şekilde paylaştırılır?*

Çözüm 1.13 *9 oyuncağı sırasıyla 3, 2, 2, 2 oyuncağı içinde bulunduran 4 göze kaç farklı şekilde yerleştireceğimizin sayısını bulmak istiyoruz. Teorem (1.6)'e göre*

$$\frac{9!}{3!2!2!2!} = 7560$$

olarak bulunur.

1.1.5 Binom Teoremi

Teorem 1.7 *n bir pozitif tamsayı olmak üzere*

$$
\begin{aligned}
(a+b)^n &= \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots \\
&\quad + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n \\
&= \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r
\end{aligned}
$$

olur.

İspat 1.7 *Tümevarım metodu kullanılarak ispat edilebilir.*

NOT: a herhangi bir sayı ve $k \geq 0$ bir tam sayı olmak üzere binom katsayıları

$$
\binom{a}{k} = \frac{a(a-1)(a-2)\cdots(a-k+1)}{k!} \tag{1.14}
$$

ile tanımlanır. Eğer $a \geq 0$ bir tamsayı ise bu katsayı

$$
\binom{a}{k} = \frac{a!}{k!(a-k)!} \tag{1.15}
$$

şekline dönüşür. Örneğin,

$$
\begin{aligned}
\binom{-2}{r} &= \frac{-2(-3)(-4)\cdots(-2-r+1)}{r!} \\
&= \frac{(-1)^r 2\cdot 3\cdot \cdots \cdot r(r+1)}{r!} \\
&= (-1)^r(r+1)
\end{aligned}
$$

dir.
$(a+b)^n$ açılımının aşağıdaki özellikleri vardır:

i. $(a+b)^n$ açılımında $n+1$ terim vardır.

ii. Her terimdeki a ve b terimlerinin üsleri toplamı n olur.

iii. a teriminin üsleri n sayısından sıfıra kadar azalır, b teriminin üsleri sıfırdan n sayısına kadar terim terim artar.

iv. k. terimin katsayısı $\binom{n}{k-1}$ olur.

v. Uçlardan eşit uzaklıktaki terimlerin katsayıları eşittir.

$(a+b)^n$ teriminin katsayıları, Paskal Üçgeni denilen aşağıdaki tablodan yararlanılarak da hesaplanır:

$$
\begin{array}{lccccccccc}
(a+b)^0 & & & & & 1 & & & & \\
(a+b)^1 & & & & 1 & & 1 & & & \\
(a+b)^2 & & & 1 & & 2 & & 1 & & \\
(a+b)^3 & & 1 & & 3 & & 3 & & 1 & \\
(a+b)^4 & 1 & & 4 & & 6 & & 4 & & 1 \\
\cdots & . & . & . & . & . & . & . & . & . & .
\end{array}
$$

Paskal Üçgeninde şu iki özellik vardır:

i. Her sıradaki birinci ve sonuncu sayı 1 dir.

ii. Üçgendeki diğer sayılar tam üsteki iki sayının toplanması ile elde edilir.

Örnek 1.14 $(3+2x)^5$ *ifadesinin açılımını yazınız.*

Çözüm 1.14

$$
\begin{aligned}
(3+2x)^5 &= \binom{5}{0}3^5 + \binom{5}{1}3^4(2x) + \binom{5}{2}3^3(2x)^2 + \binom{5}{3}3^2(2x)^3 \\
&\quad + \binom{5}{4}3(2x)^4 + \binom{5}{5}(2x)^5 \\
&= 32x^5 + 240x^4 + 720x^3 + 1080x^2 + 810x + 243
\end{aligned}
$$

elde edilir.

Binom Teoremini aşağıdaki gibi genelleştirebiliriz.

Teorem 1.8 $n = n_1 + n_2 + \cdots + n_k$ *olmak üzere,*

$$(a_1 + a_2 + \cdots + a_k)^n = \sum_{n_1+n_2+\cdots+n_k=n} \binom{n}{n_1, n_2, \cdots, n_k} a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k} \qquad (1.16)$$

yazılabilir. Daha önce Tanım (1.1) de belirtildiği gibi

$$\binom{n}{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

şeklindedir.

İspat 1.8 *Binom açılımından yararlanarak ispat yapılabilir.*

Örnek 1.15 $(a + b + c + d)^2$ *ifadesinin açılımını yazınız.*

Çözüm 1.15

$$\begin{aligned}
(a+b+c+d)^2 &= \sum_{n_1+n_2+n_3+n_4=2} \binom{2}{n_1, n_2, n_3, n_4} a^{n_1} a^{n_1} b^{n_2} c^{n_3} d^{n_4} \\
&= \binom{2}{2,0,0,0} a^2 + \binom{2}{1,1,0,0} ab + \binom{2}{1,0,1,0} ac \\
&\quad + \binom{2}{1,0,0,1} ad + \binom{2}{0,2,0,0} b^2 + \binom{2}{0,1,1,0} bc \\
&\quad + \binom{2}{0,1,0,1} bd + \binom{2}{0,0,2,0} c^2 + \binom{2}{0,0,1,1} cd \\
&\quad + \binom{2}{0,0,0,2} d^2 \\
&= a^2 + 2ab + 2ac + 2ad + b^2 + 2bc + 2bd + 2c^2 + 2cd + d^2
\end{aligned}$$

olduğu görülebilir.

Örnek 1.16 $(a + b + c)^5$ *ifadesinin açılımını yapmadan terim sayısını bulunuz.*

Çözüm 1.16 *Tekrarlı kombinasyonlardan istifade edeceğiz.* $n = 3$ $r = 5$ *olduğundan, terim sayısı*

$$\begin{aligned}
C(n+r-1, r) = \binom{n+r-1}{r} &= \frac{(n+r-1)!}{r!((n+r-1)-r)!} \\
&= \frac{(3+5-1)!}{5!(3-1)!} = \frac{7!}{5!2!} = 21
\end{aligned}$$

olarak bulunur.

1.1.6 Çözülmüş Problemler

1. Aşağıdaki ifadeleri kısaltınız.

$$i. \quad \frac{n!}{(n-1)!} \qquad\qquad ii. \quad \frac{(n+2)!}{n!}$$

Çözüm: *i.*

$$n! = (n-1)!n \Longrightarrow \frac{n!}{(n-1)!} = \frac{(n-1)!n}{(n-1)!} = n$$

Çözüm: *ii.*

$$\frac{(n-2)!}{n!} = \frac{(n+2)(n+1)n!}{n!} = (n+2)(n+1)$$

2. İlk ikisi harf, son üçü rakam olmak üzere beş sembolden oluşan kaç otomobil plakası yapılabilir.

Çözüm: Alfabenin bütün harflerini ve $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ rakamlarını kullanarak $(29)(29)(10)^3 = 841000$ tane plaka yapılır. Fakat rakamlar hanesinde 000 gelmeyeceğinden,

$$841000 - (29)(29) = 841000 - 841 = 840159$$

tane plaka yapmak mümkündür.

3. Bir öğretmen 6 öğrenci arasından 1 veya daha fazla öğrenciyi kaç şekilde seçebilir.

Çözüm: $1, 2, 3, 4, 5$ veya 6 öğrenci seçilecektir. Bu yüzden seçim sayısı:

$$\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 6 + 15 + 20 + 15 + 6 + 1 = 63$$

olarak bulunur.

NOT: Eğer A ve B olayları aynı zamanda olmayan iki olay ise ve A olayı toplam m yoldan, B olayıda toplam n yoldan oluyorsa;

i. A veya B olayı $m+n$ yoldan,

ii. A ve B olayı mn yoldan olabilir.

4. Dört evli çift arasından 3 kişilik bir kurul aşağıdaki hallere uygun olarak kaç şekilde seçilebilir.

i. Hepsi eşit seçilme şansına sahiptir.

ii. Kurulda 2 kadın ve 1 erkek olmak zorundadır.

iii. Bir karı-koca bir kurulda bulunamayacaktır.

Çözüm: *i.* Bir kurulda sıra önemli olmadığından problem, 8 şahıs arasından üçünün seçimidir. Dolayısıyla,

$$\binom{8}{3} = \frac{8!}{3!5!} = 56$$

bulunur.

Çözüm: *ii.* 2 kadın $\binom{4}{2} = 6$ çeşit seçilir. 1 erkek $\binom{4}{1} = 4$ şekilde seçilir. Böylece saymanın temel ilkesine göre 2 kadın ve 1 erkeğin seçilme yollarının sayısı

$$\binom{4}{2}\binom{4}{1} = (6)(4) = 24$$

olarak hesaplanır.

Çözüm: *iii.* 3 kişilik heyet olacağından bu kısıtlama altında bu kurulda 3 çiftten adam bulunacaktır. Bir karı-koca aynı kurulda bulunamayacaktır. Böylece, 3 çift, 4 çift arasından $\binom{4}{3}$ yolda seçilir. 3 çift seçildikten sonra, ilk çiftten $\binom{2}{1}$, ikinci çiftten $\binom{2}{1}$ ve üçüncü çiftten de $\binom{2}{1}$ seçim yapılabilir. saymanın temel ilkesine göre,

$$\binom{4}{3}\binom{2}{1}\binom{2}{1}\binom{2}{1} = 32$$

olarak hesaplanır.

5. CİHAN kelimesinin harflerinden ikişer ikişer alarak kaç kelime yapılabilir?

Çözüm: Beş harften ikisini seçip iki harfli kelime kuracağız. Kelimelerde harflerin sırası önemlidir. Kelimelerin sayısı

$$P(5,2) = \frac{5!}{(5-2)!} = \frac{5!}{3!} = 20$$

dır.

6. $(x+2y)^{12}$ ifadesinin açılımındaki 9.terimi bulunuz.

Çözüm: k.terimi bulma formülümüz $\begin{pmatrix} n \\ k-1 \end{pmatrix} a^{n-(k-1)} b^{k-1}$ idi. Dolayısıyla burada $n = 12$, $k = 9$, $a = x$ ve $b = 2y$ olduğu gözönüne alınırsa, 9.terim,

$$\begin{pmatrix} 12 \\ 8 \end{pmatrix} x^4 (2y)^8 = \frac{12!}{8!4!} 2^8 x^4 y^8$$

olarak bulunur.

7. Bir poligonun köşegeni komşu olmayan herhangi iki köşeyi birleştiren doğru olarak tanımlanıyor. Kenar sayısı aşağıda verilen poligonların köşegen sayısını bulunuz.

i. 5-gen *ii.* 6-gen *iii.* n-gen.

Çözüm: *i.* Önce 5 noktadan kaç doğru geçtiğini bulalım. İki nokta bir doğru belirteceğinden

$$\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \frac{5!}{3!2!} = 10$$

doğru çizilir. Bunlardan 5 tanesi poligonun kenarı geri kalan ise köşegen olacaktır. O halde

$$\begin{pmatrix} 5 \\ 2 \end{pmatrix} - 5 = 5$$

tane köşegen vardır.

Çözüm: *ii.* 6-gendeki köşegen sayısı da

$$\begin{pmatrix} 6 \\ 2 \end{pmatrix} - 6 = 9$$

olur.

Çözüm: *iii.* n-gendeki köşegen sayısı

$$\begin{aligned}\binom{n}{2} - n &= \frac{n(n-1)}{2} - n \\ &= \frac{n^2 - n - 2n}{2} \\ &= \frac{n(n-3)}{2}\end{aligned}$$

olur.

1.1.7 Problemler

1. Aşağıdaki ifadeleri sadeleştiriniz.

   i. $\dfrac{(n+1)!}{n!}$ *ii.* $\dfrac{(n-1)!}{(n+2)!}$ *iii.* $\dfrac{(n-r+1)!}{(n-r-1)!}$

2. $$2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

   olduğunu gösteriniz.

3. Bir öğrenci sınavda 10 sorudan 8 tanesini cevaplayacaktır.

   i. Kaç çeşit seçim yapılabilir.

   ii. İlk üç soruyu cevaplama şartı ile kaç çeşit seçim yapılabilir.

   iii. İlk beş sorudan en az dördünü cevaplama şartı ile kaç çeşit seçim yapılabilir.

4. $(4x - \frac{1}{2x})^{12}$ ifadesi veriliyor. 6.terimi bulunuz.

5. 10 oyuncudan kaç değişik voleybol takımı sahaya çıkarılabilir.

6. Spor toto da 13 ve $13 + 1$ tuturmak için kaç kolon oynanmalıdır.

7. TRABZON kelimesinin harflerini her düzende kullanma şartı ile, kaç farklı kelime elde edilebilir. Benzer soruyu MATEMATİK kelimesi için çözünüz.

8. Bir kutuda bulunan 12 toptan, 5 tanesi siyah, 4 tanesi beyaz ve 3 tanesi kırmızıdır. Her renkten en az bir topu içinde bulunduran 6 toptan oluşan grupların sayısı nedir?

9. $u = f(x, y, z)$ analitik fonksiyonunu düşünelim. 2.mertebeden farklı kısmi türevlerinin sayısı nedir?

10. Bir futbol takımı kaç farklı şekilde sahaya çıkabilir.

11. $1, 2, 3, 4, 5, 6$ rakamlarıyla (rakamlar tekrar edilmemek şartıyla) 400' den büyük kaç tane tek sayı yazabiliriz.

12. $1, 2, 3, 4$ rakamlarıyla elde edilebilecek 4 rakamlı sayıların toplamı nedir?

13. $(2x^3 - 3xy^2 + z^2)^6$ ifadesinin açılımında içinde x^{11} ve y^4 bulunan terimi açılım yapmadan bulunuz.

14. Türkiye'de birinci küme futbol karşılaşmaları 18 takım arasında yapılıyor. Karşılaşmalar sonunda takımlar ilk 3 dereceyi kaç değişik şekilde paylaşabilirler.

15. Aynı renkte 10 boncuk, bir bilezik üzerine kaç türlü sıralanabilir.

16. Bir doğru üzerinde A, B, C noktalarının birbirine göre kaç farklı konumu vardır.

17. Üçü birbirine paralel 5 tane doğru var. Bu doğruların kesim noktalarının sayısını bulunuz.

18. 9 çemberin ikişer ikişer kesim noktalarının sayısını bulunuz.

Chapter 2

OLASILIK TEORİSİ

Olasılık teorisi, şans oyunlarına bağlı olarak ortaya çıkmıştır. Günümüzde pek çok olay olasılık teorisi yardımıyla önceden tahmin edilmeye çalışılmakta ve hatta bu yol kullanılmadan yapılan tahminlere değer verilmemektedir. Günlük hayatımızda da bazı olayları (farkında olmadan) bu teori yardımıyla yorumlarız. Örneğin, 5 kırmızı ve 3 beyaz top bulunan bir kavanozdan rastgele çekilen bir topun kırmızı olma olasılığı beyaz olma olasılığından daha fazla olduğunu söyleriz.

Olasılık teorisi bir şeyin olasılığının düşük veya yüksek olduğunu söylemekle kalmaz ona bir nümerik değer karşılık getirir. Yukarıdaki örnekte (ileride bunun nasıl yapıldığını izah edeceğiz)

$$P(\text{kırmızı top}) = \frac{5}{8}, \quad P(\text{beyaz top}) = \frac{3}{8}$$

dir.

2.1 Deney, Örneklem Uzay, Olay

"Deney" kelimesi oldukça geniş manada kullanılır. Aşağıdakilerin herbiri deneye örnektir:

i. Bir lambanın kusurlu olup olmadığını kontrol etme,

ii. Atılan düzgün bir zarın üst yüzüne hangi sayının geleceğini gözetleme, *iii.* Günlük yağışların ölçümünü yapma,

iv. Rastgele seçilen bir kişiye yeni bir araba modelini beğenip beğenmediğini sorma.

Olasılıktaki deney, sonucu önceden tahmin edilemeyen şans olaylarıyla ilgilidir.

Tanım 2.1 *Bir deneyin mümkün bütün değerlerinin kümesine örneklem uzay denir ve S ile gösterilir. Örneklem uzayındaki herbir elemana sonuç (veya örnek nokta), altkümelerinden herbirine de olay adı verilir.*

S örneklem uzayı sonlu sayıda eleman ihtiva ediyorsa eleman sayısı $n(S)$ ile gösterilir.

Örnek 2.1 *Değerleri farklı üç düzgün para aynı anda atılıyor. Deneyin S örneklem uzayını yazınız. Olay ve sonuca birer örnek veriniz.*

Çözüm 2.1 $S = \{YTT, YYT, YTY, YYY, TTT, TYT, TYY, TTY\}$ *dır.* $A = \{YTT, TTT, TYT, TTY\}$ *bir olay* YYY *bir sonuçtur.*

Tanım 2.2 *A ve B, S örneklem uzayında ortak noktaları olmayan iki olay ise bu olaylara ayrık olay denir.*

Bir başka deyişle, $A \cap B = \emptyset$ ise A ve B ayrık olaylardır. Dolayısıyla A ve B ayrık olaylar ise $n(A \cap B) = 0$ dır.

Örnek 2.2 *Aşağıda verilen olayların ayrık olup olmadığını inceleyiniz.*

i. *Düzgün bir zar atılmasıyla ilgili bir deneyde* $A = \{\textit{tek sayı}\}$ *ve* $B = \{\textit{çift sayı}\}$ *olayları veriliyor.*

ii. *İki kişinin bir komiteye başkan seçilmesi deneyinde* $C = \{\textit{Ali'nin başkan seçilmesi}\}$ *ve* $D = \{\textit{Veli'nin başkan seçilmesi}\}$ *olayları veriliyor.*

iii. Düzgün bir zar atılması deneyinde $E = \{2'nin\ katı\}$ *ve* $F = \{3'ün\ katı\}$ *olayları veriliyor.*

Çözüm 2.2 *i. Olayları açık olarak* $A = \{1, 3, 5\}$ *ve* $B = \{2, 4, 6\}$ *şeklinde yazarız. Burada* $A \cap B = \emptyset$ *olduğundan olaylar ayrıktır.*

ii. İkiside aynı anda başkan seçilemeyeceğinden C *ve* D *ayrık olaylardır.*

iii. $E \cap F = \{6\} \neq \emptyset$ *olduğundan* E *ve* F *ayrık olaylar değildir.*

Deneylerin sonuçları eşit şanslı ve eşit şanslı olmayan diye iki gruba ayrılır. Bunun için şu örneği inceleyelim:

Örnek 2.3 *Aşağıdaki her bir deneyin örneklem uzayını yazınız ve sonuçların eşit şanslı olup olmadığını araştırınız:*

i. Düzgün bir paranın bir kez atılması,

ii. İki yüzünde 1 bulunan (6 bulunmayan) düzgün olmayan bir zarın atılması.

Çözüm 2.3 *i.* $S = \{Y, T\}$ *olup sonuçlar eşit şanslıdır.*

ii. $S = \{1, 2, 3, 4, 5\}$ *sonuçlar eşit şanslı değildir.*

2.2 Olasılığın Tanımı

2.2.1 Olasılığın Klasik Tanımı

Tanım 2.3 *Eğer* S *örneklem uzayı eşit şanslı sonlu sayıda sonuç ihtiva ederse bu uzaydaki bir* A *olayının olasılığı* $P(A)$ *ile gösterilir ve*

$$P(A) = \frac{n(A)}{n(S)} \tag{2.1}$$

olarak tanımlanır.

Açık olarak yazılırsa,

$$P(A) = \frac{\text{A' daki noktaların sayısı}}{\text{S' deki noktaların sayısı}}$$

demektir. Bu kümelerdeki noktaların sayısını bulmak için 1.Bölümde verilen metodlardan da istifade edeceğiz.

Örnek 2.4 *i. Rastgele atılan düzgün bir zarın 5 veya 6 gelme olasılığı nedir?*

ii. 10 erkek ve 6 bayanın bulunduğu bir topluluktan 5 kişilik bir kurul seçilecektir. Rastgele seçilen bir kurulda 3 bayan ve 2 erkek bulunma olasılığı nedir?

Çözüm 2.4 *i. Rastgele atılan düzgün bir zar için örneklem uzayımız*

$$S = \{1, 2, 3, 4, 5, 6\}$$

ve 5 veya 6 gelme olayı da $A = \{5, 6\}$ *dır. Olasılığın tanımından*

$$P(A) = \frac{n(A)}{n(S)} = \frac{2}{6} = \frac{1}{3} = 0.333$$

bulunur.

ii. Toplulukta toplam 16 kişi vardır. 16 kişiden 5 kişiyi

$$\binom{16}{5} = \frac{16!}{5!11!} = 4368$$

farklı şekilde seçebiliriz. O halde S örneklem uzayımızın eleman sayısı $n(S) = 4368$ *dir. 5 kişinin 3 ü bayan, 2 si erkek olacağından bunların seçilme sayısı*

$$\binom{10}{2}\binom{6}{3} = 900$$

dür. $A = \{\textit{kurulun 3 ü bayan 2 si erkek}\}$ *olarak alınırsa olasılığın tanımından*

$$P(A) = \frac{n(A)}{n(S)} = \frac{900}{4368} = \frac{75}{364}$$

elde edilir.

Diğer yandan, S ($n(S) = n$) örneklem uzay, A da bu uzayda r elemanlı bir olay olsun. Olasılığın tanımından,

$$P(A) = \frac{n(A)}{n(S)} = \frac{r}{n}$$

yazılır. A, S nin bir altkümesi olduğundan,

$$0 \leq r \leq n$$

dir. Eşitsizliğin her tarafını n ile bölersek,

$$0 \leq \frac{r}{n} \leq 1$$

ve buradan,

$$0 \leq P(A) \leq 1$$

elde edilir. Dolayısıyla bir olayın olasılığı $[0, 1]$ aralığında bir reel sayıdır. Eğer $P(A) = 0$ ise olayın olması mümkün değildir. Eğer $P(A) = 1$ ise olayın olması kesindir.

Olasılığın klasik tanımı, bir çok deney için kullanılabilmesine rağmen eşit şanslı olmayan olaylar için uygulanamaz. Aşağıdaki örneği açıklayıcı olması bakımından inceleyelim: Bir düzgün paranın 1000 kez atılması deneyinde 480 kez tura geldiği gözlenmiştir. Buna göre, $A = \{\text{tura gelme}\}$ ise

$$f_{\text{röl}}(A) = \frac{480}{1000} = 0.48$$

olur. Bu $f_{\text{röl}}(A)$ ya A nın **rölatif frekansı** denir.

Tanım olarak,

$$f_{\text{röl}}(A) = \frac{n_A}{n}$$

dır. Burada n_A, n bağımsız denemede A nın olduğu olay sayısını göstermektedir. Eğer denemenin sayısı yeteri kadar büyük alınırsa bir olayın rölatif frekansı yaklaşık olarak o olayın olasılığı olur. Yani, yeteri kadar büyük n için

$$P(A) = \frac{n_A}{n} = f_{\text{röl}}(A)$$

veya bu mana kastedilerek,

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

yazılır.

2.2.2 Olasılığın Aksiyomatik Tanımı

Tanım 2.4 *S, örneklem uzay; Σ, olayların sınıfı ve $\mathcal{R}$ reel sayılar kümesi olmak üzere $P : \Sigma \longrightarrow \mathcal{R}$ fonksiyonu verilsin. Aşağıdaki aksiyomlar sağlanırsa P fonksiyonuna olasılık fonksiyonu $P(A)$ değerine de A olayının olasılığı denir:*

i. *Her $A \in \Sigma$ için $0 \leq P(A) \leq 1$ dir.*

ii. *$P(S) = 1$ dir.*

iii. *A ve B ayrık olaylar ise $P(A \cup B) = P(A) + P(B)$ olur. Bu aksiyomlara olasılık aksiyomları adı verilir.*

Eğer S örneklem uzayı sonsuz sayıda nokta ihtiva ederse *iii.* aksiyomunun yerine aşağıdaki aksiyomu yazarız:

iii'. $A_1, A_2, \ldots$ olayları ayrık olmak üzere

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$$

dır.

2.3 Kümeler Cebirinin Olaylara Uygulanması

Olasılıkla ilgili temel teoremleri ispatlarken kullanacağımız bazı kavramları gözden geçireceğiz.

Verilen bir S örneklem uzayındaki $A, B, C, \ldots$ olaylarından yeni olaylar aşağıdaki şekilde türetilebilir:

i. A ve B olaylarının birleşimi $A \cup B$ ile gösterilir ve A veya B veya her ikisindeki sonuçları ihtiva eder.

ii. A ve B olaylarının arakesiti $A \cap B$ ile gösterilir ve hem A hemde B de olan sonuçları ihtiva eder.

iii. A olayının tümleyeni $\bar{A}$ ile gösterilir ve A da olmayan S deki bütün sonuçları ihtiva eder. Dolayısıyla $A \cap \overline{A} = \emptyset$ olduğundan A ve $\bar{A}$ olayları ayrık olaylardır.

iv. B de bulunmayıp A da bulunan sonuçları kümesine A ile B olayının farkı denir ve $A - B$ ile gösterilir. B ve $A - B$; A ve $B - A$ ayrık olaylardır.

A ve B olayları ile ilgili aşağıdaki eşitlikler vardır:

i. $A - B = A \cap \bar{B}$,

ii. $B - A = \bar{A} \cap B$,

iii. $A \cup B = A \cup (B - A) = B \cup (A - B)$.

2.4 Olasılığın Temel Teoremleri

Olasılık aksiyomları, olasılık teorisini inşa etmek ve onu istatistiğe uygulamak için bize yardımcı olacaktır. Olasılığın temel teoremlerini ispat ederken bu aksiyomlardan istifade edeceğiz.

Teorem 2.1 *Herhangi bir A olayı için $P(A) = 1 - P(\bar{A})$ olur.*

İspat 2.1 *Tümleyenin tanımına göre $S = A \cup \bar{A}$ ve $A \cap \bar{A} = \emptyset$ yazılır. Böylece,*

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}) \Longrightarrow P(A) = 1 - P(\bar{A})$$

elde edilir.

Bu teorem, $P(\bar{A})$, $P(A)$ dan daha kolay hesaplanabildiği zaman, uygulamalarda oldukça faydalıdır.

Teorem 2.2 *A ve B herhangi iki olay olsun. $A \cap B \neq \emptyset$ (ayrık olmayan olaylar) ise*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

olur.

İspat 2.2 .

Şekil 2.1:

Şekile göre, C, D, E ayrık olaylardır. $A = C \cup D$ ve $B = D \cup E$ olduğundan iii. olasılık aksiyomuna göre

$$P(A) = P(C) + P(D), \quad P(B) = P(D) + P(E)$$

yazılır. Bu eşitlikler taraf tarafa toplanırsa

$$P(A) + P(B) = P(C) + P(D) + P(D) + P(E)$$

bulunur. Burada her iki tarafa $-P(D)$ ilave edersek

$$P(A) + P(B) - P(D) = P(C) + P(D) + P(E)$$

olur. $D = A \cap B$ olduğundan eşitliğin sol tarafı $P(A) + P(B) - (A \cap B)$ dir. C, D, E ayrık olaylar ve $B = C \cup D \cup E$ olduğundan eşitliğin sağ tarafı $P(A \cup B)$ dir.Böylece

$$P(A \cup B) = P(A) + P(B) - (A \cap B)$$

elde edilir.

Sonuç: Teorem 2.2 şu şekilde genelleştirilebilir:

$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i \prec j} P(A_i \cup A_j) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap ... \cap A_n).$$

Teorem 2.3 *A ve B olayları verilsin bu takdirde*

$$P(A-B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

olur.

İspat 2.3 *B ile $\bar{B}$ ve $A \cap B$ ile $A \cap \bar{B}$ olayları ayrık olaylardır. Buna göre,*

$$A = A \cap S = A \cap (B \cup \bar{B}) = (A \cap B) \cup (A \cap \bar{B})$$

olduğundan

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A \cap B) + P(A-B)$$

yazılır. Buradan da

$$P(A-B) = P(A) - P(A \cap B)$$

elde edilir.

Teorem 2.4 *A ve B olayları verilsin $B \subseteq A$ ise*

$$P(A-B) = P(A) - P(B)$$

olur.

İspat 2.4 *$B \subseteq A$ ise $A \cap B = B$ olur. Teorem 2.3 den*

$$P(A-B) = P(A) - P(A \cap B) = P(A) - P(B)$$

elde edilir.

Sonuç : $B \subseteq A$ ise $P(B) \leq P(A)$ dır.

2.5 Şartlı (Koşullu) Olasılık

Bir B olayının olması bir başka A olayının olup olmamasını etkileyebilir. Onun için B olayının gerçekleşmesi şartı altında A olayının olasılığını bulma ihtiyacı duyulur. Bu olasılığa "B bilindiğine göre A nın şartlı olasılığı" denir ve $P(A/B)$ ile gösterilir. Bu durumda, B yeni (indirgenmiş) örneklem uzay vazifesi görür. Formül olarak

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad (P(B) \neq 0) \tag{2.2}$$

şeklinde gösterilir. Benzer şekilde,

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \quad (P(A) \neq 0) \tag{2.3}$$

yazılabilir. Özel olarak $P(A/A) = 1$ ve $P(B/B) = 1$ olur.

Örnek 2.5 *Bir zar atma denemesinde sonucun çift olduğu bilindiğine göre, 2 gelme olasılığı nedir.*

Çözüm 2.5 $A = \{\textit{2 gelmesi}\}$, $B = \{\textit{Çift gelmesi}\}$ *olayları olsun. Bir kere,*

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{3}{6} = \frac{1}{2} \quad \textit{ve} \quad P(A \cap B) = \frac{1}{6} = 0.16666$$

olduğundan,

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{(1/6)}{(1/2)} = \frac{1}{3} = 0.3333$$

elde edilir.

(2.2) ve (2.3) ifadeleri kullanılarak,

$$P(B \cap A) = P(B)P(A/B)$$

yazılabilir. Bu formül:

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)...P(A_n/A_1 \cap A_2 \cap ... \cap A_{n-1})$$

şeklinde genelleştirilebilir.

Örnek 2.6 *4 top, 4 göze yerleştiriliyor. İlk iki top ayrı gözlerde olduğuna göre, bir gözde tam 3 top bulunması olasılığı nedir?*

Çözüm 2.6 *A={Bir gözde tam 3 top bulunması} ve B={İlk iki topun ayrı gözlere gelmesi } olayları olsun. Bu takdirde B verilmişken A olayının şartlı olasılığı isteniyor. Formülümüz;*

$$P(A/B) \;=\; \frac{P(A\cap B)}{P(B)} \tag{2.4}$$

şeklindeydi. Bu formülü kullanabilmemiz için $P(A \cap B)$ *ve* $P(B)$ *değerlerini bulmak zorundayız. Bir kere herbir top dört kutuya da konabileceğinden örneklem uzayımızdaki sonuç sayı* 4^4 *olur.* $A \cap B = \{$*Bir gözde tam 3 top bulunacak ve ilk iki top ayrı gözlerde olacak*$\}$ *olduğundan* $A \cap B$*'nin eleman sayısını bulalım: 1.top 4 kutuya da konabilir. 2.top 1.topun konduğu kutu hariç üç kutuya konabilir. Geriye kalan 2 top ise ya 1.topun konduğu kutuya veya 2.topun konduğu kutuya konulabileceğinden 2 seçme şansımız vardır. O halde* $A \cap B$ *arakesitindeki olay sayısı saymanın temel ilkesine göre* $(4)(3)(2)$ *kadardır. Dolayısıyla,*

$$P(A\cap B) \;=\; \frac{n(A\cap B)}{n(S)} = \frac{(4)(3)(2)}{(4)(4)(4)(4)} = \frac{3}{32} = 0.09375$$

olarak hesaplanır. Diğer yandan, B kümesinin eleman sayısını şu muhakemeyle bulabiliriz: 1.top 4 kutuya, 2.top 3 kutuya(1.topun konduğu kutu hariç) 3.top 4 kutuya ve 4.top da 4 kutuya konabileceğinden saymanın temel ilkesine göre B kümesinin eleman sayısı $(4)(3)(4)(4)$ *olur. O halde,*

$$P(B) \;=\; \frac{n(B)}{n(S)} = \frac{(4)(3)(4)(4)}{(4)(4)(4)(4)} = \frac{3}{4} = 0.75$$

olarak hesaplanır. Bulunan bu değerler (2.4) de yerine yazılırsa

$$P(A/B) \;=\; \frac{P(A\cap B)}{P(B)} = \frac{(3/32)}{(3/4)} = \frac{1}{8} = 0.125$$

elde edilir.

2.6 Bağımsız Olaylar

İki veya daha çok olaydan birinin oluşu veya olmayışı diğerlerini hiç bir şekilde etkilemezse bu olaylara bağımsız olaylar denir. A ve B olayları bağımsız ise $P(A/B) = P(A)$ olacağı açıktır. Bu eşitliği bağımsızlığın tanımı olarak kullanacağız.

Tanım 2.5 *A ve B, S örneklem uzayında iki olay olsun. Eğer,*

$$P(A \cap B) = P(A)P(B) \tag{2.5}$$

oluyorsa, A ve B'ye bağımsız olaylar denir.

Burada şu hususu belirtmek yerinde olur: Daha önce "aynı anda gözlenmesi mümkün olmayan olaylar" olarak tanımladığımız ayrık olaylarla, bağımsız olayları karıştırmamak gerekir. Ayrık olayların ortak noktası yoktur. Halbuki bağımsız olayların ortak noktası olabilir. Aşağıdaki örneği inceleyelim:

Örnek 2.7 *Hilesiz bir para üç kez atılsın. $A = \{$Birinci atış turadır$\}$, $B = \{$İkinci atış turadır$\}$ ve $C = \{$Üç kez atışın sadece ikisinde tura gelmiştir$\}$ olayları verilsin. A ile B, A ile C ve B ile C olaylarının bağımsız olup olmadıkları hakkında ne söyleyebilirsiniz.*

Çözüm 2.7
Örneklem uzayımız $S = \{TTT, TTY, TYT, TYY, YTT, YTY, YYT, YYY\}$ şeklindedir. $A = \{TTT, TTY, TYT, TYY\}$, $B = \{TTT, TTY, YTY\}$ ve $C = \{TTY, TYT, YTT\}$ olur. Buradan $P(A) = \frac{4}{8} = \frac{1}{2}$, $P(B) = \frac{4}{8} = \frac{1}{2}$ ve $P(C) = \frac{3}{8}$ olur.

Önce A ve B olaylarını ele alalım: $P(A \cap B) = \frac{2}{8} = \frac{1}{4}$ ve $\frac{1}{4} = P(A \cap B) = P(A)P(B) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$ olup (2.5) den dolayı A ve B olayları bağımsızdır. Burada, $A \cap B = \{TTT, TTY\}$ olduğuna dikkat ediniz.

Şimdi de A ve C olaylarını inceleyelim: $A \cap C = \{TTY, TYT\}$ olup $P(A \cap C) = \frac{2}{8} = \frac{1}{4}$ ve $\frac{1}{4} = P(A \cap C) \neq P(A)P(C) = \frac{1}{2}\frac{3}{8} = \frac{3}{16}$ olup A ve C olayları bağımsız değildir(bağımlıdır).

Son olarak B ve C olaylarını inceleyelim: $B \cap C = \{TTY, YTT\}$ *olup* $P(B \cap C) = \frac{2}{8} = \frac{1}{4}$ *ve* $\frac{1}{4} = P(B \cap C) \neq P(B)P(C) = \frac{1}{2}\frac{3}{8} = \frac{3}{16}$ *olup B ve C olayları bağımsız değildir(bağımlıdır).*

Ağaç Diyagramı: Bir çok olasılık problemini çözmek için olasılık ağacı çizmek faydalıdır. Bu metod aşağıdaki örneklerde izah edilmiştir:

Örnek 1. Bir torbada 8 beyaz, 3 siyah top vardır. Ardarda iki top çekiliyor.

i. ilk topu iade etmek şartıyla,

ii. ilk topu iade etmeksizin,

çekilen toplardan birinin beyaz birinin siyah olma olasılığı nedir?

Çözüm: p ile bir beyaz ve bir siyah top çekme olasılıklarını gösterelim.

i. İade ederek çekiliş yapılıyor, W_1 birinci çekilen top beyaz, W_2 ikinci çekilen top beyaz, B_1 birinci çekilen top siyah, B_2 ikinci çekilen top siyah, olayları olsun. Birinci ve ikinci çekilişlerin sonuçları farklı dallarda gösterilir. Ağacın herhangi bir dalı boyunca görülen olaylar bağımsızdır. Böylece, $P(W_1 \cap W_2) = P(W_1)p(W_2)$ yazılır.

Şekil 2.2:

$$
\begin{aligned}
p &= P(W_1 \cap B_2) + P(W_2 \cap B_1) \\
&= \frac{24}{121} + \frac{24}{121} \\
&= \frac{48}{121}
\end{aligned}
$$

bulunur.

ii. İade etmeksizin çekiliş yapılıyor, Ağacın herhangi bir dalı boyunca görülen olaylar bağımsız olmadığı halde

$$P(W_1 \cap W_2) = P(W_1)P(W_2/W_1)$$

olduğunu kullanarak dal boyunca görülen olasılıkları çarpabiliriz. Böylece, aşağıdaki ağaç diyagramı yapılabilir.

Şekil 2.3:

$$\begin{aligned} p &= P(W_1 \cap B_2) + P(W_2 \cap B_1) \\ &= \frac{24}{110} + \frac{24}{110} \\ &= \frac{24}{55} \end{aligned}$$

bulunur.

NOT: İkinci dal üzerinde $P(W_2/W_1) = 7/10$ yerine $P(W_2) = 7/10$ yazılırsa, W_1'in meydana geldiği aşikar olarak anlşılır. Böylece diyagram aşağıdaki gibi olur.

.

Şekil 2.4:

Örnek 2. A ve B, $P(A) = 1/3$, $P(B/A) = 1/4$ ve $P(\bar{B}/\bar{A}) = 4/5$ olacak şekilde iki olay olsun. Diyagram çizerek,

i. $P(\bar{B}/A)$, *ii.* $P(A \cap B)$,
iii. $P(B)$, *iv.* $P(A \cup B)$

olasılıklarını bulunuz.

Çözüm: Ağaç diyagramı çizer ve verilen bilgileri yerleştirirsek aşağıdaki diyagram bulunur.

Şekil 2.5:

Diyagramdan, $P(\bar{A}) = 2/3$, $P(\bar{B}/A) = 3/4$ ve $P(B/\bar{A}) = 1/5$ yazılır. Çünkü dalın herbir kümesi için olasılıkların toplamı 1'dir. Şekil aşağıdaki gibi tamamlanır.

.

Şekil 2.6:

Buna göre,

i. $P(\bar{B}/A) = 3/4$

ii. $P(A \cap B) = 1/12$

iii. $P(B) = P(B \cap A) + P(B \cap \bar{A}) = 1/12 + 2/15 = 13/60$

iv. $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 1 - 8/15 = 7/15$ bulunur.

Ağaç Diyagramının Özellikleri:

i. Bir noktadan yapılan dallanmadaki olasılıkların toplamı 1'dir.

ii. Herbir dallanma sonucunda bulunan olasılıkların toplamı 1'dir. Bunları daha açık olarak şöyle görebiliriz:

.

Şekil 2.7:

i. özelliğe göre,

$$p_1 + p_2 = 1; \;\; p_3 + p_4 = 1; \;\; p_5 + p_6 = 1$$

ve ii. özelliğe göre,

$$P_1 + P_2 + P_3 + P_4 = 1$$

yazılır.

2.7 Toplam Olasılık Formülü

Tanım 2.6 *S örneklem uzayında B_1, B_2,..., B_k olayları verilsin. Eğer aşağıdaki şartlar sağlanıyorsa B_1, B_2,$\cdots$, B_k olayları S örneklem uzayının bir parçalanışıdır denir:*

i. Her $i \neq j$ için $B_i \cap B_j = \emptyset$,

ii. $\bigcup\limits_{i=1}^{k} B_i = S$,

iii. Her i için $P(B_i) > 0$ dır.

Bunu şekil olarak aşağıdaki gibi gösterebiliriz:

.

Şekil 2.8:

Örneğin, düzgün bir zar atılması deneyi için $B_1 = \{1\}$, $B_2 = \{2, 4\}$ ve $B_3 = \{3, 5, 6\}$ olayları, $S = \{1, 2, 3, 4, 5, 6\}$ örneklem uzayının bir parçalanışıdır.

A, S örneklem uzayında bir olay ve B_1, B_2,$\cdots$, B_k; S örneklem uzayının bir parçalanışı olsun. O halde,

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$

yazabiliriz. $B_1, B_2, \ldots, B_k$, S nin bir parçalanışı olduğundan $(A \cap B_1)$, $(A \cap B_2)$,$\cdots$, $(A \cap B_k)$ olaylarının tümü çifter çifter ayrıktır. Böylece, olasılığın aksiyomatik tanımından

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$$

yazılır. Ayrıca her j için $P(A \cap B_j) = P(A/B_j)P(B_j)$ olduğu dikkate alınarak

$$P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \cdots + P(A/B_k)P(B_k)$$

elde edilir. Buna **toplam olasılık formülü** denir.

B, $\bar{B}$, S örneklem uzayının parçalanışı olduğu özel durumunu göz önüne alalım. Buna göre $A = (A \cap B) \cup (A \cap \bar{B})$ olup

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

ve buradan

$$P(A) = P(A/B)P(B) + P(A/\bar{B})P(\bar{B})$$

elde edilir.

Örnek 2.8 *Dört kavanoz ile 16 siyah ve 15 beyaz bilye veriliyor. Bu kavanozların ikisinde 3 siyah, 4 beyaz bilye; bir tanesinde 9 siyah, 5 beyaz bilye; bir tanesinde de 1 siyah, 6 beyaz bilye vardır. Kavanozların herhangi biri rastgele seçiliyor ve içinden bir bilye rastgele çekiliyor. Çekilen bilyenin siyah olması olasılığını bulunuz.*

Çözüm 2.8 *Kavanozlarla ilgili olayları sırasıyla B_0, B_1, B_2 ve B_3 ile gösterelim. Böylece,*

$$B_0 = B_1 = \{\textit{Kavanozda 3 siyah, 4 beyaz bilye olması}\},$$
$$B_2 = \{\textit{Kavanozda 9 siyah, 5 beyaz bilye olması}\},$$
$$B_3 = \{\textit{Kavanozda 1 siyah, 6 beyaz bilye olması}\}$$

ve $A = \{\textit{Siyah bilye çekme}\}$ olayları olsun. Aynı bilye B_1, B_2, B_3 kümelerinde olmayacağından B_1, B_2, B_3 olayları ayrıktır. Dolayısıyla.

$$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$$

olduğundan,

$$P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + P(A/B_3)P(B_3)$$

yazabiliriz. Şimdi bu formüldeki değerleri bulmaya çalışalım:

$$P(A/B_1) = \frac{3}{7}, \quad P(A/B_2) = \frac{9}{14}, \quad P(A/B_3) = \frac{1}{7}$$

ve

$$P(B_1) = \frac{2}{4}, \quad P(B_2) = \frac{1}{4}, \quad P(B_3) = \frac{1}{4}$$

olup

$$P(A) = \left(\frac{3}{7}\right)\left(\frac{2}{4}\right) + \left(\frac{9}{14}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{7}\right)\left(\frac{1}{4}\right) = \frac{23}{56} = 0.4107142$$

olarak hesaplanır.

Şekil 2.9:

Örnek 2.9 *Yarın havanın güneşli olma olasılığı* $\frac{1}{3}$ *dür. Canan'ın tenis oynama olasılığı hava güneşli olursa* $\frac{4}{5}$, *güneşli olmazsa* $\frac{2}{5}$ *dir. Buna göre yarın Canan'ın tenis oynama olasılığını bulunuz.*

Çözüm 2.9 *"Yarın Canan'ın tenis oynaması" olayını* A *ve "yarın havanın güneşli olması" olayını da* B *ile gösterelim. Bu durumda "yarın havanın güneşli olmaması" olayı* $\bar{B}$ *olur. Buna göre*

$$P(B) = \frac{1}{3}, \quad P(\bar{B}) = \frac{2}{3}, \quad P(A/B) = \frac{4}{5}, \quad P(A/\bar{B}) = \frac{2}{5}$$

dır. O halde,

$$\begin{aligned} P(A) &= P(A/B)P(B) + P(A/\bar{B})P(\bar{B}) \\ &= \left(\frac{4}{5}\right)\left(\frac{1}{3}\right) + \left(\frac{2}{5}\right)\left(\frac{2}{3}\right) \\ &= \frac{8}{15} \\ &= 0.53333 \end{aligned}$$

olur.

2.8 Bayes Formülü

Hatırlanacağı gibi,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad P(B/A) = \frac{P(B \cap A)}{P(A)}$$

formüllerini daha önce gördük. Bu formülleri kullanarak,

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$

yazılır. Şimdi bu formülü genelleştireceğiz.

Teorem 2.5 *$B_1, B_2, \ldots, B_n$ ayrık olaylar ve $S = B_1 \cup B_2 \cup \ldots \cup B_n$ ve A, S de keyfi bir olay olsun. Bu durumda:*

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \cdots + P(A/B_n)P(B_n)}, \quad i = 1, 2, \ldots n$$

dır.

İspat 2.5 *Toplam olasılık formülünden,*

$$P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \cdots + P(A/B_n)P(B_n)$$

yazılır. Diğer yandan,

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{P(A)}$$

dır. Böylece, $i = 1, 2, \ldots n$ için,

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \cdots + P(A/B_n)P(B_n)}$$

elde edilir.

.

Şekil 2.10

Örnek 2.10 *Ayşe, Burcu ve Candan bir fabrikada bisküvit paketleme işinde çalışıyor. Onlara ayrılan bisküvit yığınının %55'ini Ayşe, %30'unu Burcu ve %15'ini de Candan paketliyor. Ayşe, Burcu ve Candan'ın bir paketteki bisküvitleri kırma olasılığı sırasıyla 0.7, 0.2 ve 0.1 dir. Kontrol sırasında kırık bisküvit bulunan bir paket bulunuyor. Bunun Ayşe tarafından paketlenmiş olması olasılığı nedir?*

Çözüm 2.10 *A={Ayşe tarafından paketlendi}, B={Burcu tarafından paketlendi}, C={Candan tarafından paketlendi} ve D={Pakette kırık bisküvit var} diyelim. Buna göre,*

$$P(A) = 0.55, \quad P(B) = 0.3, \quad P(C) = 0.1$$

ve

$$P(D/A) = 0.7, \quad P(D/B) = 0.2, \quad P(D/C) = 0.1$$

dir. $P(A/D)$ yi bulmak istiyoruz. Bayes formülüne göre

$$P(A/D) = \frac{P(D/A)P(A)}{P(D)}$$

yazılır. $P(D)$, *D nin toplam olasılığıdır. Bu ağaç diyagramından kolayca bulunur.*

Şekil 2.11

Şekili de gözönüne alarak,

$$\begin{aligned} P(D) &= P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) \\ &= (0.7)(0.55) + (0.2)(0.3) + (0.1)(0.15) \\ &= 0.46 \end{aligned}$$

ve

$$P(D/A)P(A) = (0.7)(0.55)$$

dir. Böylece,

$$P(A/D) = \frac{(0.7)(0.55)}{0.46} = 0.837$$

olur.

Örnek 2.11 *Önümüzde 1,2,3 numaralı 3 tane torba ve bu torbalarda beyaz ve kırmızı toplar şöyle dağılmış olsun:*

i. 1 beyaz 3 kırmızı top,

ii. 2 beyaz 2 kırmızı top,

iii. 3 beyaz 1 kırmızı top.

Rastgele bir torba seçiliyor ve seçilmiş olan torbadan da bir top çekiliyor. Eğer çıkan top beyaz ise, bu topun 1 numaralı torbadan çekilmiş olması olasılığı nedir?

Çözüm 2.11 *$A = \{Beyaz\ top\ çekilmesi\}$; B_1, 1 numaralı torbadan; B_2, 2 numaralı torbadan; B_3, 3 numaralı torbadan çekilmesi olayları olsun. Bizden istenen $P(B_1/A)$ sayısıdır. Hatırlanacağı gibi,*

$$P(B_1/A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A/B_1)}{P(A)}$$

dir. S örneklem uzayını torbalarla 3 ayrık sınıfa ayırdık ve bir A olayı verdik. Dolayısıyla

$$\begin{aligned} P(A) &= P(B_1 \cap A) + P(B_2 \cap A) + P(B_3 \cap A) \\ &= P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + P(A/B_3)P(B_3) \end{aligned}$$

dır. Burada $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$ ve $P(A/B_1) = \frac{1}{4}$, $P(A/B_2) = \frac{2}{4}$, $P(A/B_3) = \frac{3}{4}$ olup

$$P(A) = \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{4}\right) = \frac{1}{2} = 0.500$$

bulunur. Bunları formülde yerine yazarsak

$$P(B_1/A) = \frac{(1/3)(1/4)}{1/2} = \frac{1}{6}$$

elde edilir. $P(B_2/A) = \frac{1}{3}$ ve $P(B_3/A) = \frac{1}{2}$ olduğunu siz gösteriniz.

Bayes Formülü esas itibariyle sonuç bilindiğinde sebeblerden bir tanesinin bunu sağlamış olması olasılığını bulmamıza yarar.

2.9 Çözülmüş Problemler

1. İki zarın bir kez atıldığında üst yüzdeki sayıların toplamının 7 veya 10 olması olasılığı nedir?

Çözüm: İki zar deneyi için örneklem uzayı şöyledir:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Toplamın 7 olması olayı A, toplamın 10 olması olayı B olsun. Buna göre

$$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

ve

$$B = \{(4,6),(5,5),(6,4)\}$$

dır. Ayrıca $A \cap B = \emptyset$ olduğundan 2.2 Teoreme göre,

$$P(A \cup B) = P(A) + P(B) = \frac{6}{36} + \frac{3}{36} = \frac{9}{36} = \frac{1}{4}$$

bulunur.

2. 52 kartlık standart bir desteden rastgele bir kart çekildiğinde, bu kartın birli veya karo olması olasılığı nedir?

Çözüm: Karo çekilmesi olayı A, birli çekilmesi olayı B olsun. Bu örneklem uzayda 52 nokta bulunduğundan ve

$$P(A) = \frac{13}{52},\ P(B) = \frac{4}{52} \text{ ve } A \cap B = \{\text{Karo birli}\} \Longrightarrow P(A \cap B) = \frac{1}{52}$$

olduğundan, 2.2. Teoreme göre,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

bulunur.

3. Bir para üç kez atılsın. En az bir kez tura gelmesi olasılığını bulunuz?

Çözüm 1: En az bir kez tura gelmesi olayı A olsun. Bu deney için örneklem uzay

$$S = \{TTT, YTT, TYT, TTY, YYT, YTY, TYY, YYY\}$$

kümesidir. $n(A) = 7$ olduğundan istenen olasılık

$$P(A) = \frac{7}{8} = 0.875$$

olur.

Çözüm 2: $P(A) + P(\bar{A}) = 1$ olduğundan, A olayını düşündüğümüz gibi, A olayının tümleyenini de düşünebiliriz. $\bar{A}$ yalnız bir noktaya sahiptir. Buna göre,

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{1}{8} = \frac{7}{8} = 0.875$$

elde edilir.

4. 52 kart arasından 1 kart çekiliyor ve aşağıdaki olaylar tanımlanıyor.

$A =$ Çekilen kart kupa papazıdır.

$B =$ Çekilen kart kupadır.

$C =$ Çekilen kart maça ası veya kupadır.

$D =$ Çekilen kart maça veya kupadır.

Bu bilgilere göre,

i. $A \cup B$ ve $A \cup C$ olayları olasılıklarını,

ii. $B \cup C$ ve $B \cup D$ olayları olasılıklarını,

iii. $A \cup D$ ve $C \cup D$ olayları olasılıklarını hesaplayınız.

Çözüm: Önce problemde belirtilen olayların olasılıklarını eleman sayılarından hareketle yazalım:

$$P(A) = \frac{1}{52},\ P(B) = \frac{13}{52},\ P(C) = \frac{1}{52} + \frac{13}{52} = \frac{14}{52},\ P(D) = \frac{13}{52} + \frac{13}{52} = \frac{26}{52}$$

olur. Aranılan olasılıklar:

$i.\ P(A\cup B)=\frac{1}{52}+\frac{13}{52}-\frac{1}{52}=\frac{13}{52};\ \ P(A\cup C)=\frac{1}{52}+\frac{14}{52}-\frac{1}{52}=\frac{14}{52};$

$ii.\ P(B\cup C)=\frac{13}{52}+\frac{14}{52}-\frac{13}{52}=\frac{14}{52};\ \ P(B\cup D)=\frac{13}{52}+\frac{26}{52}-\frac{13}{52}=\frac{26}{52}=\frac{1}{2};$

$iii.\ P(A\cup D)=\frac{1}{52}+\frac{26}{52}-\frac{1}{52}=\frac{26}{52};\ \ P(C\cup D)=\frac{14}{52}+\frac{26}{52}-\frac{14}{52}=\frac{26}{52}=\frac{1}{2};$

bulunur.

5. "en az" durumu ihtiva eden problemler:

   i. Düzgün 5 zar atılıyor. En az birinin 6 gelme olasılığını bulunuz.

   ii. Düzgün n zar atılıyor. En az birinin 6 gelme olasılığını bulunuz.

   iii. Kaçtane düzgün zar atılmalıdır ki en az birinin 6 gelme olasılığı en az 0.99 olsun.

Çözüm: Verilen değerlere göre çözüm aşağıdaki şekilde elde edilir.

i. Bir atışta $P(6)=\frac{1}{6}$ ve $P(\bar{6})=\frac{5}{6}$ dır. 5 zar atıldığında,

$$\begin{aligned} P(\text{En az biri 6}) &= 1-P(\text{hiç biri 6 değil}) \\ &= 1-P(\bar{6}\bar{6}\bar{6}\bar{6}\bar{6}) \\ &= 1-\left(\frac{5}{6}\right)^5 \\ &= 0.598 \end{aligned}$$

olur.

ii. n zar atıldığında,

$$P(\text{En az biri 6}) \ = \ 1-\left(\frac{5}{6}\right)^n$$

olur.

iii. Öyle bir n arıyoruz ki

$$1-\left(\frac{5}{6}\right)^5 \geq 0.99$$

veya

$$\left(\frac{5}{6}\right)^5 \leq 0.01$$

olur. Her iki yanın logaritmasını alırsak

$$(n)log\left(\frac{5}{6}\right) \leq log(0.01)$$

ve buradan

$$n \geq \frac{log(0.01)}{log(\frac{5}{6})} \Longrightarrow n \geq 25.3$$

bulunur. Böylece n en az 26 olması gerekir.

6. 4 tanesi bozuk olan 12 nesneden 2 tanesi rasgele çekiliyor. $A = \{2$ nesne de bozuktur$\}$ ve $B = \{2$ nesne sağlamdır$\}$ olduğuna göre $P(A)$ ve $P(B)$ olasılıklarını bulunuz. $C = \{$En az 1 nesne bozuktur$\}$ olduğunda $P(C)$ olasılığını bulunuz.

Çözüm: Önce S örneklem uzayının eleman sayısını bulalım. 12 nesneden 2 nesnenin çekilmesi olan S, $\binom{12}{2} = 66$ nesneden oluşabilir. A kümesinin eleman sayısı ise, 4 bozuk nesneden 2 nesnenin çekilmesi olan $\binom{4}{2} = 6$ olur. B kümesinin eleman sayısı da 8 bozuk olmayan nesneden 2 bozuk olmayan nesnenin çekilmesi olan $(\,8\ cr2\,) = 28$ dir. Buna göre,

$$P(A) = \frac{6}{66} \text{ ve } P(B) = \frac{28}{66}$$

olarak bulunur.

C, B kümesinin tümleyenidir. Dolayısıyla $S = C \cup B$ olur. Buradan

$$P(S) = P(C) + P(B) \Longrightarrow P(C) = 1 - P(B) = 1 - \frac{28}{66} = 1 - \frac{14}{33} = \frac{19}{33}$$

elde edilir.

7. 6 istatikçi ve 5 biyologdan 7 kişilik bir komisyon seçilecektir.

i. Komisyonda 4 istatistikçi,

ii. Komisyonda en az 4 istatistkçi bulunması olasılığı nedir?

Çözüm: İstenen olasılıklar:

i.

$$P_1 = \frac{\binom{6}{4}\binom{5}{3}}{\binom{11}{7}} = \frac{5}{11}$$

ii.

$$P_2 = \frac{\binom{6}{4}\binom{5}{3} + \binom{6}{5}\binom{5}{2} + \binom{6}{6}\binom{5}{1}}{\binom{11}{7}} = \frac{43}{66}$$

olur.

8. Bir otobüs kazasında arabadaki 20 yolcudan 4'ü yaralanmıştır. Otobüste 4 tenis oyuncusu olduğu bilinmektedir.Yaralıların tenis oyuncuları olması olasılığı nedir?

Çözüm: Yaralıların tenis oyuncuları olmaları olasılığı:

$$P = \frac{\binom{4}{4}}{\binom{20}{4}} = \frac{1}{\binom{20}{4}} = \frac{1}{4845} = 0.0002063$$

dir.

9. Bir kavanozda 1 den 8'e kadar numaralanmiş 8 top vardır. 4 top (yerine koymaksızın) aynı anda çekiliyor. Çekilen en küçük sayının 2 olması olasılığı nedir?

Çözüm: 1 olmayacağına ve 2 de kesinlikle olacağına göre 6 top arasından 3 tane çekilecektir. Buna göre,

$$P = \frac{\binom{6}{3}}{\binom{8}{4}} = \frac{2}{7}$$

bulunur.

10. Oyuncuların toplam sayısını azaltmak amacıyla, $2n$ takım iki eşit gruba ayrılıyor. En kuvvetli iki takımın ,

i. Farklı gruplarda,

ii. Aynı grupta bulunmaları olasılığı nedir?

Çözüm:

i. En kuvvetli iki takımın farklı gruplarda bulunması olasılığı,

$$P = \frac{\binom{2}{1}\binom{2n-2}{n-1}}{\binom{2n}{n}} = \frac{n}{2n-1}$$

olur.

ii. En kuvvetli iki takımın aynı grupta bulunması olasılığı,

$$P = \frac{\binom{2}{2}\binom{2n-2}{n-2}}{\binom{2n}{n}} = \frac{n-1}{2(2n-1)}$$

olur.

11. Her biri 1 lira değerinde 5 bilet, 3 lira değerinde 3 bilet, 5 lira değerinde 2 bilet arasından rastgele 3 bilet çekiliyor.

i. En az ikisinin aynı fiatta,

ii. Üçünün toplam fiatının 7 lira olması olasılığını bulunuz.

Çözüm: Aşağıdaki olayları tanımlayabiliriz:

$A = \{$En az ikisinin aynı fiyatta olması$\}$ ise

$\bar{A} = \{$Tümünün farklı fiyatta olması$\}$ olur. Buna göre,

i.

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{\binom{5}{1}\binom{3}{1}\binom{2}{1}}{\binom{10}{3}} = \frac{3}{4}$$

ii.

$$P = 1 - P(\bar{A}) = 1 - \frac{\binom{5}{1}\binom{3}{2} + \binom{2}{1}\binom{5}{2}}{\binom{10}{3}} = \frac{7}{24}$$

olarak bulunur.

12. Rastgele alınan 9 kişiden en az 2 kişinin aynı ayda doğmuş olması olasılığı nedir?

Çözüm: $A = \{$En az iki kişinin aynı ayda doğması$\}$ ise $\bar{A} = \{$Herhangi ikisinin aynı ayda doğmaması$\}$ olur. $P(\bar{A})$'yı hesaplamak daha kolaydır. Böylece,

$$\begin{aligned} P(\bar{A}) &= \frac{12(12-1)(12-2)(12-3)(12-4)(12-5)(12-6)(12-7)(12-8)}{12^9} \\ &= \frac{(12)(11)(10)\cdots(4)}{12^9} = \frac{12!}{3!12^9} \end{aligned}$$

bulunur. İstenen olasılık

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{12!}{3!12^9}$$

olur.

13. Bir şiirin 3 kopyası 3 şahsa, diğer bir şiirinde 3 kopyası diğer 3 şahsa gönderilecektir. Bu iki şiirin 6 kopyası bu 6 şahsın adreslerinin yazıldığı zarflara rastgele konuyor. Şiirlerin doğru adrese gönderilme olasılığı nedir?

Çözüm: $AAA\ BBB$ ile şiirlerin kopyalarını gösterelim. Bunların farklı permütasyonlarının sayısı,

$$\binom{6}{3,3} = \frac{6!}{3!3!} = 20$$

olur.

1	2	3	4	5	6
A	A	A	B	B	B

olacağından istenen olasılık,

$$P = \frac{1}{20}$$

olur.

14. Üç koşucu yarışmaktadır. Birincinin kazanma şansı ikincinin $1/3$ katı, ikincinin kazanma şansı üçüncünün $1/4$ katı ise her birinin kazanma şansını hesaplayınız?

Çözüm: Yarışçıları sırasıyla A, B, C ile gösterelim. $P(A) = \frac{1}{2}P(B)$, $P(B) = \frac{1}{4}P(C)$ yazılabilir. Öteyandan

$$
\begin{aligned}
P(A) + P(B) + P(C) = 1 \quad &\Longrightarrow \quad \frac{1}{3}P(B) + P(B) + 4P(B) = 1 \\
&\Longrightarrow \quad P(B) = \frac{3}{16}
\end{aligned}
$$

elde edilir. Bundan yararlanarak $P(A) = \frac{1}{16}$ ve $P(C) = \frac{12}{16} = \frac{3}{4}$ bulunur.

15. A, B ve C aynı olasılık uzayında tanımlanmış olaylar olsun. Aşağıdaki bilgiler verilmektedir: A ve B bağımsız, A ve C bağımsız,

$$
P(A \cap B) = \frac{1}{5}, \quad P(A \cup C) = \frac{4}{5}, \quad P(B \cap C) = \frac{1}{6}, \quad P(C/B) = \frac{1}{3}
$$

Bu bilgileri kullanarak,

i. $P(A)$, $P(B)$ ve $P(C)$ değerlerini bulunuz.

ii. B ve C olaylarının bağımsız olup olmadığını belirtiniz.

Çözüm

i.

$$
\begin{aligned}
P(C/B) = \frac{P(B \cap C)}{P(B)} \quad &\Longrightarrow \quad P(B \cap C) = P(B)P(C/B) \\
&\Longrightarrow \quad \frac{1}{6} = P(B)\frac{1}{3} \Longrightarrow P(B) = \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B) \Longrightarrow \frac{1}{5} = P(A)\frac{1}{2} \Longrightarrow P(A) = \frac{2}{5} \\
P(A \cup C) &= P(A) + P(C) - P(A \cap C) \\
&= P(A) + P(C) - P(A)P(C) \\
&= P(A) + P(C)[1 - P(A)] \\
\frac{4}{5} &= \frac{2}{5} + P(C)[1 - \frac{2}{5}] \\
&\Longrightarrow \frac{2}{5} = P(C)\frac{3}{5} \Longrightarrow P(C) = \frac{2}{3}
\end{aligned}
$$

elde edilir.

ii. B ve C olaylarının bağımsız olması için $P(B \cap C) = P(B)P(C)$ olması gereklidir. Oysa

$$P(B \cap C) = \frac{1}{6} \neq P(B)P(C) = \left(\frac{1}{2}\right)\left(\frac{2}{3}\right) = \frac{1}{3}$$

bulunur. Dolayısıyla B ve C olayları bağımlıdır.

16. Bir zar atılıyor. Gelen yüzün tek sayı olduğu bilindiğine göre bu sayının 5 olma olasılığını hesaplayınız.

Çözüm: Tek sayı gelme olayı B_1 ve 5 gelmesi olayıda B_2 olsun. B_1 bilindiğine göre bunun B_2 olması olasılığı

$$P(B_2/B_1) = \frac{P(B_1 \cap B_2)}{P(B_1)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}$$

olur veya olayların eleman sayısını hesaplayarak da

$$P(B_2/B_1) = \frac{n(B_1 \cap B_2)}{n(B_1)} = \frac{1}{3}$$

olduğu görülebilir.

17. Birinde üç kırmızı ve iki beyaz, diğerinde iki kırmızı ve beş beyaz bilye bulunan iki kutudan rastgele, önce bir kutu sonrada bu kutudan bir bilye seçilip ikinci kutuya konulduktan sonra ikinci kutudan rastgele yeni bir bilye seçilse; birinci ve ikinci kutudan seçilen iki bilyeninde aynı renkte(KK veya BB) olma olasılığını bulunuz.

Çözüm: Şu dört hali incelememiz gerekir.

i. Birinci kutu seçildiğinde çekilen bilyenin kırmızı olması ve ikinci kutuya konduktan sonra ikinci kutudan kırmızı bilye çekilmesi,

ii. Birinci kutu seçildiğinde çekilen bilyenin beyaz olması ve ikinci kutuya konduktan sonra ikinci kutudan beyaz bilye çekilmesi,

iii. İkinci kutu seçildiğinde çekilen bilyenin kırmızı olması ve birinci kutuya konduktan sonra birinci kutudan kırmızı bilye çekilmesi,

iv. İkinci kutu seçildiğinde çekilen bilyenin beyaz olması ve birinci kutuya konduktan sonra birinci kutudan beyaz bilye çekilmesi, Bu olayları sırasıyla L_i, L_{ii}, L_{iii} ve L_{iv} ile kırmızı bilye çekilmesini K, beyaz bilye çekilmesini B, birinci kutuyu A_1, ikinci kutuyu A_2 ile gösterelim. Böylece,

$$\begin{aligned}
P(L_i) &= P(A_1)P(K/A_1)P(K/K\cap A_1) = \left(\frac{1}{2}\right)\left(\frac{3}{5}\right)\left(\frac{3}{8}\right) = \frac{9}{80} \\
P(L_{ii}) &= P(A_1)P(B/A_1)P(B/B\cap A_1) = \left(\frac{1}{2}\right)\left(\frac{2}{5}\right)\left(\frac{6}{8}\right) = \frac{12}{80} \\
P(L_{iii}) &= P(A_2)P(K/A_2)P(K/K\cap A_2) = \left(\frac{1}{2}\right)\left(\frac{2}{7}\right)\left(\frac{4}{6}\right) = \frac{8}{84} \\
P(L_{iv}) &= P(A_2)P(B/A_2)P(B/B\cap A_2) = \left(\frac{1}{2}\right)\left(\frac{5}{7}\right)\left(\frac{3}{6}\right) = \frac{15}{84}
\end{aligned}$$

olarak hesaplanır. Dolayısıyla

$$\begin{aligned}
P(L_i \text{ veya } L_{ii} \text{ veya } L_{iii} \text{ veya } L_{iv}) &= P(L_i) + P(L_{ii}) + P(L_{iii}) + P(L_{iv}) \\
&= \frac{9}{80} + \frac{12}{80} + \frac{8}{84} + \frac{15}{84} = \frac{901}{1680}
\end{aligned}$$

olur.

NOT: Yukarıdaki problemde $P(K/K\cap A_1)$ ile "1. kutudan kırmızı çekildiği bilindiğine göre ikinci kutudan kırmızı çekme olasılığı" anlaşılacaktır. Benzer olarak $P(B/B\cap A_2)$ ile de "2. kutudan beyaz çekildiği bilindiğine göre birinci kutudan beyaz çekme olasılığı" anlaşılacaktır.

18. Eğer A ve B iki bağımsız olay ise $\bar{A}$ ve $\bar{B}$ (A nın tümleyeni ve B nin tümleyeni) olaylarınında bağımsız olacaklarını gösteriniz.

Çözüm: A ve B bağımsız iki olay ise $P(A\cap B) = P(A)P(B)$ olur. Buradan

$$\begin{aligned}
P(\bar{A}\cap\bar{B}) = P(A\,\bar{\cup}\,B) &= 1 - P(A\cup B) = 1 - [P(A) + P(B) - P(A\cap B)] \\
&= 1 - P(A) - P(B) + P(A)P(B) \\
&= 1 - P(A) - P(B)[1 - P(A)] \\
&= [1 - P(A)][1 - P(B)] = P(\bar{A})P(\bar{B})
\end{aligned}$$

elde edilir.

19. İki kavanozdan birincisinde 4 siyah ve 6 kırmızı top, ikincisinde 3 siyah ve 2 kırmızı top vardır. Rastgele bir kavanoz seçer ve seçilen kavanozdan yine rastgele bir top seçilirse:

i. Siyah bir top çekilmiş olması,

ii. Siyah topun çekildiği bilindiğinde birinci kavanozun seçilmiş olması olasılığı nedir?

Çözüm: $B_1 = \{\text{I. kavanozun çekilmesi}\}$, $B_2 = \{\text{II. kavanozun çekilmesi}\}$, $A = \{\text{Siyah topun çekilmesi}\}$ olarak tanımlansın.

i. $P(Siyah) = P(\text{I. kavanoz ve siyah}) + P(\text{II. kavanoz ve siyah})$

$$\begin{aligned} P(A) = P(B_1 \cap A) + P(B_2 \cap A) &= P(A/B_1)P(B_1) + P(A/B_2)P(B_2) \\ &= \left(\frac{1}{2}\right)\left(\frac{4}{10}\right) + \left(\frac{1}{2}\right)\left(\frac{3}{5}\right) = \frac{10}{20} \end{aligned}$$

bulunur.

ii. $P(B_1/A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A/B_1)}{P(A)} = \frac{4/20}{10/20} = \frac{4}{10}$ elde edilir.

Şekil 2.12:

20. Arkadaşımız ziyaret etmek üzere A, B ve C illerinden birini seçti. A ilinde yağmur yağması olasılığı $1/2$, B ilinde $1/4$ ve C ilinde $1/6$ idi. Seyahatten döndüğünde arabası çamurlu olduğuna göre C ilini ziyaret etmiş olması olasılığını bulunuz.

Çözüm: A_1, A iline gitmesi olayı; A_2, B iline gitmesi olayı; ve A_3, C iline gitmesi olayı olsun. $P(A_1) = P(A_2) = P(A_3) = 1/3$ olacağı açıktır. K havanın yağmurlu olması olayı ise

$$P(K/A_1) = \frac{1}{3}, \quad P(K/A_2) = \frac{1}{4}, \quad P(K/A_3) = \frac{1}{6}$$

olur. Bayes formülünü kullanırsak,

$$\begin{aligned} P(A_3/K) &= \frac{P(A_3)P(K/A_3}{P(A_1)P(K/A_1) + P(A_2)P(K/A_2) + P(A_3)P(K/A_3)} \\ &= \frac{(1/3)(1/6)}{(1/3)(1/3) + (1/3)(1/4) + (1/3)(1/6)} = \frac{2}{9} \end{aligned}$$

sonucu bulunur.

2.10 Problemler

1. Bir torbada 4 beyaz, 6 siyah ve 2 kırmızı bilye bulunmaktadır. Gelişi güzel çekilen bir bilyenin beyaz olma olasılığı nedir.

2. Bir kutuda 6 kırmızı, 4 yeşil, 3 beyaz top vardır. Çekilen üç toptan 2 tanesi kırmızı ve bir tanesinin yeşil olma olasılığı nedir.

3. 52 lik bir desteden gelişigüzel seçilen 8 kağıtın 3 tanesi sinek, 3 tanesinin maça olması ihtimali nedir.

4. 4 ciltlik bir kitap gelişigüzel diziliyor. Bunların doğru sıra içinde olmaları ihtimali nedir.

5. Bir hayvanat bahçesinde bulunan 10 kediden 3 tanesinin gözleri mavidir. rastgele seçilen 2 kedinin:

   i. ikisinin de mavi gözlü olması,

   ii. Hiçbirinin mavi gözlü olmaması,

   iii. En az birinin mavi gözlü olması olasılıklarını bulunuz.

6. 52 adet kart 1 den 52 ye kadar numaralanıyor ve karıştırıldıktan sonra masaya konuyor. Destedeki ilk üç kağıtın küçükten büyüğe doğru bir sırada olması olasılığı nedir.

7. Dört ayrık olaydan meydana gelen $S = \{A, B, C, D\}$ örneklem uzayı verildiğinde, aşağıdaki hangi fonksiyon S örneklem uzayı üzerinde olasılık fonksiyonu tanımlar?

   i. $F(A) = \frac{1}{2}$, $F(B) = \frac{1}{3}$, $F(C) = \frac{1}{4}$, $F(D) = \frac{1}{5}$,

   ii. $H(A) = \frac{1}{2}$, $H(B) = \frac{1}{4}$, $H(C) = \frac{1}{8}$, $H(D) = \frac{1}{8}$

8. 52'lik bir kağıt destesinden yerine koymaksızın 3 kart çekiliyor. Bu üç kartında as olması olasılığı nedir?

9. $1, 2, 3, 4, 5$ rakamlarıyla tekrar etmeksizin yazılabilen 5 rakamlı bir sayının

   i. Tek sayı,

   ii. 5 ile bölünebilir bir sayı olması olasılığı nedir?

10. S örneklem uzayında bir F olayı $P(F) > 0$ olasılığına sahip olsun.

    i. $P(A/F) = 1$,

    ii. $F \subset E$ ise $P(E/F) = 1$

    iii. $F \subset E$ ise $P(F/E) = P(F)/P(E)$ olduğunu gösteriniz?

11. Evli bir çift iki çocuk sahibi olmayı ümit ediyorlar. Çocuğun kız olması olasılığının $1/2$ olduğunu kabul edelim,

    i. İlk çocuğun kız olması,

    ii. Her ikisinde aynı cinsiyetten olması,

    iii. Çocukların farklı cinsiyetten olmaları,

    iv. Çocukların ikisinin de oğlan olması olasılığı nedir?

12. Bir alarm sisteminin tehlikeli olduğu zaman çalışması olasılığı 0.98; tehlike olmadığında alarm vermemesi olasılığı 0.99 ve tehlike olması olasılığı 0.001 dir.

    i. Sistem alarm verdiğine göre, tehlike nedeniyle çalışmış olması olasılığı nedir?

    ii. Tehlike olması ve sistemin alarm vermemesi olasılığı nedir?

13. E, S örneklem uzayında herhangi bir olay ise E ve S nin bağımsız olduğunu gösteriniz. E ve $\emptyset$ bağımsız olabilir m?

14. A ve B, S örneklem uzayında iki olay ve $P(A) = a$, $P(B) = b$ ise

    $$P(A/B) \geq \frac{a+b-1}{b}$$

    olduğunu gösteriniz.

15. 1 ile 30 arasında doğal sayılardan oluşan bir küme var. Bu kümeden herhangi bir sayı seçiliyor. Bu sayının 5 ile bölünebilme olasılığı nedir.

16. Televizyondaki bir çekilişte, bir büyük küre içinde üzerinde, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 rakamlarından biri yazılı olan 10 tane top vardır. Makine çalıştığı zaman bu toplar karışır ve bir tanesi düşer. Bu düşen top üzerinde yazılı olan sayının;

    i. 5 ten küçük olması olasılığı,

    ii. 3 ten küçük olması olasılığı,

    iii. 6 dan büyük olması olasılığı,

    iv. Çift sayı olması olasılığı nedir.

17. Ekrem ve Murat adlarında iki avcı bir tavşan avlıyorlar. Ekrem'in tavşanı vurma olasılığı 3/4, Murat'ın ise 4/5 dir. Tavşanın vurulması olasılığını bulunuz.

Chapter 3

RASTLANTI DEĞİŞKENLERİ

3.1 Rastlantı Değişken Kavramı

Bazı denemelerin örneklem uzayının elemanları sayılardır. Örneğin bir zarın atılmasıyla ilgili örneklem uzay $S = \{1, 2, 3, 4, 5, 6\}$ dır. Bazı örneklem uzaylarda ise elemanlar sayısal olarak değerlendirilemez. Örneğin bir paranın iki kez atılması ile ilgili örneklem uzay $S = \{YY, TT, YT, TY\}$ dır. Bu bölümde örneklem uzayının her noktasına bir reel sayı karşılık getireceğiz.

Tanım 3.1 *Bir deneyin S örneklem uzayı üzerinde tanımlı reel değerli X fonksiyonuna rastlantı değişkeni denir.*

Örnek 3.1 *Bir paranın iki kez atılması deneyini düşünelim. Bunun için bir rastlantı değişkeni bulunuz.*

Çözüm 3.1 *Bu deneyin örneklem uzayı*

$$S = \{YY, TT, YT, TY\}$$

şeklindedir. X rastlantı değişkeni "turaların sayısını" göstermek üzere

$$X(YY) = 0,\ X(TT) = 2,\ X(YT) = 1,\ X(TY) = 1$$

olur. Böylece X rastlantı değişkeninin mümkün değerleri 0, 1 ve 2 dir.

Örnek 3.2 *Tura gelinceye kadar bir paranın atılması olayını gözönüne alalım. X rastlantı değişkeni ilk tura gelinceye kadar yapılan atışların sayısı olduğuna göre bu rastlantı değişkeninin alacağı değerleri belirtiniz.*

Çözüm 3.2

2cm	*Atış sayısı*
T	1
YT	2
YYT	3
$\vdots$	$\vdots$

olur. Görüldüğü gibi mümkün sonuçların sayısı sonlu değildir. Bununla beraber bütün mümkün sonuçların cümlesini numaralayabiliriz. Buradaki X rastlantı değişkeni 1, 2, 3, ... sayılabilir sonsuz değer alır.

3.2 Kesikli Rastlantı Değişkeninin Olasılık Fonksiyonu ve Dağılımı

Tanım 3.2 *X bir rastlantı değişkeni olsun. X değişkeninin mümkün değerlerinin sayısı sonlu veya sayılabilir sonsuz ise X rastlantı değişkenine kesikli rastlantı değişkeni denir.*

Kesikli bir rastlantı değişkeninin olasılık fonksiyonu, mümkün değerlerin, karşılık gelen olasılıklarla birlikte belirtilmesidir. Rastlantı değişkenleri büyük harflerle ve belirtilen değerlerinide küçük harflerle göstereceğiz.

Tanım 3.3 *X sayılabilir sayıdaki $x_1, x_2, \ldots,$ değerlerini alan rastlantı değişkeni ve bunlara karşılık gelen olasılıklar*

$$f(x_i) = P(X = x_i), i = 1, 2, \ldots,$$

olsun. Ayrıca aşağıdaki şartlar sağlanırsa, $f(x)$ fonksiyonuna X 'in **olasılık fonksiyonu** *denir:*

i. *Bütün* x *değerleri için* $f(x) \geq 0$,

ii. $\sum_{i=1}^{\infty} f(x_i) = 1$

O halde X rastlantı değişkeninin olasılık fonksiyonu aşağıdaki gibi yazılır:

$X = x$	x_1	x_2	$\cdots$
$f(x) = P(X = x)$	$f(x_1)$	$f(x_2)$	$\cdots$

Örnek 3.3 *Bir paranın iki kez atılması deneyini düşünelim. X rastlantı değişkeni, bulunan turaların sayısı olmak üzere X in olasılık fonksiyonunu bulunuz.*

Çözüm 3.3 *Bu deney için örneklem uzay*

$$S = \{YY, TT, YT, TY\}$$

şeklindedir. X değişkeninin mümkün değerleri 0,1 ve 2 olur. X 'in x değerini alma olasılığı $f(x)$ olsun. O halde

$$f(x) = P(X = x)$$

olur. Böylece

$$\begin{aligned} f(0) &= P(X = 0) = \frac{1}{4} \\ f(1) &= P(X = 1) = \frac{1}{2} \\ f(2) &= P(X = 2) = \frac{1}{4} \end{aligned}$$

olur. Burada $f(0) + f(1) + f(2) = 1$ olduğuna dikkat ediniz. X 'in mümkün değerleri ve karşılık gelen olasılıklar (olasılık fonksiyonu) aşağıdaki tabloda gösterilmiştir:

$X = x$	*0*	*1*	*2*
$f(x) = P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Örnek 3.4 *Bir para tura gelinceye kadar atılıyor. X atışların sayısını gösteren rastlantı değişkeni olsun. X rastlantı değişkeninin olasılık fonksiyonunu bulunuz.*

Çözüm 3.4 *T ile turayı, Y ile yazıyı gösterelim. Bu deney için mümkün sonuçlar ve karşılık gelen olasılıklar aşağıdadır:*

$$\begin{array}{lcc}
\textit{1. atış} & T & \frac{1}{2} \\
\textit{2. atış} & YT & \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4} \\
\textit{3. atış} & YYT & \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8} \\
\vdots & \vdots & \vdots \\
\textit{N. atış} & \underbrace{YY\ldots Y}_{N-1}T & \left(\frac{1}{2}\right)^{N-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{N}
\end{array}$$

Böylece X rastlantı değişkeninin olasılık fonksiyonu

$X = x$	1	2	$\cdots$	N	$\cdots$
$f(x) = P(X = x)$	$\frac{1}{2}$	$(\frac{1}{2})^2$	$\cdots$	$(\frac{1}{2})^N$	$\cdots$

dır. Gerçekten

i. *Her x için $f(x) > 0$ dır.*

ii.

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x$$

sonsuz geometrik serinin toplamı

$$\sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

dir.

X rastlantı değişkeninin olasılık fonksiyonu

$$f(x) = \left(\frac{1}{2}\right)^x, \ x = 1, 2, \ldots$$

şeklinde de yazılabilir.

X rastlantı değişkeninin mümkün değerlerinin birine eşit olması olasılığı ile ilgilendiğimiz gibi eşit veya küçük olması olasılığı ile de ilgileneceğiz. Bu toplamalı olasılığa X rastlantı değişkeninin dağılım fonksiyonu denir ve aşağıdaki gibi tanımlanır:

Tanım 3.4 *X, $f(x)$ olasılık fonksiyonuna sahip kesikli rastlantı değişkeni ve*

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

olsun. $F(x)$'e X rastlantı değişkeninin dağılım fonksiyonu denir.

Örnek 3.5 *Düzgün bir zar bir kez atılıyor. Üste gelen yüzdeki noktaların olasılık ve dağılım fonksiyonunu bulunuz. Bu fonksiyonların grafiklerini çiziniz.*

Çözüm 3.5 *Bu deney için örneklem uzay $S = \{1,2,3,4,5,6\}$ şeklindedir. Örneklem uzayda 6 nokta var ve her biri $\frac{1}{6}$ olasılıkla elde edilir. O halde olasılık fonksiyonu*

$X = x$	*1*	*2*	*3*	*4*	*5*	*6*
$f(x) = P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(x) = P(X \leq x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

olur.

Şekil 3.1:

Uygulamalarda, X in değer aldığı $a < x \leq b$ uygun aralığı için $P(a < X \leq b)$ olasılıkları ile ilgilenilir. Bu olasılık

$$P(a < X \leq b) = F(b) - F(a)$$

dır. Gerçekten

$$\begin{aligned} F(b) &= P(X \leq b) \\ &= P(X \leq a) + P(a < X \leq b) \\ &= F(a) + P(a < X \leq b) \end{aligned}$$

olup ilk ve son terimden istenen eşitlik elde edilir. Dağılım fonksiyonu da gözönüne alınarak

$$P(a < X \leq b) = \sum_{a < x_i \leq b} f(x_i)$$

yazılır (burada $<$ ve $\leq$ işaretlerine dikkat ediniz).

Örnek 3.6 *Düzgün iki zarın atılması deneyinde, X rastlantı değişkeni elde edilen sayıların toplamı olmak üzere toplamın enaz 4 ve en çok 8 olma olasılığı nedir.*

Çözüm 3.6 *Bu X değişkeninin olasılık ve dağılım fonksiyonu aşağıda gösterilmiştir.*

$X = x$	2	3	4	5	6	7	8	9	10	11	12
$f(x) = P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$
$F(x) = P(X \leq x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

Buna göre

$$P(3 < X \leq 8) = F(8) - F(3) = \frac{26}{36} - \frac{3}{36} = \frac{23}{36} = 0.63888$$

olur.

3.3 Sürekli Rastlantı Değişkeninin Olasılık Fonksiyonu ve Dağılımı

Tanım 3.5 *X rastlantı değişken olsun. X değişkeninin mümkün değerleri bir aralık veya aralıkların bir kolleksiyonu ise X değişkenine sürekli rastlantı değişkeni denir.*

Tanım 3.6 *X, $(-\infty, \infty)$ aralığında tanımlanan sürekli rastlantı değişken olsun. Eğer aşağıdaki şartlar sağlanıyorsa $f(x)$ fonksiyonuna X rastlantı değişkeninin olasılık yoğunluk fonksiyonu denir:*

i. $-\infty < x < \infty$ *için* $f(x) \geq 0$,

ii. $f(x)$ *eğrisi altında kalan ve x-ekseni ile sınırlanan alan 1 'e eşittir. Yani,*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

dir.

Tanım 3.7 *X, $f(x)$ olasılık yoğunluk fonksiyonuna sahip sürekli rastlantı değişkeni ve*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

olsun. $F(x)$'e X rastlantı deişkeninin dağılım fonksiyonu dnir.

Tanımdaki $F(x) = \int_{-\infty}^{x} f(s)ds$ eşitliğinden türev alınırsa,

$$F'(x) = f(x)$$

olduğu görülür.

Ayrıca $f(x)$, olasılık yoğunluk fonksiyonu ise X sürekli rastlantı değişkeninin a ve b $(-\infty < a < b < \infty)$ arasında bulunması olasılığı $P(a < X < b)$ ile gösterilir. Bu olasılık $f(x)$ eğrisi, x-ekseni ve $x = a$, $x = b$ doğruları ile sınırlanan alana eşittir, yani

$$P(a < X < b) = \int_{a}^{b} f(s)ds = F(b) - F(a)$$

dır. X sürekli rastlantı değişkeni için $a < X < b,\ a \leq X < b,\ a < X \leq b,\ a \leq X \leq b$ aralıklarına karşılık gelen olasılık aynıdır. Bu durumun kesikli rastlantı değişkenleri için aynı olmadığına dikkat ediniz.

Örnek 3.7 $f(x) = 1,\ 0 < x < 1$ *olasılık yoğunluk fonksiyonu verilsin. Aşağıdaki olasılıkları bulunuz.*

i. $P(0.25 < X < 0.75)$

ii. $P(X > 0.25)$

Çözüm 3.7 $f(x) = 1,\ 0 < x < 1$ *fonksiyonunun grafiği Şekilde 3.2 de çizilmiştir.* $x = 0,\ x = 1$ *doğruları ve* $f(x) = 1$ *doğrusu ve* x*-ekseni tarafından sınırlanan bölgenin alanı 1 dir. Böylece*

i. $P(0.25 < X < 0.75) = \int_{0.25}^{0.75} dx = x\big|_{0.25}^{0.75} = 0.75 - 0.25 = 0.50$ *olup taralı alandır.*

ii. $P(X > 0.25) = \int_{0.25}^{1} dx = x\big|_{0.25}^{1} = 1 - 0.25 = 0.75$ *olur.*

Şekil 3.2:

Örnek 3.8 *Belli bir tipteki elektrik ampullerinin dayanma süresi (saat olarak)* X *olsun.* X *'i sürekli bir rastlantı değişkeni olarak kabul edelim.* X *değişkeninin olasılık yoğunluk fonksiyonu*

$$f(x) = \begin{cases} \frac{a}{x^3} & , \quad 1500 \leq x \leq 2500 \\ 0 & , \quad x < 1500 ve x > 2500 \end{cases}$$

ise a sabitinin değerini bulunuz.

Çözüm 3.8 *Problemde $x < 1500$ ve $x > 2500$ için $f(x) = 0$ olarak veriliyor. $f(x)$ olasılık yoğunluk fonksiyonu olduğundan a sabitini hesaplamak için*

$$\int_{-\infty}^{\infty} f(x)dx = \int_{1500}^{2500} \frac{a}{x^3}dx = 1$$

eşitliği kullanılır. Buradan

$$1 = \int_{1500}^{2500} \frac{a}{x^3}dx = -\frac{a}{2x^2}\Big|_{1500}^{2500} = -\frac{a}{2(2500)^2} + \frac{a}{2(1500)^2}$$

olup $a = 7\,031\,250$ olarak bulunur.

Teorem 3.1 *$F(x)$, X 'in dağılım fonksiyonu ise aşağıdaki şartlar sağlanır:*

i. *F azalmayan bir fonksiyon yani, $x \leq y$ ise $F(x) \leq F(y)$ dir.*

ii. *$\lim_{x\to-\infty} F(x) = 0$ ve $\lim_{x\to\infty} F(x) = 1$ dir (Bu genellikle $F(-\infty) = 0, F(\infty) = 1$ olarak yazılır.*

İspat 3.1 *i.* *A ve B olaylarını $A = \{X \leq x\}$, $B = \{X \leq y\}$ olarak tanımlayalım. $x \leq y$ olduğundan $A \subset B$ yazılır. Teorem 2.2 'un sonucundan dolayı $P(A) \leq P(B)$ dir. Buradan*

$$\begin{aligned} P(A) &= P(X \leq x) = F(x) \\ P(B) &= P(X \leq y) = F(y) \end{aligned}$$

olduğundan $F(x) \leq F(y)$ yazılır.

ii. *$F(x) = \int_{-\infty}^{x} f(s)ds$ olduğunu biliyoruz. Böylece*

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \int_{-\infty}^{x} f(s)ds = \lim_{x\to-\infty} \int_{x}^{x} f(s)ds = 0$$

olur. Diğer yandan

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{\infty} f(s)ds = 1$$

olduğu $f(s)$ 'nin olasılık yoğunluk fonksiyonu olmasından yazılır.

Örnek 3.9 *X rastlantı değişkeninin olasılık yoğunluk fonksiyonu*

$$f(x) = \begin{cases} 0.75(1-x^2) & ,-1 \leq x \leq 1 \\ 0 & ,x \text{ 'in diğer değerleri} \end{cases}$$

olsun.

i. Dağılım fonksiyonunu,

ii. $P(-\frac{1}{2} \leq X \leq \frac{1}{2})$ olasılığını,

iii. $P(\frac{1}{4} \leq X \leq 2)$ olasılığını,

iv. $P(X \leq x)$ olacak şekilde x'i

bulunuz.

Çözüm 3.9 *i. $F(x) = P(X \leq x)$ olduğunu biliyoruz. Buna göre,*

a. $x < -1$ için

$$F(x) = \int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{x} 0ds = 0$$

olur.

b. $-1 \leq x \leq 1$ için

$$F(x) = \int_{-\infty}^{x} f(s)ds = \int_{-1}^{x} 0.75(1-s^2)ds = 0.5 + 0.75x - 0.25x^3$$

olur.

c. $x > 1$ için durumu inceleyelim. $P(X \geq x)$ için $F(x)$ tanımı uygulayarak doğrudan bulamayız. Onun için, Teorem 3.1 e göre $x > 1$ olduğundan

$$F(x) = P(X \leq x) \geq F(1) = 1$$

ve diğer yandan $P(X \leq x)$ olasılık olduğundan

$$F(x) = P(X \leq x) \leq 1$$

yazılır. Bu iki eşitsizlikten $F(x) = 1$ bulunur. O halde,

$$F(x) = \begin{cases} 0 & , \quad x < -1 \\ 0.75 \int_0^x (1-s^2)ds = 0.5 + 0.75x - 0.25x^3 & , \quad -1 \leq x \leq 1 \\ 1 & , \quad x > 1 \end{cases}$$

istenen dağılım fonksiyonudur.

ii. Teorem 3.1 e göre

$$P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = F\left(\frac{1}{2}\right) - F\left(-\frac{1}{2}\right) = 0.68$$

bulunur.

iii. Teorem 3.1 e göre,

$$P(\frac{1}{4} \leq X \leq 2) = F(2) - F\left(\frac{1}{4}\right) = 0.31$$

olur.

iv.

$$P(X \leq x) = F(x) = 0.5 + 0.75x - 0.25x^3 = 0.95$$

ve buradan,

$$3x - x^3 = 1.8$$

olur. Bu denklemin bir çözümü $x = 0.73$ dür.

3.4 Bir Rastlantı Değişkeninin Beklenen Değeri

Tanım 3.8 *X aşağıdaki olasılık fonksiyonuna sahip kesikli bir rastlantı değişkeni olsun.*

$X = x$	x_1	x_2	$\cdots$	x_N
$f(x) = P(X = x)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_N)$

X değişkeninin $E(X)$ ile gösterilen beklenen değeri aşağıdaki gibi tanımlanır:

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_N f(x_N) = \sum_{i=1}^{N} x_i f(x_i)$$

X rastlantı değişkeni sayılabilir sonsuzluktaki $x_1, x_2, \ldots, x_N, \ldots$ değelerini alıyorsa

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_N f(x_N) + \cdots = \sum_{i=1}^{\infty} x_i f(x_i)$$

olur.

Bu $E(X)$, X 'in mümkün değerlerinin ortalamasıdır. Beklenen değeri veya ortalamayı μ ille göstereceğiz. $X, Y, \ldots$ gibi birden fazla rastlantı değişkenleri ile çalışırken karışıklık olmaması için $\mu_X = E(X), \mu_Y = E(Y), \ldots$ yazacağız. Eğer rastlantı değişkeni bir tane ise karışıklık sözkonusu olmayacağından $E(X) = \mu_X = \mu$ yazacağız.

Örnek 3.10 *Düzgün bir zar atıldığında üste gelen noktaların beklenen değeri nedir?*

Çözüm 3.10 *Zarın üste gelen noktaların sayısını X ile gösterelim. X değişkeninin mümkün değerleri 1, 2, 3, 4, 5, 6 dır. Bunların her birinin elde edilme olasılığı $\frac{1}{6}$ dır. O halde beklenen değer*

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2} = 3.5$$

olur.

Örnek 3.11 *İçinde 3 beyaz ve 2 siyah top bulunan bir kavanozdan çekileni yerine koyma şartı ile iki top çekilmiştir. Çekilen her beyaz top için 100 lira kazanılacak ve çekilen her siyah top için 50 lira kaybedilecektir. Bu oyunda beklenen kâr nedir?*

Çözüm 3.11 *X rastlantı değişkeni kazanılan liranın sayısı olsun. X 'in beklenen değerini bulursak problemi çözmüş olacağız. Aşağıdaki tablo ile mümkün sonuçların örneklem uzayı, karşılık gelen olasılıklar ve kâr gösterilmiştir.*

Şekil 3.3:

X rastlantı değişkeninin beklenen değeri

$$E(X) = 200\left(\frac{9}{25}\right) + 50\left(\frac{6}{25}\right) + 50\left(\frac{6}{25}\right) - 100\left(\frac{4}{25}\right) = 80 \quad lira$$

olarak bulunur.

Tanım 3.9 *$f(x)$, X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu olsun. X 'in beklenen değeri $E(x)$ ile gösterilir ve*

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

ile tanımlanır.

Örnek 3.12 *X sürekli rastlantı değişkeni ve*

$$f(x) = \begin{cases} e^{-x} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

X 'in olasılık yoğunluk fonksiyonu olsun. $\mu = E(X)$ 'i bulunuz.

Çözüm 3.12 *Tanım gereğince ve $x < 0$ için $f(x) = 0$ olduğundan*

$$\mu = E(X) = \int_{-\infty}^{\infty} xe^{-x}dx = \int_{0}^{\infty} xe^{-x}dx$$

dir. Bu integrali kısmi integrasyon metodu ile hesaplayacağız. Bunun için $u = x$, ve $dv = e^{-x}dx$ alırsak $du = dx$, ve $v = -e^{-x}$ olur. Buradan

$$E(X) = -xe^{-x}\Big|_0^{\infty} + \int_0^{\infty} e^{-x}dx$$

bulunur. Hospital kuralının uygulanması ile

$$\lim_{x\to\infty} xe^{-x} = \lim_{x\to\infty} \frac{1}{e^x} = 0$$

ve böylece

$$E(X) = \int_0^{\infty} xe^{-x}dx = \int_0^{\infty} e^{-x}dx = -e^{-x}\Big|_0^{\infty} = 1$$

bulunur.

3.5 Beklenen Değerin Özellikleri

Tanım 3.10 *i. X kesikli rastlantı değişkeni için olasılık fonksiyonu*

$X = x$	x_1	x_2	$\cdots$	x_N
$f(x) = P(X = x)$	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_N)$

olarak verilsin. $g(X)$ fonksiyonunun beklenen değeri,

$$E[g(X)] = g(x_1)f(x_1) + g(x_2)f(x_2) + \cdots + g(x_N)f(x_N)$$

veya

$$E[g(X)] = \sum_{i=1}^{N} g(x_i)f(x_i)$$

olarak tanımlanır.

ii. X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu $f(x)$ olsun. $g(X)$ fonksiyonunun beklenen değeri,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

olarak tanımlanır.

Örnek 3.13 *Bir zarın atılması ile ilgili bir deneyde X rastlantı değişkeninin olasılık fonksiyonu,*

$X = x$	1	2	3	4	5	6
$f(x) = P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

olarak veriliyor. Aşağıda verilen her fonksiyon için beklenen değerleri bulunuz.

i. $g(X) = 3$ *ii.*$g(X) = X$ *iii.* $g(X) = X^2$ *iv.* $g(X) = 2X$

Çözüm 3.13 *Önce rastlantı değişkeninin durumuna göre tablo yapalım:*

x	*1*	*2*	*3*	*4*	*5*	*6*
x^2	*1*	*4*	*9*	*16*	*25*	*36*
$2x$	*2*	*4*	*6*	*8*	*10*	*12*
$f(x) = P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

i.

$$\begin{aligned} E(3) &= \sum_{i=1}^{6} 3f(x_i) \\ &= 3\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) \\ &= 3 \end{aligned}$$

olur.

ii.

$$\begin{aligned} E(X) &= \sum_{i=1}^{6} x_i f(x_i) \\ &= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) \\ &= \frac{7}{2} \end{aligned}$$

olarak bulunur.

iii.

$$\begin{aligned} E(X^2) &= \sum_{i=1}^{6} {x_i}^2 f(x_i) \\ &= 1\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) + 16\left(\frac{1}{6}\right) + 25\left(\frac{1}{6}\right) + 36\left(\frac{1}{6}\right) \\ &= \frac{91}{6} \end{aligned}$$

dır.

iv.

$$\begin{aligned} E(2X) &= \sum_{i=1}^{6} 2x_i f(x_i) \\ &= 2\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) + 8\left(\frac{1}{6}\right) + 10\left(\frac{1}{6}\right) + 12\left(\frac{1}{6}\right) \\ &= \frac{42}{6} = 7 \end{aligned}$$

olarak bulunur.

Teorem 3.2 *a, b sabit ve X (kesikli veya sürekli) rastlantı değişkeni olsun. Bu taktirde*

$$E(aX + b) = aE(X) + b$$

dir.

İspat 3.2 *İspatı kesikli rastlantı değişkeni için yapacağız. Sürekli rastlantı değişkeni için de benzer muhakeme uygulanır. $g(X) = aX + b$ diyelim. Tanıma 4.10 'a göre,*

$$\begin{aligned} E[g(X)] = E(aX + b) &= \sum_{i=1}^{N} (ax_i + b) f(x_i) \\ &= a \sum_{i=1}^{N} x_i f(x_i) + \sum_{i=1}^{N} b f(x_i) \end{aligned}$$

yazılır. Hatırlanacağı gibi,

$$\sum_{i=1}^{N} x_i f(x_i) = E(X)$$

ve $f(x)$ olasılık fonksiyonu olduğundan,

$$\sum_{i=1}^{N} f(x_i) = 1$$

dır. Bu değerler yerine yazılırsa,

$$E(aX + b) = aE(X) + b$$

elde edilir.

Bu teoremin sonuçu olarak şunlar yazılabilir:

i. b sabit olmak üzere $E(b) = b$ dir.

ii. a sabit olmak üzere $E(aX) = aE(X)$ dir.

iii. $E[X - E(X)] = 0$ dır.

3.6 Bir Rastlantı Değişkeninin Varyansı

Varyans, bir rastlantı değişkeninin, kendi ortalaması (beklenen değeri) civarındaki dağılımının (saçılmasının) derecesini gösterir. Varyans ne kadar küçük olursa değişkenin değeri kendi ortalamasına o kadar yakın olur.

Tanım 3.11 *X bir rastlantı değişkeni ve bu değişkenin beklenen değeri $E(X) = \mu$ olsun. Bu takdirde X' in varyansı $Var(X)$ (veya ${\sigma_X}^2$) ile gösterilir ve*

$$Var(X) = {\sigma_X}^2 = E[(X - \mu)^2]$$

olarak tanımlanır.

Bu tanıma göre X kesikli rastlantı değişkeni ise,

$$Var(X) = {\sigma_X}^2 = \sum_{i=1}^{N} (x_i - \mu)^2 f(x_i)$$

ve X sürekli rastlantı değişkeni ise,

$$Var(X) = {\sigma_X}^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

olur.

$Var(X) = {\sigma_X}^2 = E[(X - \mu)^2]$ formülünden istifade ederek,

$$\begin{aligned} Var(X) &= {\sigma_X}^2 = E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + E(\mu^2)] \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

elde edilir. Bu da varyansı bulmak için bir başka formüldür.

Tanım 3.12 *${\sigma_X}^2$ varyansının karekökü olan ${\sigma_X}'$ e X rastlantı değişkeninin standart sapması denir.*

Örnek 3.14 *X rastlantı değişkeninin olasılık fonksiyonu aşağıdaki gibi verilmektedir:*

$X = x$	*1*	*2*	*3*	*4*	*5*
$f(x) = P(X = x)$	*0.1*	*0.3*	*0.2*	*0.3*	*0.1*

Buna göre,

i. $E(X)$ *'i,*

ii. $Var(X) = E[(X - E(X))^2]$ *formülünü kullanarak* $Var(X)$ *'i,*

iii. $E(X^2)$ *'yi ,*

iv. $Var(X) = E(X^2) - [E(X)]^2$ *formülünü kullanarak* $Var(X)$ *'i*

hesaplayınız.

Çözüm 3.14 *i. Kesikli rastlantı değişkenler için beklenen değeri bulma formülünü kullanarak,*

$$E(X) = 1(0.1) + 2(0.3) + 3(0.2) + 4(0.3) + 5(0.1) = 3$$

elde edilir.

ii. Bilindiği üzere,

$$E[(X - E(X))^2] = E[(X - 3)^2] = \sum_{i=1}^{5}(x_i - 3)^2 f(x_i)$$

dir.

x	*1*	*2*	*3*	*4*	*5*
$(x - 3)$	*-2*	*-1*	*0*	*1*	*2*
$(x - 3)^2$	*4*	*1*	*0*	*1*	*4*
$f(x) = P(X = x)$	*0.1*	*0.3*	*0.2*	*0.3*	*0.1*

Bu tablodan yararlanarak,

$$E[(X-3)^2] = 4(0.1) + 1(0.3) + 0(0.2) + 1(0.3) + 4(0.1) = 1.4$$

olur. Yani,

$$Var(X) = 1.4$$

dür.

iii. Beklenen değer formülünü kullanarak,

$$E(X^2) = 1(0.1) + 4(0.3) + 9(0.2) + 16(0.3) + 25(0.1) = 10.4$$

bulunur.

iv.

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= 10.4 - 9 \\ &= 1.4 \end{aligned}$$

elde edilir.

Örnek 3.15 *3 kırmızı ve 4 beyaz top bulunan bir torbadan geri koymaksızın 2 top çekiliyor. Eğer X, kırmızı top çekmenin sayısı ise X 'in beklenen değerini ve standart sapmasını bulunuz.*

Çözüm 3.15 *X, kırmızı top çekmenin sayısı olduğundan alabileceği değerler 0, 1, 2 'dir. Önce olasılık fonksiyonunu bulalım. Bunun için $P(X=0)$, $P(X=1)$ ve $P(X=2)$ değerlerini bulacağız. B ile beyaz, K ile kırmızı topları gösterirsek*

$$P(X=0) = P(B_1B_2) = \left(\frac{4}{7}\right)\left(\frac{3}{6}\right) = \frac{2}{7}$$

$$P(X=1) = P(B_1K_2) + P(K_1B_2) = \left(\frac{4}{7}\right)\left(\frac{3}{6}\right) + \left(\frac{3}{7}\right)\left(\frac{4}{6}\right) = \frac{4}{7}$$

ve

$$P(X=2) = P(K_1K_2) = \left(\frac{3}{7}\right)\left(\frac{2}{6}\right) = \frac{1}{7}$$

olur. O halde X 'in olasılık fonksiyonu

$X = x$	*0*	*1*	*2*
$f(x) = P(X = x)$	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

olur.

$$E(X) = 0\left(\frac{2}{7}\right) + 1\left(\frac{4}{7}\right) + 2\left(\frac{1}{7}\right) = \frac{6}{7}$$

dir. Yani, kırmızı topun beklenen değeri $\frac{6}{7}$ *'dir.*

Şimdi de X *'in standart sapmasını bulalım: Bilindiği gibi* $\sigma_X = \sqrt{Var(X)}$ *'dir. Onun için önce* $Var(X)$ *'i bulmalıyız.* $Var(X) = E(X^2) - [E(X)]^2$ *eşitliğinde* $[E(X)]^2 = \frac{36}{49}$ *ve*

$$E(X^2) = 0\left(\frac{2}{7}\right) + 1\left(\frac{4}{7}\right) + 4\left(\frac{1}{7}\right) = \frac{8}{7}$$

olduğundan

$$Var(X) = \frac{8}{7} - \frac{36}{49} = \frac{20}{49} = 0.408$$

dır. Buradan standart sapma

$$\sigma_X = \sqrt{0.408} = 0.639$$

olarak bulunur.

Varyans ile ilgili şu özellikler vardır:

i. a bir sabit olmak üzere $Var(a) = 0$ 'dır.

ii. a bir sabit olmak üzere $Var(aX) = a^2 Var(X)$ 'dir.

iii. a, b birer sabit olmak üzere $Var(aX + b) = a^2 Var(X)$ 'dir.

3.7 Rastlantı Değişkenlerin Fonksiyonları

X, olasılık yoğunluk fonksiyonu $f(x)$ olan sürekli rastlantı değişken ve H sürekli bir fonksiyon olduğunda çok önemli bir durumla karşılaşılır. $Y = H(X)$ sürekli rastlantı değişkeninin fonksiyonu $g(y)$'yi bulmaya çalışacağız. Genel metod aşağıdaki gibidir:

i. $\{Y \leq y\}$ olayına eşdeğer olan bir A olayını belirterek Y değişkeninin dağılım fonksiyonu $G(y) = P(Y \leq y)$ yi buluruz.

ii. $G(y)$ nin y ye göre türevini alarak $g(y)$ yi buluruz.

iii. $g(y) > 0$ olacak şekilde Y nin tanım bölgesinde y nin değerlerini belirtiriz.

Örnek 3.16 *X rastlantı değişkeninin olasılık yoğunluk fonksiyonu*

$$f(x) = \begin{cases} 2x & , \quad 0 < x < 1 \\ 0 & , \quad \textit{başka yerde} \end{cases}$$

olsun. $H(x) = 3x + 1$ alalım. Böylece $Y = H(X) = 3x + 1$ sürekli rasrgele değişkeninin fonksiyonunu bulunuz.

Çözüm 3.16 *$H(x) = 3x + 1$ fonksiyonunun grafiği aşağıdaki gibidir.*

Şekil 3.4:

Dağılım fonksiyonu

$$\begin{aligned} G(y) = P(Y \leq y) &= P(3x + 1 \leq y) = P(X \leq \frac{y-1}{3}) \\ &= \int_0^{\frac{y-1}{3}} 2x dx = [\frac{y-1}{3}]^2 \end{aligned}$$

olarak bulunur. Böylece $g(y) = G'(y) = \frac{2}{9}(y-1)$ bulunur. $0 < x < 1$ için $f(x) > 0$ olduğundan ($x = 0$ için $y = 3x+1 = 1$ ve $x = 1$ için $y = 3x+1 = 4$) $1 < y < 4$ için $g(y) > 0$ bulunur. O halde

$$g(y) = \begin{cases} \frac{2}{9}(y-1) & , \quad 1 < y < 4 \\ 0 & , \quad y \geq 4, \ y \leq 1 \end{cases}$$

yazılır. Bunun grafiği şöyledir.

Şekil 3.5:

NOT: A olayı $Y \leq y$ yada basit olarak $X \leq \frac{y-1}{3}$'e eşdeğerdir.

Örnek 3.17 *X rastlantı değişkeninin fonksiyonu*

$$f(x) = \begin{cases} 2x & , \quad 0 < x < 1 \\ 0 & , \quad x \geq 1, \ x \leq 0 \end{cases}$$

olsun. $H(x) = e^{-x}$ alalım. $Y = H(X)$ fonksiyonunun olasılık fonksiyonunu bulunuz.

Çözüm 3.17

$$\begin{aligned} G(y) &= P(Y \leq y) = P(E^{-x} \leq y) = P(X \geq -ln(y) \\ &= \int_{-ln(y)}^{1} 2x dx = 1 - (-ln(y))^2 \end{aligned}$$

bulunur. O halde

$$g(y) = G'(y) = -2\frac{ln(y)}{y}$$

elde edilir. $0 < x < 1$ için $f(x) > 0$ olduğundan $1/e < y < 1$ için $g(y) > 0$ bulunur. O halde

$$g(y) = \begin{cases} -\frac{2ln(y)}{y} & , \quad 1/e < y < 1 \\ 0 & , \quad y \leq 1/e, \ y > 1 \end{cases}$$

yazılır.

3.8 Çözülmüş Problemler

1. 4 parayı bir kez atalım. X rastlantı değişkeni turaların sayısını göstersin. X değişkeninin olasılık fonksiyonunu ve dağılım fonksiyonunu bulunuz?

Çözüm: Örneklem uzay S={TYYY, TYYT, TYTY, TYTT, TTYY, TTYT, TTTT, TTTY, YYYY, YYYT, YYTT, YYTY, YTYY, YTYT, YTTY, YTTT} şeklindedir. Olasılık fonksiyonu,

$X = x$	0	1	2	3	4
$f(x) = P(X = x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

olarak yazılır. Daha kısa olarak,

$$f(x) = \frac{\binom{4}{x}}{2^4}, \quad x = 0, 1, 2, 3, 4$$

olarak da yazılabilir. Dağılım fonksiyonu da

$X = x$	0	1	2	3	4
$F(x) = P(X \leq x)$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	$\frac{16}{16}$

veya

$$F(x) = P(X \leq x) = \sum_{s=0}^{x} \frac{\binom{4}{s}}{2^4}$$

olur.

2. X rastlantı değişkeninin olasılık fonksiyonu aşağıdaki gibi verilmektedir:

$X = x$	0	1	2	3	4	5	6	7
$f(x) = P(X = x)$	0	c	2c	2c	3c	c^2	$2c^2$	$7c^2 + c$

i. c değerini hesaplayınız?

ii. $P(X \leq k) > \frac{1}{2}$ ise k'nın minimum değerini bulunuz.

Çözüm:

i. $0 + c + 2c + 2c + 3c + c^2 + 2c^2 + 7c^2 + c = 1 \Longrightarrow 10c^2 + 9c = 1$ $c_1 = -1$,$c_2 = 1/10$ olur. Olasılık negatif olamayacagından $c = 1/10$ olur.

ii. (a) şıkkında bulduğumuz $c = 1/10$ değerini yerine yazdığımızda

$X = x$	0	1	2	3	4	5	6	7
$f(x) = P(X = x)$	0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{1}{100}$	$\frac{2}{100}$	$\frac{7}{100} + \frac{1}{10}$

bulunur. Burada

$$P(X=0) + P(X=1) = \frac{1}{10} < \frac{1}{2}$$
$$P(X=0) + P(X=1) + P(X=4) = 0 + \frac{1}{10} + \frac{2}{10}$$
$$= \frac{3}{10} < \frac{1}{2}$$
$$P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$= 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} = \frac{5}{10} = \frac{1}{2}$$
$$P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$$
$$= 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} = \frac{8}{10} > \frac{1}{2}$$

olduğundan,

$$F(4) = P(X \leq 4) > \frac{1}{2} \Longrightarrow k = 4$$

yazılır.

3. Bir X rastlantı değişkeninin mümkün değerleri $1, 2, 3, \cdots, n$ tamsayılarıdır ve c belli bir sabit olduğuna göre

$$f(x) = P(X = x) = cx$$

olarak veriliyor. $c = \frac{2}{n(n+1)}$ olduğunu gösteriniz.

Çözüm: X değişkeninin olasılık fonksiyonu tablo olarak

$X = x$	1	2	3	$\cdots$	n
$f(x) = P(X = x)$	c	2c	3c	$\cdots$	nc

şeklinde yazılabilir. Dolayısıyla

$$\begin{aligned} 1 = \sum_{i=1}^{n} f(x_i) &= c + 2c + 3c + \cdots + nc = c(1 + 2 + \cdots + n) \\ &= \frac{n(n+1)}{2} c \Longrightarrow c = \frac{2}{n(n+1)} \end{aligned}$$

olduğu görülür.

4. X rastlantı değişkeninin olasılık fonksiyonu

$$p(x) \quad = \quad \frac{a}{e^x + e^{-x}}$$

olduğuna göre a sabitini bulunuz.

Çözüm:

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

olduğuna göre,

$$\int_{-\infty}^{\infty} \frac{a}{e^x + e^{-x}} dx = \int_{-\infty}^{\infty} \frac{ae^x}{e^{2x} + 1} dx$$

$u = e^x$ değişken değiştirmesi yapılırsa, $x \longrightarrow -\infty$ için $u \longrightarrow 0$, $x \longrightarrow +\infty$ için $u \longrightarrow +\infty$ ve böylece

$$1 = \int_{-\infty}^{\infty} \frac{ae^x}{e^{2x} + 1} dx = a \int_0^{\infty} \frac{du}{1 + u^2} = (a) arctg(u) \Big|_0^{\infty} = a\frac{\pi}{2}$$

elde edilir. Buradan

$$a = \frac{2}{\pi}$$

bulunur.

5. X bir tüpün elektronlarının hayat uzunluğunu göstersin. Farzedelimki X, $f(x) = be^{-bx}$ $(x \geq 0)$, olasılık fonksiyonuna sahip sürekli bir rastlantı değişkenidir.

$$p_j = p(j \leq x \leq j + 1)$$

veriliyor. p_j nin $(1 - a)a^j$ şeklinde olduğunu gösteriniz.

Çözüm:

$$\begin{aligned} p_j = p(j \leq x \leq j+1) &= \int_j^{j+1} be^{-bx} dx \\ &= -e^{-bx}\Big|_j^{j+1} \\ &= -e^{-b(j+1)} + e^{-bj} \\ &= e^{-bj}(1 - e^{-b}) \end{aligned}$$

olur. $a = e^{-b}$ alınırsa,

$$p_j = a^j(1 - a)$$

elde edilir.

6. X rastlantı değişkeninin olasılık yoğunluk fonksiyonu,

$$f(x) = ax^2e^{-kx}, \quad (k > 0, 0 \leq k < \infty)$$

olarak verilsin.

i. a katsayısını,

ii. X değişkeninin dağılım fonksiyonunu,

iii. X değişkeninin $(0, 1/2)$ aralığında bulunması olasılığını bulunuz.

Çözüm:

i.

$$\int_0^\infty ax^2e^{-kx}dx = 1$$

eşitliğinden,

$$a = \frac{1}{\int_0^\infty x^2e^{-kx}dx} = \frac{k^3}{2}$$

bulunur.

ii.

$$F(x) = \int_0^x \frac{k^3}{2}u^2e^{-ku}du = 1 - \frac{k^2x^2 + 2kx + 2}{2}e^{-kx}$$

olur.

iii.

$$P(0 < x < \frac{1}{k}) = F\left(\frac{1}{k}\right) = 1 - \frac{5}{2e} \approx 0.086$$

bulunur.

7. $1, 2, \cdots, 10$ tamsayıları arasından rastgele seçilen bir sayının bölen sayısı X olsun. X değişkeninin ortalamasını(beklenen değerini) ve varyansını bulunuz.

Çözüm:

Tamsayılar	1	2	3	4	5	6	7	8	9	10
Bölen sayısı	1	2	2	3	2	4	2	4	3	4

X değişkeninin olasılık fonksiyonu:

X=x	1	2	3	4
P(X=x)	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

şeklindedir. Ohalde

$$E(X) = 1\left(\frac{1}{10}\right) + 2\left(\frac{4}{10}\right) + 3\left(\frac{2}{10}\right) + 4\left(\frac{3}{10}\right) = \frac{27}{10}$$

ve $Var(X) = E(X^2) - [E(X)]^2$ olduğundan

$$E(X^2) = 1\left(\frac{1}{10}\right) + 4\left(\frac{4}{10}\right) + 9\left(\frac{2}{10}\right) + 16\left(\frac{3}{10}\right) = \frac{83}{10}$$

yazılır. Dolayısıyla

$$Var(X) = \frac{83}{10} - \left(\frac{27}{10}\right)^2 = 0.91$$

bulunur.

8. X rastlantı değişkeninin olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} \frac{1}{b-a} & , \ a < x < b \\ 0 & , \ \text{aralığın dışında} \end{cases}$$

olduğuna göre dağılım fonksiyonunu, ortalama değerini ve varyansını bulunuz.

Çözüm: Dağılım fonksiyonu,

$$P(a \leq X \leq x) = F(x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt = \frac{1}{b-a}t\Big|_a^x = \frac{x-a}{b-a}$$

bulunur. Sürekli rastlantı değişkenler için ortalama tanımından

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} f(x)dx = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx \\ &= \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} \end{aligned}$$

elde edilir. X değişkeninin varyansı ise

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}dx = \frac{(b-a)^2}{12}$$

olarak bulunur.

9. $\mu = E(X)$ sonlu bir değer, c bir sabit olsun. Bu takdirde,

   i. $E(X-\mu) = 0$,

   ii. $E(X-c)^2 = E(X-\mu)^2 + (\mu - c)^2$ olduğunu gösteriniz.

Çözüm:

i.

$$\begin{aligned} E(X-\mu) &= \sum_{x=1}^{n}[x - E(X)]f(x) = \sum_{x=1}^{n}\{x - \sum_{u=1}^{n} uf(u)\}f(x) \\ &= \sum_{x=1}^{n} xf(x) - \sum_{x=1}^{n}\sum_{u=1}^{n} uf(u)f(x) \\ &= E(X) - \sum_{u=1}^{n} uf(u)\sum_{x=1}^{n} f(x) \\ &= E(X) - E(X) = 0 \end{aligned}$$

Burada $f(x)$ olasılık fonksiyonu olduğundan $\sum f = 1$ dir.

ii.

$$\begin{aligned} E(X-c)^2 &= E(X^2 - 2cX + c^2) \\ &= E(X^2) - 2cE(X) + c^2 - [E(X)]^2 + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 + c^2 - 2cE(X) + [E(X)]^2 \\ &= E(X-\mu)^2 + (c-\mu)^2 \end{aligned}$$

bulunur.

3.9 Problemler

1. X rastlantı değişkeninin olasılık fonksiyonu

$$f(x) = c\binom{5}{x}, \quad x = 0, 1, 2, 3, 4, 5$$

verilsin. c bir sabit olmak üzere c değerini bulunuz.

2. X rastlantı değişkeninin olasılık fonksiyonu

$$f(x) = cx, \quad x = 3, 4, 5, 6$$

olsun. Burada c bir sabittir.

$i.$ c değerini hesaplayınız,

$ii.$ $E(X)$ hesaplayınız,

$iii.$ $Var(X)$ hesaplayınız.

3. X rastlantı değişkeni İstanbul, Ankara yolunda verilmiş bir günde meydana gelen trafik kazalarının sayısının göstersin. c bir sabit olmak üzere X değişkeninin olasılık fonksiyonu aşağıdadır:

$$f(x) = \begin{cases} c, & x = 0 \ \ ise \\ 2c, & x = 1 \ \ ise \\ 3c, & x = 2 \ \ ise \\ 4c, & x = 3 \ \ ise \\ 1.5c, & x = 4 \ \ ise \\ 0.5c, & x = 5 \ \ ise \end{cases}$$

   $i.$ c değerini bulunuz.

   $ii.$ $P(X > 3)$,

   $iii.$ $P(0 < x < 4)$,

   $iv.$ $P(0 < x < 2)$

   olasılıklarını bulunuz.

4. 52'lik bir desteden 4 kart çekilmiştir.

   $i.$ Kupaların sayısının dağılımı,

   $ii.$ İkililerin sayısının dağılımı

   nedir?

5. 4 tanesi bozuk olan bir düzine yumurtadan yerine koymaksızın 3 tane çekiliyor. X rastlantı değişkeni bu örnekteki bozuk yumurtaların sayısı olsun. X'in olasılık fonksiyonunu bulunuz.

6. İçinde 5 beyaz 3 yeşil top bulunan bir kavanozdan rastgele iki top çekiliyor. Çekilen yeşil topların sayısı X olsun.

   $i.$ X değişkeninin olasılık fonksiyonunu, *ii.* X değişkeninin dağılım fonksiyonunu bulunuz.

7. $Y = 3x - 5$ ve $E(X) = 4$, $Var(X) = 2$ olsun. $E(Y)$ ve $Var(Y)$ değerlerini bulunuz.

8. X rastlantı değişkeninin olasılık yoğunluk fonksiyonu

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

olsun.

i. $Y = X^{1/3}$ değişkeninin olasılık yoğunluk fonksiyonunu bulunuz.

ii. $Y = X^2$ değişkeninin olasılık yoğunluk fonksiyonunu bulunuz.

9.

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

fonksiyonunun bir olasılık yoğunluk fonksiyonu olduğunu gösteriniz.

Chapter 4

BAZI KESİKLİ RASTLANTI DEĞİŞKENLERİNİN DAĞILIMLARI

Önceki bölümde genel olarak rastlantı değişkenlerinin dağılımları üzerinde durduk. Bu bölümde önemlerinden dolayı bazı kesikli rastlantı değişkenleri üzerinde duracağız.

4.1 Binom Dağılımı

Bir deney n defa bağımsız olarak yapıldığında, bir A olayının meydana gelme sayısı ile ilgileneceğiz. Tek bir deneme için $P(A) = p$ olduğunu kabul edelim. Bu takdirde, aynı deneme için A olayının olmama olasılığı $q = 1 - p$ dir. n defa yapılan bir deneyde X rastlantı değişkeni "A olayının meydana gelme sayısı" olarak alınsın. Buna göre, X in mümkün değerleri $0, 1, 2, \ldots, n$ olur. Bu değerlere karşılık gelen olasılıkları bulmak istiyoruz. Bunun için, bu değerlerin herhangi birini, örneğin, $X = x$'i gözönüne alalım. Bu şu demektir: n denemenin x inde A olayı meydana gelmekte, $n - x$ denemede ise A olayı gözükmemektedir. Yani,

$$\underbrace{AA\ldots A}_{x}\underbrace{BB\ldots B}_{n-x}$$

dır. Burada $\overline{A} = B$ dir. Şu hususu tekrar hatırlatalım ki denemeler bağımsız olarak (biri diğerini etkilemeden) yapılmaktadır. Buna göre $P(A) = p$ ve $P(B) = q$ için yukarıdaki olayın olasılığı

$$\underbrace{pp\ldots p}_{x}\underbrace{qq\ldots q}_{n-x} = p^x q^{n-x}$$

dır. x i seçmenin $\binom{n}{x}$ farklı yolu olduğundan $X = x$ in olasılığı

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

olur. Bu olasılık fonksiyonlu dağılıma **Binom Dağılımı** denir. A olayının meydana gelmesine başarı, meydana gelmemesine de başarısızlık adı verilir.

Eğer özel olarak $p = q = \frac{1}{2}$ alınırsa

$$f(x) = P(X = x) = \binom{n}{x} \left(\frac{1}{2}\right)^n, \quad x = 0, 1, \ldots, n$$

yazılır.

Örnek 4.1 *Bir deneydeki bir olayın başarı olasılığı p ve başarısızlık olasılığı da $q = 1 - p$ olsun. X rastlantı değişkeni "başarılı olma sayısı" olsun. Bu deney 5 kez bağımsız olarak yapıldığında X in $0, 1, 2, 3, 4, 5$ değerleri için olasılığı bulunuz ve sonucu binom açılımı ile mukayese ediniz.*

Çözüm 4.1 *İstenilen olasılık*

$$f(x) = P(X = x) = \binom{n}{x} \left(\frac{1}{2}\right)^n, \quad x = 0, 1, 2, 3, 4, 5$$

formülü ile bulunur. Buradan

$$\begin{aligned}
P(X = 0) &= \binom{5}{0} q^5 p^0 = q^5 \\
P(X = 1) &= \binom{5}{1} q^4 p^1 = 5q^4 p \\
P(X = 2) &= \binom{5}{2} q^3 p^2 = 10q^3 p^2
\end{aligned}$$

$$\begin{aligned} P(X=3) &= \binom{5}{3} q^2p^3 = 10q^2p^3 \\ P(X=4) &= \binom{5}{4} qp^4 = 5qp^4 \\ P(X=5) &= \binom{5}{5} q^0p^5 = p^5 \end{aligned}$$

$q^5, 5q^4p, 10q^3p^2, 10q^2p^3, 5qp^4, p^5$; $(p+q)^5$ *in binom açılımındaki terimleridir.*

Örnek 4.2 *Düzgün bir zar 4 kez atılıyor.*

i. İki defa 6,

ii. En az iki defa 6,

gelmesi olasılığı nedir?

Çözüm 4.2 *"6 gelme sayısını" X ile gösterelim. $p=\frac{1}{6}, q=\frac{5}{6}$ ve $n=4$ olur.*

i. Burada $x=2$ olup

$$\binom{4}{2} p^2q^2 = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$$

bulunur.

ii. En az iki defa 6 gelmesi düşünüldüğünden istenilen olasılık

$$\begin{aligned} P(X\geq 2) &= P(X=2)+P(X=3)+P(X=4) \\ &= \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2+\binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)+\binom{4}{4}\left(\frac{1}{6}\right)^4 \\ &= \frac{171}{1296} = 0.1319444 \end{aligned}$$

elde edilir.

Örnek 4.3 *Kız veya erkek çocuk doğması olasılığı eşit olduğu düşünülüyor. Buna göre 5 çocuklu bir ailede iki erkek çocuk olması olasılığı nedir?*

Çözüm 4.3 *Erkeklerin sayısı X olsun. Verilenlere göre*

$$n = 5,\ p = \tfrac{1}{2} \text{ ve } x = 2$$

dır. Böylece binom dağılımındaki olasılık fonksiyonundan

$$P(X = 2) \quad = \quad \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{5-2} = \frac{5!}{3!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{10}{12}$$

bulunur.

Teorem 4.1 *X rastlantı değişkeni binom dağılımına sahip olsun. Bu takdirde binom dağılımının ortalaması ve varyansı sırayla*

$$\mu = E(X) = np,\ \sigma^2 = E(X^2) - [E(x)]^2 = npq$$

olur.

İspat 4.1 *X in olasılık fonksiyonu*

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

dır. Ortalama (beklenen) değerin tanımından

$$\mu = E(x) = \sum_{x=0}^{n} x f(x)$$

yazılır. O halde

$$\begin{aligned}
\mu &= \sum_{x=0}^{n} \frac{xn!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)![n-1-(x-1)]!} p^x q^{n-1-(x-1)} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)![n-1-(x-1)]!} p^{x-1} q^{n-1-(x-1)} \\
&= np(p+q)^{n-1} \\
&= np
\end{aligned}$$

bulunur. Varyansı hesaplamak için önce $E(X^2)$ değerini bulmalıyız.

$$E(X^2)=\sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \frac{x^2 n! p^x q^{n-x}}{x!(n-x)!}$$

yazılır. $x^2 = x(x-1)+x$ *yukarıda yerine yazılırsa*

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} \frac{[x(x-1)+x]n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \frac{x(x-1)n!}{x!(n-x)!} p^x q^{n-x} + \sum_{x=0}^{n} \frac{xn!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)![n-2-(x-2)]!} p^x q^{n-x} + E(X) \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)![n-2-(x-2)]!} p^{x-2} q^{n-2-(x-2)} + E(X) \\
&= n(n-1)p^2 (p+q)^{n-2} + np \\
&= n(n-1)p^2 + np \\
&= n^2p^2 - np^2 + np
\end{aligned}$$

bulunur. O halde

$$\begin{aligned}
\sigma^2 = E(X^2) - [E(X)]^2 &= n^2p^2 - np^2 + np - (np)^2 \\
&= np(1-p) \\
&= npq
\end{aligned}$$

elde edilir.

Örnek 4.4 *Bir günün güneşli olma olasılığı 0.4 ise bir haftadaki güneşli günlerin sayısının ortalamasını ve standart sapmasını bulunuz.*

Çözüm 4.4 X *rastlantı değişkeni "bir haftadaki güneşli günlerin sayısı" olsun. Verilenlere göre,* $p = 0.4$ $q = 0.6$ *ve* $n = 7$ *dir. O halde*

$$\mu = E(X) = np = 7(0.4) = 2.8$$

ve

$$\sigma = \sqrt{npq} = \sqrt{7(0.4)(0.6)} = 1.30$$

elde edilir.

4.1.1 Çok Terimli Dağılım

Bir denemede sırasıyla $p_1, p_2, \ldots, p_k$ olasılıklarına sahip $A_1, A_2, \ldots, A_k$ ayrık olaylarının meydana geldiğini kabul edelim (Burda $p_1+p_2+\ldots+p_k = 1$ dir). n bağımsız deneme yapıldığını farzedelim. x_1 tane A_1, x_2 tane A_2, …, x_k tane A_k meydana gelme olasılığı

$$f(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, \quad x_i = 0, 1, \ldots, n; \quad i = 1, 2, \ldots, k$$

dır. Burada

$$\textstyle\sum_{i=1}^{k} x_i = n \quad \text{ve} \quad \sum_{i=1}^{k} p_i = 1$$

şeklindedir.

Örnek 4.5 *Bir zar 12 kez atılsın. İki kez 1, üç kez 2, bir kez 3, iki kez 4, üç kez 5, bir kez altı gelmesi olasılığı nedir?*

Çözüm 4.5 $A_1 = 2,\ A_2 = 3,\ A_3 = 1,\ A_4 = 2,\ A_5 = 3,\ A_6 = 1,\ n = 12$ *ve*

$$p_i = \tfrac{1}{6},\ i = 1, \ldots, 6$$

olup

$$\begin{aligned} P &= f(2,3,1,2,3,1) \\ &= \frac{12!}{2!3!1!2!3!1!} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^1 \\ &= \frac{11!}{12} \left(\frac{1}{6}\right)^{12} \end{aligned}$$

elde edilir.

4.2 Poisson Dağılımı

Binom dağılımında $\mu = np$ ve olasılık fonksiyonu

$$f(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

dır. Burada, $p = \frac{\mu}{n}$ olduğu dikkate alınarak

$$\begin{aligned}\binom{n}{r} p^x q^{n-x} &= \frac{n(n-1)\ldots(n-x+1)}{x!}\mu\ n^x 1 - \frac{\mu^{n-x}}{n} \\ &= \frac{n(n-1)\ldots(n-x+1)}{n^x}\mu^x\ x! 1 - \frac{\mu^n}{n} 1 - \frac{\mu^{-x}}{n}\end{aligned}$$

yazılır. $n \to \infty$ için limit alınırsa

$$\frac{\mu^x}{x!}e^{-\mu}$$

elde edilir. Olasılık fonksiyonu

$$f(x) = P(X = x) = \frac{\mu^x}{x!}e^{-\mu}, \quad n = 0, 1, 2, \ldots$$

olan dağılıma Poisson dağılımı denir. Burada μ Poisson dağılımının ortalamasıdır. Varyansı da $\sigma^2 = \mu$ ile bulunur.

Poisson dağılımı görülen olaylara aşağıdaki gibi örnekler verilebilir:

i. Yolun belirli dar yerinde bir günde meydana gelen araba kazalarının sayısı,

ii. Bir fabrikada bir haftada meydana gelen kazaların sayısı,

iii. Verilen bir dakika içinde telefon santraline gelen telefon sayısı,

iv. Verilen belirli bir zamanda bir şirkete yapılan sigorta isteği sayısı,

v. Verilen bir zamanda bir radyoaktif kaynağı tarafından saçılan atomların sayısı

Örnek 4.6 *Bir mililitre sıvıdaki bakteri sayısı ortalama olarak 4 olduğu bilinmektedir. Bakterilerin sayısının Poisson dağılımı gösterdiği kabul edilerek 1 mililitre*

i. Hiç bakteri olmaması,

ii. 4 bakteri olması

iii. 3 den az bakteri olması

olsılığı nedir.

Çözüm 4.6 *X rastlantı değişkeni "1 mililitredeki bakterilerin sayısı" olsun. Bakteri sayısı ortalama 4 olduğundan $\mu = 4$ dür. O halde istenen olasılıklar*

$$f(x) = P(X = x) = \frac{4^x}{x!}e^{-4}, \quad n = 0, 1, 2, \ldots$$

formülü ile bulunur.

i. $P(X = 0) = e^{-4} = 0.0183$

ii. $P(X = 4) = (4^4/4!)e^{-4} = 0.195$

iii. 3 den az bakteri olması olasılığı,

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-4} + 4e^{-4} + \frac{4^2}{2!}e^{-4} \\ &= 0.238 \end{aligned}$$

olarak bulunur.

Örnek 4.7 *Civata fabrikasında bir bozuk civata üretme olasılığı $p = 0.01$ dir. 100 tane civatadan 2 den daha fazla bozuk civata çıkma olasılığı nedir.*

Çözüm 4.7 *A, "ikiden fazla bozuk civata olması" olayı ise $\bar{A}$, "ikiden fazla olmayan bozuk civata" olayı olur. Burada X rastlantı değişkenini "bozuk civata sayısı" olarak alalım. Buna göre*

$$\begin{aligned} P(\bar{A}) = P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= e^{-1}(1 + 1 + 0.5) \\ &= 0.919 \end{aligned}$$

bulunur. O halde

$$P(A) = 1 - P(\bar{A}) = 0.081$$

elde edilir.

4.3 Geometrik Dağılım

Bir deneyde A olayının meydana gelme olasılığı p olsun. A olayını elde edinceye kadar denemelere devam edelim. X rastlantı değişkeni ile bu denemelerin sayısını gösterlim. $X = x$ için A olayının meydana geldiği kabul edilirse $X = x$ in olasılığı $q = 1 - p$ olmak üzere

$$f(x) = P(X = x) = q^{x-1}p, \quad x = 0, 1, 2, \ldots$$

dır. Olasılık fonksiyonu bu olan dağılıma geometrik dağılım denir.

Örnek 4.8 *1 elde edinceye kadar bir zarı atalım.*

i. Bağımsız denemeler dizisinde, ilk 1 in elde edilmesi için gereken denemelerin sayısının olasılık fonksiyonu nedir?

ii. denemede 1 bulmanın olasılığı nedir?

Çözüm 4.8 *İlk 1'in elde edilmesi için gereken denemelerin sayısı X rastlantı değişkeni olsun. Bu takdirde $p = \frac{1}{6}$ olmak üzere, X geometrik dağılıma sahiptir.*

i. X'in olasılık fonksiyonu

$$f(x) = P(X = x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right), \quad x = 1, 2, \cdots$$

şeklindedir.

ii. üçüncü denemede 1 elde edilmesinin olasılığı

$$f(3) = P(X = 3) = \left(\frac{5}{6}\right)^{3-1}\left(\frac{1}{6}\right) = \frac{25}{256} = 0.0976562$$

bulunur.

Teorem 4.2 *Geometrik dağılım için ortalama değer ve varyans*

$$E(X) = \frac{1}{p} \quad ve \quad \sigma^2 = E(X^2) - [E(X)]^2 = \frac{q}{p^2}$$

olarak verilir.

İspat 4.2 *Ortalama değer için ispatı verelim.*

$$E(X) = \sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} xq^{x-1}p = 1p + 2pq + 3q^2p + \cdots$$

yazılır. Bilindiği gibi

$$1 + a + \cdots + a^n = \frac{1 - a^{n+1}}{1 - a}$$

olur. O halde ortalama değer

$$\mu = p\sum_{x=1}^{\infty} xq^{x-1} = p\frac{d}{dp}(\sum_{x=0}^{\infty} q^x) = p\frac{d}{dp}\left(\frac{1}{1-q}\right) = p\frac{1}{(1-q)^2} = \frac{1}{p}$$

elde edilir.

4.4 Problemler

1. Dört çocuklu bir ailede en az bir erkek çocuk olması olasılığını bulunuz?

2. Bir makinanın ürettiği civataların %20'si kusurlu çıkıyorsa, şansa bağlı olarak seçilen 4 civatadan,

   i. Birisinin kusurlu olma olasılığını,

   ii. Hiç birinin kusurlu olmama olasılığını,

   iii. En fazla ikisinin kusurlu olma olasılığını bulunuz.

3. Bir şirketin ürettiği pillerin %3'ünün kusurlu olması halinde, 100 pilden meydana gelen bir örnekte,

   i. 0   *ii*. 1   *iii*. 3

   iv. 3'ten daha fazla

   v. 2 veya daha fazla pilin kusurlu olması olasılıklarını bulunuz.

4. Bir hastanenin acil servisine günde ortalama 60 hasta gelmektedir. Buna göre, günün belli bir saatinde acil servise en az iki hasta gelmesi olasılığını bulunuz.

5. X rastlantı değişkeni Bernouilli dağılımına sahipse, $k = 1, 2, \cdots$ için

$$E(X^k) = p$$

olduğunu gösteriniz.

6. $X_1, X_2, \cdots, X_n$, $E(X_i) = p$ olan Bernouilli değişkenlerinin bir dizisi ise,

$i.$ $\quad E(\sum_{i=1}^{n} X_i) = np,$ $\qquad$ $ii.$ $\quad Var(\sum_{i=1}^{n} X_i) = npq$

olduğunu gösteriniz.

7. Bir zar 4 kez atılsın.

   $i.$ 4 atışta 3 kez 2 gelmesi,

   $ii.$ 4 atışta 2 kez 2 gelmesi olasılığı nedir.

8. Bir basketbol oyuncusunun tek atışta basket elde etme olasılığı1/5 tir. 25 atışta en az 4 basket elde etme olasılığı nedir.

9. %10'u kusurlu olan elektrik ampullerinin teslim anında 3 ampul rastgele çekiliyor,

   $i.$ 3'ü de sağlam,

   $ii.$ 1 kusurlu,

   $iii.$ 2 kusurlu ampul çekilmiş olması olasılıklarını bulunuz.

10. Bir fabrikadaki belli bir makinada üretilen parçaların kusurlu olması olasılığı 0.1 dir. Parçalar arasından 8'i seçiliyor.

    $i.$ En az birinin kusurlu olması,

    $ii.$ Parçaların tümünün kusursuz olması olasılığı nedir.

11. Bir üniversitedeki öğrencilerin %60'ı lisans, %30'ı yüksek lisans ve %10'u doktora öğrencisidir. 5 kişilik bir kurul senatoya katılacaktır. Her bir şahıs diğerlerinden bağımsız olarak seçiliyor. Kurulda 2 lisans, 2 yüksek lisans ve 1 de doktora öğrencisi bulunması olasılığı nedir.

12. Bir sigortacı hepsi aynı yaşta ve sıhatli 10 kişiyle sigorta anlaşması yapmıştır. Yaşama süresi olasılık tablosuna göre bu şahısların yaşında olan birisinin 2/3 olasılıkla 40 yıl yaşayacağı tahmin edilmektedir. Bu 10 kişinin,

    i. Hepsinin,

    ii. En az üçünün,

    iii. Sadece ikisinin ve

    iv. En az birisinin 30 yıl yaşama olasılığını bulunuz.

13. Üniversiteye giren erkek öğrencilerin üçte birinin en az 70 kg olduğu saptanmıştır. Rastgele seçilen 4 öğrencinin en az üçünün 70 kg dan az olması olasılığı nedir.

Chapter 5

SÜREKLİ RASLANTI DEĞİŞKENLERİN DAĞILIMLARI

5.1 Normal Dağılım

Bu kısımda, olasılık teorisi ve istatistiksel analizde çok kullanılan normal dağılımdan bahsedeceğiz.

5.1.1 Normal Yoğunluk Fonksiyonu

Tanım 5.1 *X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad (\sigma > 0), \tag{5.1}$$

şeklinde ise, X **normal dağılıma** *sahiptir denir.*

Burada μ ve σ sırasıyla normal dağılımın ortalaması ve standart sapmasıdır. $f(x)$ fonksiyonunun eğrisi $x = \mu$ ya göre simetriktir. Eğer $\mu = 0$ ise, $f(x)$ eğrisi ordinata göre simetrik olur. Ordinata göre simetrik olan bu eğriye **çan eğrisi** denir. σ^2 ye de dağılımın varyansı adı verilir. Aşağıdaki şekilde görüldüğü gibi çan eğrisi varyansın değerine göre yüksek ve daha sarp yada alçak ve daha yayvan olur.

Şekil 5.1:

Tanım 5.2 *(5.1) fonksiyonunun* $-\infty$ *'dan* x*'e kadar integralini alırsak*

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{v-\mu}{\sigma})^2} dv. \tag{5.2}$$

bulunur. Bu fonksiyona **normal dağılım fonksiyonu** *denir.*

$a < x \leq b$ olmak üzere, X rastlantı değişkeninin olasılığı

$$P(a < X \leq b) = F(b) - F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}(\frac{v-\mu}{\sigma})^2} dv. \tag{5.3}$$

şeklindedir. $f(x)$ olasılık yoğunluk fonksiyonu olduğundan

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\frac{v-\mu}{\sigma})^2} dv = 1 \tag{5.4}$$

yazılır.

(5.2) bildiğimiz integral metodları ile hesaplanamaz. Bu integrali pratik olarak hesaplayabilmek için özel olarak hazırlanmış bir tablodan faydalanılır.

Fakat genel olarak μ ve σ değerleri için (5.2) integralini hesap etmek oldukça zordur. Bu zorluğu gidermek için aşağıdaki düzenleme yapılır, Eğer $\mu = 0$, $\sigma^2 = 1$ ve $Z = (X - \mu)/\sigma$ alınırsa, Z nin olasılık yoğunluk fonksiyonu

$$\Phi(z) = P(Z < z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du. \tag{5.5}$$

olur. Z değişkenine **standart normal rastlantı değişkeni** denir. Çeşitli Z değerleri için $\Phi(z)$ 'nin tablosu kitabın sonunda (z cetveli)[1] verilmiştir.

Şimdi de $F(x)$ fonksiyonunu $\Phi(z)$ cinsinden ifade etmeye çalışalım: Eğer (5.2) ve (5.5) in üstel kısımlarını karşılaştıracak olursak

$$\frac{v-\mu}{\sigma} = u \Longrightarrow \frac{dv}{\sigma} = du \tag{5.6}$$

olur. Bu değişken değiştirme yapıldıktan sonra (5.2) de $v = x$ integral üst limiti yeni değişkene göre $u = (x - \mu)/\sigma$ olur. Buradan

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\mu^2/2} \sigma du \tag{5.7}$$

ve ya daha sade olarak

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\mu^2/2} du \tag{5.8}$$

yazılır. (5.5) de $z = (x - \mu)/\sigma$ alınırsa

$$F(x) = \Phi(\frac{x-\mu}{\sigma}) \tag{5.9}$$

elde edilir. Burada z ye **standart normal değişken** denir. Böylece (5.9) ve (5.3) formülleri kullanılarak

$$P(a < X \leq b) = F(b) - F(a) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}) \tag{5.10}$$

[1] z cetvelindeki ölçümler standart ölçümlerdir. Farklı dağılımlardaki farklı ölçümlerin karşılaştırılması için kullanılır. Örneğin, bir öğrenci Matematik sınavından 60, Bilgisayar Proğramlama sınavından 70 puan almış olsun. Bu sonuçlara bakarak, öğrencinin Bilgisayar Proğramlama dersinde daha başarılı olduğunu söyleyemeyiz. Çünkü 60 sınav yapılan grubun en yüksek puanlarından 70 de grubun en düşük puanlarından olabilir. O halde sadece puanlara bakarak bir karşılaştırma yapmak yerine onları ortak bir puan sistemine çevirmek uygun olacaktır. Standart ölçümler gerçek ölçümlerin ortalamadan olan farklarının standart sapmaya bölünmesiyle elde edilir.

bulunur. Özel olarak, eğer $a = \mu - \sigma$ ve $b = \mu + \sigma$ ise

$$P(a < X \leq b) = \Phi(1) - \Phi(-1);$$

$a = \mu - 2\sigma$ ve $b = \mu + 2\sigma$ ise

$$P(a < X \leq b) = \Phi(2) - \Phi(-2)$$

olur.
z cetvelini kullanarak,

i. $P(\mu - \sigma < X \leq \mu + \sigma) \approx \%68$,

ii. $P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx \%95.5$,

iii. $P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx \%99.7$

yazarız. *i.* ve *ii.* şıkkın şekilleri aşağıda verilmiştir.

Şekil 5.2:

Örnek 5.1 *Ortalaması* $\mu = 0$ *ve varyansı* $\sigma^2 = 1$ *olan* X *normal rastlantı değişkeni için aşağıdaki olasılıkları bulunuz?*

i. $P(X \leq 2.44)$ *ii.* $P(X \leq -1.16)$ *iii.* $P(X \geq 1)$ *iv.* $P(2 \leq X \leq 10)$.

Çözüm 5.1 $\mu = 0$ *ve* $\sigma^2 = 1$ *olduğundan direkt olarak* z *cetvelinden;*

i. $P(X \leq 2.44) = 0.9927$,

ii. $P(X \leq -1.16) = 0.1230$,

iii. $1 - P(X \leq 1) = 1 - 0.8413 = 0.1587,$

iv. $\Phi(10) = 1.0000,\ \Phi(2) = 0.9772,\ \Phi(10) - \Phi(2) = 0.0228$

bulunur.

Örnek 5.2 *Ortalaması $\mu = 0.8$ ve varyansı $\sigma^2 = 4$ olan X normal rastlantı değişkeni için aşağıdaki olasılıkları bulunuz?*

i. $P(X \leq 2.44)$ *ii.* $P(X \leq -1.16)$ *iii.* $P(X \geq 1)$ *iv.* $P(2 \leq X \leq 10)$.

Çözüm 5.2 $\mu = 0.8$ *ve* $\sigma^2 = 4$ *olduğundan 5.10 formülünü kullanarak z cetvelinden;*

i. $F(2.44) = \Phi(\frac{2.44-0.80}{2}) = \Phi(0.82) = 0.7939,$

ii. $F(-1.16) = \Phi(-0.98) = 0.1635,$

iii. $1 - P(X \leq 1) = 1 - F(1) = 1 - \Phi(0.1) = 0.4602,$

iv. $F(10) - F(2) = \Phi(4.6) - \Phi(0.6) = 1 - 0.7257 = 0.2743$

yazılır.

Örnek 5.3 *Ortalaması $\mu = 0$ ve varyansı $\sigma^2 = 1$ olan X normal rastlantı değişkeninin aşağıdaki olasılıkları için c değerini bulunuz?*

$$\begin{array}{llll} i. & P(X \geq c) = \%10 & ii. & P(0 \leq X \leq c) = \%45 \\ iii. & P(-c \leq X \leq c) = \%99 & iv. & P(X \leq c) = \%5. \end{array}$$

Çözüm 5.3 *z cetveli kullanılırsa:*

$$\begin{array}{ll} i. & 1 - P(X \leq c) = 1 - \Phi(c) = 0.1,\ \Phi(c) = 0.9,\ c = 1.282 \\ ii. & \Phi(c) - \Phi(0) = \Phi(c) - 0.5 = 0.45,\ \Phi(c) = 0.95,\ c = 1.645 \\ iii. & c = 2.576 \\ iv. & c = -1.645 \end{array}$$

bulunur.

Örnek 5.4 *Ortalaması $\mu = -2$ ve varyansı $\sigma^2 = 0.25$ olan X normal rastlantı değişkeninin aşağıdaki olasılıkları için c değerini bulunuz.*

i. $P(-2-c \leq X \leq -2+c) = \%99.6$ *ii.* $P(-c \leq X \leq -1) = 0.5$
iii. $P(-2-c \leq X \leq -2+c) = 0.9$ *iv.* $P(X \geq c) = 0.2.$

Çözüm 5.4 *z cetveli kullanılırsa;*

i. $\Phi(2c) - \Phi(-2c) = \%99.6,\ 2c = 2.878,\ c = 1.439,$
ii. $\Phi(\frac{-1+2}{0.5}) - \Phi(\frac{-c+2}{0.5}) = 0.9772 - \Phi(4-2c) = 0.5,$
$\Phi(4-2c) = 0.4772,\ 4-2c = -0.057,\ c = 2.03.$
iii. $\Phi(\frac{-1+2+c}{0.5}) - \Phi(\frac{-2-c+2}{0.5}) = \Phi(2c) - \Phi(-2c) = 0.9,\ 2c = 1.645,$
$c = 0.823.$
iv. $1 - P(X \leq c) = 1 - \Phi(\frac{c+2}{0.5}) = 0.2,$
$\Phi(2c+4) = 0.8,\ 2c+4 = 0.842,\ c = -1.579$

bulunur.

5.2 Düzgün Dağılım

X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu

$$f(x) = \frac{1}{b-a},\ a \leq x \leq b \tag{5.11}$$

şeklinde ise X, $[a,b]$ aralığında **düzgün dağılıma sahiptir** denir. X bir sürekli rastlantı değişkeni olduğundan,

$$\int_a^b f(x)dx = \int_a^b \frac{1}{b-a}dx = 1 \tag{5.12}$$

dir. $y = f(x)$'in grafiği aşağıdaki gibidir.

.

Şekil 5.3:

Teorem 5.1 *X rastlantı değişkeni $[a, b]$ aralığında düzgün dağılıma sahipse,*

$$i.\ E(X) \;=\; \tfrac{1}{2}(a+b) \qquad ii.\ V(X) \;=\; \tfrac{1}{12}(b-a)^2$$

dir.

İspat 5.1 *: i.*

$$E(X) \;=\; \int_a^b x\frac{1}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{1}{2}(a+b)$$

ii.

$$\begin{aligned} V(X) \;&= E(X^2) - [E(X)]^2 = \int_a^b x^2\frac{1}{b-a}dx - (\frac{a+b}{2})^2 \\ &= \frac{b^3-a^3}{3(b-a)} - (\frac{a+b}{2})^2 = \frac{1}{12}(b-a)^2 \end{aligned}$$

elde edilir.

Örnek 5.5
Bir dikdörtgenin bir kenarının uzunluğu x cm, çevre uzunluğunun toplamı 12 cm ve alanı ise A cm^2 dir. X, $[0, 2]$ aralığında düzgün dağılıma sahipse A 'nın olasılık yoğunluk fonksiyonunu bulunuz.

Çözüm 5.5 *A'nın olasılık yoğunluk fonksiyonu $f(a)$ olsun. Dikdörtgenin bir keneri x cm ve çevre uzunluğu 12 cm olduğuna göre diğer kenarının uzunluğu $(6-x)$ cm dir. Buna göre;*

$$\begin{aligned} a &= x(6-x) \\ x^2-6x &= -a \\ (x-3)^2 &= 9-a \\ x-3 &= \pm\sqrt{9-a} \\ dx &= \pm\frac{da}{2\sqrt{9-a}} \end{aligned}$$

olur. X, $[0,2]$ aralığında düzgün dağılıma sahip olduğundan, $x=0$ ve $a=0$ olduğunda;

$$1 = \int_0^2 \frac{1}{2}dx$$

olur. Burada $a = x(6-x)$ değişken değiştirmesi yapılırsa, sınırlar da $x=0$ için $a=0$ ve $x=2$ için $a=8$ olur. Böylece,

$$1 = \int_0^8 \frac{1}{2}\frac{da}{2\sqrt{9-a}}$$

yazılır. Buradan A nın olasılık yoğunluk fonksiyonu:

$$f(a) = \frac{1}{4\sqrt{9-a}}, \quad 0 \le a \le 8$$

olur. Yani A, düzgün dağılıma sahiptir.

5.3 Üstel Dağılım

Tanım 5.3 *Pozitif değerler alan X sürekli rastlantı degişkeninin olasılık yoğunluk fonksiyonu,*

$$f(x) = \lambda e^{-\lambda x}, \quad x \ge 0$$

olsun. Bu takdirde, λ pozitif bir parametre olmak üzere, X sürekli rastlantı değişkeni **üstel dağılıma sahiptir** *denir.*

X bir rastlantı değişken olduğundan;

$$\begin{aligned}
\int_0^\infty f(x)dx &= \int_0^\infty \lambda e^{-\lambda x} dx \\
&= -[e^{-\lambda x}]_0^\infty \\
&= 1 \quad (\text{Çünkü} \lim_{x\to\infty} e^{-\lambda x} = 0)
\end{aligned} \tag{5.13}$$

dir. $y = f(x)$'in grafiği aşağıdaki gibidir.

Şekil 5.4:

Teorem 5.2 *X rastlantığişkeni $[a, b]$ aralığında üstel dağılıma sahip ise,*

$$i.\ E(X) = \tfrac{1}{\lambda} \qquad ii.\ V(X) = \tfrac{1}{\lambda^2}$$

dir.

İspat 5.2 *: i.*

$$\begin{aligned}
E(X) &= \int_x x f(x) dx \\
&= \int_0^\infty x(\lambda e^{-\lambda x}) dx \\
&= [x(-e^{-\lambda x})]_0^\infty - \int_0^\infty (-e^{-\lambda x}) dx \\
&= 0 + \int_0^\infty (e^{-\lambda x}) dx, \quad (\textit{Çünkü} \lim_{x\to\infty} x e^{-\lambda x} = 0) \\
&= -\frac{1}{\lambda}[e^{-\lambda x}]_0^\infty \\
&= -\frac{1}{\lambda}(0-1) \\
&= \frac{1}{\lambda}
\end{aligned}$$

ii.

$$
\begin{aligned}
E(X^2) &= \int_x x^2 f(x)dx \\
&= \int_0^{\infty} x^2(\lambda e^{-\lambda x})dx \\
&= [x^2(-e^{-\lambda x})]_0^{\infty} - \int_0^{\infty} 2x(-e^{-\lambda x})dx \\
&= 0 + 2\int_0^{\infty} x(e^{-\lambda x})dx, \quad (\textit{Çünkü} \lim_{x\to\infty} x^2 e^{-\lambda x} = 0) \\
&= \frac{2}{\lambda^2}, \quad (\textit{Çünkü} \int_0^{\infty} x(\lambda e^{-\lambda x})dx = \frac{1}{\lambda})
\end{aligned}
$$

$$
\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

elde edilir.

5.3.1 Üstel Dağılımın Özellikleri

Eğer X üstel dağılıma sahipse:

i.

$$
\begin{aligned}
P(X > a) &= \int_a^{\infty} \lambda e^{-\lambda x}dx \\
&= [-e^{-\lambda x}]_a^{\infty} \\
&= e^{-\lambda a}
\end{aligned}
$$

ii.

$$
\begin{aligned}
P(X > a + b | X > a) &= \frac{P(X > a + b \bigcap X > a)}{P(X > a)} \\
&= \frac{e^2 - \lambda(a + b)}{e^{-\lambda a}} \\
&= e^{-\lambda b} \\
&= P(X > b)
\end{aligned}
$$

yazılır. Benzer şekilde:

iii. $P(X \leq a) = 1 - P(X > a), \quad x \geq 0$ ve $P(X \leq a) = 0, \quad x < 0$ olur.

Örnek 5.6 *T rastlantı değişkeni bir televizyon tüpünün dayanıklılık süresi ve T'nin olasılık yoğunluk fonksiyonu:*

$$f(t) = \begin{cases} Ae^{-kt} & , \quad 0 \leq t \leq \infty, \ (k > 0) \\ 0 & , \quad \textit{diğer yerlerde} \end{cases}$$

olarak veriliyor. A'yı k cinsinden hesaplayınız. Buna göre,

i. Bir fabrikada yapılan bir araştırmaya göre üretilen 1000 tüpün 371'i ilk iki yıl içinde bozulmaktadır. Buna göre k değerini bulunuz.

ii. (i) de elde edilen k değerini kullanarak T'nin Ortalaması ve Varyansını hesaplayınız.

iii. Satılan iki tüpten birisinin ilk yıl içinde bozulma ve diğer tüpün ise 6 yıldan fazla dayanma olasılığı nedir.

Çözüm 5.6 *T rastlantı değişkeni için $\int_0^\infty f(t)dt = 1$ olduğunu biliyoruz. Böylece,*

$$\begin{aligned} \int_0^\infty Ae^{-kt}dt &= 1 \\ A\left[-\frac{1}{k}e^{-kt}\right]_0^\infty &= 1 \\ \frac{A}{k}e^{-0} &= 1 \\ A &= k \end{aligned}$$

olur. Yani T, olasılık yoğunluk fonksiyonu $f(t) = ke^{-kt}, \ \ 0 \leq t \leq \infty$ şeklinde üstel dağılıma sahiptir.

i. $P(T \leq 2) = 0.371$ *olduğundan,*

$$\begin{aligned} \int_0^2 ke^{-kt}dt &= 0.371 \\ \left[-e^{-kt}\right]_0^2 &= 0.371 \end{aligned}$$

$$\begin{aligned}
-e^{-2k} + e^{0} &= 0.371 \\
e^{-2k} &= 1 - 0.371 \\
&= 0.629 \\
e^{2k} = \frac{1}{0.629} & \\
2k &= ln\left(\frac{1}{0.629}\right) \\
&= 0.464 \\
k &= 0.232
\end{aligned}$$

olarak bulunur.

ii.

$$\begin{aligned}
E(T) &= \frac{1}{k} \\
&= \frac{1}{0.232} \\
&= 4.3\, yıl
\end{aligned}$$

ve

$$\begin{aligned}
V(T) &= \frac{1}{k^2} \\
&= \frac{1}{(0.232)^2} \\
&= 19
\end{aligned}$$

olur.

iii.

$$\begin{aligned}
P(T < 1) &= \int_0^1 ke^{-kt}dt \\
&= \left[-e^{-kt}\right]_0^1 \\
&= -e^{-k} + 1 \\
&= 1 - 0.793 \\
&= 0.207
\end{aligned}$$

ve

$$P(T > 6) = 1 - P(T \leq 6)$$

$$
\begin{aligned}
&= 1 - \int_0^6 ke^{-kt}dt \\
&= 1 - \left[-e^{-kt}\right]_0^6 \\
&= 1 + e^{-6k} - 1 \\
&= 0.249
\end{aligned}
$$

şeklinde bulunur. İki tüp satıldığına göre istenen olasılık

$$
\begin{aligned}
P[(T_1 < 1) \cap (T_2 > 6)] + P[(T_2 < 1) \cap (T_1 > 6)] &= 2(0.207)(0.249) \\
&= 0.103
\end{aligned}
$$

olur.

5.4 Gamma Dağılımı

Tanım 5.4

$$\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx, \quad p > 0 \tag{5.14}$$

şeklinde tanımlanan fonksiyona **Gamma Fonsiyonu** *denir.*

Yukarıdaki integrale kısmi integrasyon metodu uygulanırsa, $e^{-x}dx = dv$ ve $x^{p-1} = u$ olmak üzere,

$$
\begin{aligned}
\Gamma(p) &= e^{-x}x^{p-1}|_0^\infty - \int_0^\infty \left[x^{p-2}(-e^{-x})\right] dx \\
&= 0 + (p-1)\int_0^\infty \left[x^{p-2}e^{-x}\right] dx \\
&= (p-1)\Gamma(p-1)
\end{aligned} \tag{5.15}
$$

bulunur. p pozitif tam sayı ise

$$\Gamma(p) = (p-1)(p-2)\cdots\Gamma(1) \tag{5.16}$$

elde edilir. Burada,

$$\Gamma(1) = \int_0^\infty e^{-x}dx = 1 \tag{5.17}$$

dir. Böylece,

$$\Gamma(p) = (p-1)! \tag{5.18}$$

olarak bulunur.

Tanım 5.5 *X pozitif değerler alan sürekli rastlantı değişkeni olsun. X'in olasılık yoğunluk fonksiyonu,*

$$f(x) \;=\; \begin{cases} \frac{1}{\Gamma(n)\lambda^n} x^{n-1} e^{-x/\lambda} & , \quad x > 0 \;\; \textit{için}, \\ 0 & , \quad x \leq 0 \;\; \textit{için} \end{cases} \tag{5.19}$$

ise, X, **Gamma dağılımına sahiptir** *denir.*

Burada λ ve n parametrelerdir. Gamma dağılımında $n = 1$ alındığında, elde edilen dağılımın üstel dağılım olduğu görülür.

X Gamma dağılımına sahip ise, X'in ortalaması $E(X) = \lambda n$ ve varyansı $V(X) = \lambda^2 n$' dir.

5.5 Beta Dağılımı

Tanım 5.6

$$B(\alpha, \beta) \;=\; \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx, \;\; \alpha > 0, \;\; \beta > 0 \tag{5.20}$$

fonksiyonuna **Beta fonksiyonu** *denir.*

Ayrıca,

$$B(\alpha, \beta) = B(\beta, \alpha)$$

ve

$$B(\alpha, \beta) \;=\; \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \;\; \alpha > 0, \beta > 0 \tag{5.21}$$

dır.

Tanım 5.7 *Eğer X rastlantı değişkeninin olasılık yoğunluk fonksiyonu*

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & , \quad 0 < x < 1 \\ 0 & , \quad \textit{aralığın dışında} \end{cases} \tag{5.22}$$

ise, X rastlantı değişkeni Beta dağılımına sahiptir denir.

Burada α, β parametreleri pozitif reel sayılardır.

5.6 Problemler

1. Ortalaması $\mu = 80$ ve varyansı $\sigma^2 = 9$ olan X normal rastlantı değişkeninin aşağıdaki olasılıklarını bulunuz.

   i. $P(X > 83)$ *ii.* $P(X < 81)$ *iii.* $P(X < 80)$ *iv.* $P(78 < X < 82)$.

2. Ortalaması $\mu = 14$ ve varyansı $\sigma^2 = 4$ olan X normal rastlantı değişkeninin aşağıda verilen olasılıkları için c değerini bulunuz.

   i. $P(X \leq c) = \%95$ *ii.* $P(X \leq c) = \%5$
   iii. $P(X \leq c) = \%99.5$ *iv.* $P(-c < X - 3.6 \leq c) = \%99.9$.

3. Belli bir insan grubunun, baş ölçümü X'in şahıslar arasında normal dağıldığını kabul edelim. Grubun %50'si ($X \leq 75$) uzun kafalı, %38'i ($75 < X \leq 80$) orta kafalı, ve %4 ü ($80 < X$) yuvarlak kafalıdır. X'in ortalama ve varyansını bulunuz.

4. Bir basketbol oyuncusu faul atışlarının ortalama olarak %60 ında sayı elde edebiliyor. Bir sezon boyunca 100 atış yapıyor.

   i. 70 veya daha çok sayı elde etmesi olasılığını bulunuz.

   ii. Tam 60 sayı elde edebilmesi olasılığı nedir.

5. Bir sınavda 10 soru soruluyor ve notlar 10 üzerinden değerlendiriliyor. Verilen notlar ortalaması $\mu = 6.7$ olan normal olasılık fonksiyonuna uygun olarak dağılıyorlar ve $P(X > 8) = 0.10$ dur.

   i. $P(|X| \leq 7)$ olasılığını bulunuz.

ii. En düşük not alan %15 üğrenci içinde en yüksek not nedir?

6. Normal dağılıma uyan bir rastlantı değişkeninin 50'den küçük değer alması olasılığı 0.10 ve 100'den büyük değer alması olasılığı 0.05'tir.

   i. $P(X > 7)$ olasılığını bulunuz?

   ii. $P[(X \leq 70)/(60 \leq X \leq 80)]$ olasılığını bulunuz?

   iii. $P(X > a) = 0.15$ eşitsizliğini sağlayan a değerini bulunuz?

7. X sürekli rastlantı değişkeninin olasılık yoğunluk fonksiyonu, $1/\lambda = 1/50$ olan üstel olasılık yoğunluk fonksiyonuna uymaktadır. $Y = 2X+1$ rastlantı değişkeninin olasılık yoğunluk ve dağılım fonksiyonlarını bulunuz?

8. Bir banliyö treni bir istasyona, programa göre saat 11.00' de varmaktadır. Bu trenin istasyona varma zamanı olan X rastlantı değişkeni $10.55 - 11.10$ aralığında değişen düzgün dağılıma uymaktadır. Trenin programda belirtilen saatten en çok üç dakika sonra istasyonda olması olasılığını bulunuz?

Chapter 6

ÖRNEKLEM SEÇİMİ VE VERİLERİN DÜZENLENMESİ

İstatistik, verilerin toplanması, düzenlenmesi, takdimi, analizi ve bu analizlerle elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir. Buna göre istatistiği ikiye ayırabiliriz:
i. Deskriptif (tasviri) istatistik,
ii. İndaktif (analiz edici, tümevarımsal) istatistik.
Frekans dağılımları, yer ölçüleri (aritmetik ortalama, geometrik ortalama, harmonik ortalama, kareli ortalama, mod, medyan,...), dağılım ölçüleri (ortalama sapma, standart sapma, varyans, değişim aralığı, ...), çarpıklık ve basıklık gibi konular, verilerin özetlenmesi ve tasviri ile ilgili olduklarından, tasviri istatistiğin konusuna girerler. Örnekleme teorisi, hipotez testleri, regresyon ve korelesyon analizleri gibi konular ise indaktif istatistiği meydana getirirler.

6.1 Örneklem Kavramı ve Seçimi

Bir istatistiki araştırmada, araştırmayı konu alan bütün birimlere **kitle** denir. Uygulamada bir kitlenin bütünüyle incelenmesi hem zaman hem de işgücü bakımından büyük güçlükler ve masraflar doğuracağından, çoğu zaman kitlenin tamamı yerine bu kitleden alınan bir grubun incelenmesi yoluna gidilir. Bu gruptan elde edilen sonuçlar bütün kitleye genelleştirilir. İşte kitleyi temsil etmek üzere seçilen söz konusu gruba **örneklem** denilir. Örneğin, bir üniversitedeki bütün öğrenciler kitle, bu öğrencilerden alınan 100 kişilik bir şans grubu ise örneklemdir. Eğer araştırmamızı bir fakülte için yapıyorsak, fakültenin bütün öğrencileri kitle, bu öğrenciler içerisinden alacağımız, örneğin, 30 kişilik bir grup örneklemdir.

Örneklem seçerken, örneklem istenilen oranın altında olmamalıdır. Örneklem kötü bir şekilde seçilmişse, matematiksel ve istatistiksel metodlar bizi istenilen sonuca götürmez. Örneklem seçiminde en çok dikkat edilmesi gereken nokta örneklemin kitleyi temsil edici olmasıdır.

6.2 Frekans Dağılımları

Bir araştırmaya konu olan bütün verileri

i. Gözlendiği şekilde (basit veri),

ii. Her bir veriyi frekansları ile belirterek (sınıflandırılmış veri),

iii. Belirli aralıktaki verileri frekansları ile belirterek (gruplandırılmış veri)

düzenleyebiliriz.

Basit verileri büyüklük sırasına göre de düzenleyebiliriz. Bunlara dizi adını vereceğiz. Sınıflandırılmış ve gruplandırılmış verilerde kullanılan frekans (f), bir verinin tekrar sayısını göstermektedir. Örneğin, bir gözlemde bir A olayı 7 defa görülmüşse bu A olayının frekansı 7 dir. Gruplandırma, diğer iki düzenlemeye göre daha kullanışlıdır.

Örnek 6.1 *23 kişilik bir öğrenci grubunun ayakkabı numaraları ile ilgili veriler,*

$$35, 35, 37, 37, 37, 37, 38, 38, 38, 38, 38, 40, 40, 40, 40, 40, 40, 41, 41, 41, 42, 42, 43$$

olsun (bunlar basit verilerdir). Bu verileri sınıflandırınız.

Çözüm 6.1 *2 tane 35, 4 tane 37, 5 tane 38, 6 tane 40, 3 tane 41, 2 tane 42 ve 1 tane 43 numara ayakkabı olduğundan şu şekilde sınıflandırma yapılır:*

Ayakkabı numarası (X)	*35*	*37*	*38*	*40*	*41*	*42*	*43*	
Frekans (f)	*2*	*4*	*5*	*6*	*3*	*2*	*1*	*Toplam* = 23

Örnek 6.2 *30 kişilik bir sınıfta Matematik sınavının sonuçları,*

$$3, 4, 4, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9, 10, 10, 10, 11, 11, 11, 11, 11, 13$$

şeklindedir (notlar 20 üzerinden verilmiştir). Bu verileri gruplandırınız.

Çözüm 6.2

X	*3-5*	*6-8*	*9-11*	*12-14*	
f	*4*	*10*	*15*	*1*	*Toplam* = 30

Şimdi verileri gruplandırmada kullanılan bazı terimleri tanıtalım:

Tanım 6.1 *En büyük gözlem değeri ile en küçük gözlem değeri arasındaki farka, değişimin genişliği (veya değişimin büyüklüğü) denir.*

Örnek 6.2 deki gözleme göre değişimin büyüklüğü (D.B): D.B $= 13 - 3 = 10$ olur.

Tanım 6.2 *N tane verinin değişim büyüklüğünün bölündüğü aralıklara sınıf aralıkları denir.*

Örnek 6.2 deki 3-5, 6-8, 9-11, 12-14 her biri sınıf aralığıdır.

Tanım 6.3 *Bir sınıf aralığının iki uç noktasına sınıf limitleri denir.*

Örnek 6.2 de 3-5 aralığındaki 3 ve 5 bu aralığın sınıf limitleridir. Burada 3 'e aralığın alt limiti, 5 'e ise aralığın üst limiti denir.

Tanım 6.4 *Bir sınıf aralığının alt limiti ile kendisinden önceki sınıf aralığının üst limitinin aritmetik ortalamasına o sınıf aralığının alt sınırı; kendisinin üst limiti ile kendisinden sonraki sınıf aralığının alt limitinin aritmetik ortalamasına da o sınıf aralığının üst sınırı denir.*

Örnek 6.2 deki sınıf aralıklarının sınırları 2.5-5.5, 5.5-8.5, 8.5-11.5, 11.5-14.5 olur.

Tanım 6.5 *Bir sınıf aralığının merkezine veya sınıf limitleri arasındaki orta noktaya sınıf ortası (veya sınıf değeri) denir.*

Örnek 6.2 de 3-5 sınıf aralığının sınıf değeri $\frac{3+5}{2} = 4$ dür.

Tanım 6.6 *Gruplandırılmış bir veride her sınıf aralığının ilk rakamına alt sınıf ucu, son rakamına ise üst sınıf ucu denir.*

Örnek 6.2 de, birinci grubun sınıf ucları 3 ile 5 tir.

Tanım 6.7 *Bir sınıf aralığının alt ve üst sınırları arasındaki farka sınıf büyüklüğü denir.*

Örnek 6.2 de gruplandırılmış verilerde birinci sınıfın sınıf büyüklüğü, $5.5 - 2.5 = 3$ tür.

Gruplamada hesap kolaylığı bakımından genellikle, gruplar eşit büyüklükte alınır. Bununla beraber, uygulamada farklı büyüklükte gruplardan meydana gelen diziler yanında, ilk veya son sınıf aralıkları açık olan gruplara da rastlanmaktadır. Gruplandırılmış verilerle matematiksel işlemler yapabilmemiz için, her bir sınıf aralığını bir sayı ile temsil etmek mecburiyetinde kalırız. Bu sayı sınıf ortası veya sınıf değeri dediğimiz sayıdır.

Nisbi Frekans: Verideki her bir frekansın toplam frekansa bölünmesiyle nisbi frekanslar elde edilir. Bu frekanslar *nisbi frekans dağılımını* oluştururlar.

Örnek 6.2' de gruplandırılmış veriler için "nisbi frekans dağılımı" aşağıdaki gibidir.

Gruplar	3-5	6-8	9-11	12-14	
Frekans	4	10	15	1	Toplam $= 30$
N.Frekans	$4/30$	$10/30$	$15/30$	$1/30$	1

Kümülatif Frekans: Kümülatif, birikmiş veya yığılmış manasına gelmektedir. Kümülatif frekans, ".....den az kümülatif frekans " ve "...... den çok kümülatif frekans" olmak üzere ikiye ayrılır. Örnek 6.2 'de gruplandırılmış veriler için bu kümülatif frekanslar aşağıdaki gibi olur.

...den az	küm. fre.	....den çok	küm.fre.
2.5 den az	0	2.5 den çok	30
5.5 "	4	5.5 "	26
8.5 "	14	8.5 "	16
11.5 "	29	11.5 "	1
14.5 "	30	14.5 "	0

6.3 Frekans Histogramı ve Frekans Poligonu

Histogram bir dikdörtgenler dizisidir. Bu dikdörtgenlerin tabanları gruplandırılmış verilerdeki herbir sınıfın sınıf büyüklüğünü, yükseklikleri ise sınıf frekansını gösterir. Bu dikdörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe "frekans poligonu" denilir. Frekans poligonu yatay ekseni, ilk sınıftan bir önceki sınıfın orta noktasıyla, son sınıftan bir sonraki sınıfın orta noktasında keser. Frekans histogramı ve frekans poligonunun altında kalan toplam alan birbirine eşittir.

Örnek 6.3 *Aşağıdaki veriler,* 100 *öğrencinin istatistik dersinden aldıkları notları göstermektedir. Buna göre,*

i. Bu verileri bir dizi haline getiriniz.

ii. Notların değişim aralığını bulunuz.

iii. 50 *den az puan alan öğrenci sayısını bulunuz.*

iv. 70 *veya daha fazla puan alan öğrenci sayısını bulunuz.*

v. Bu verileri, sınıf büyüklüğü 10 *olacak şekilde gruplandırarak bir frekans dağılımı meydana getiriniz.*

vi. (v).ci şıktaki sınıflandırma için her sınıfa ait sınıf değerini ve sınıf sınırlarını bulunuz.

vii. Bu veriler için, (v).ci şıktaki sınıfları dikkate alarak bir frekans histogramı ve frekans poligonu oluşturunuz.

89	57	57	93	97	65	95	33	71	42
85	47	63	76	49	88	82	50	62	93
39	76	84	54	87	56	39	49	70	74
77	81	99	73	48	57	89	55	73	86
35	66	79	85	72	91	39	49	84	66
58	65	81	84	43	62	95	58	64	77
93	74	68	76	65	61	81	37	45	89
22	50	68	87	51	45	89	95	87	91
43	71	93	67	49	65	96	84	79	77
80	93	65	89	99	64	53	62	86	69

Çözüm 6.3 *i.*

22	33	35	37	39	39	39	42	43	43
45	45	47	48	49	49	49	49	50	50
51	53	54	55	56	57	57	57	58	58
61	62	62	62	63	64	64	65	65	65
65	65	66	66	67	68	68	69	70	71
71	72	73	73	74	74	76	76	76	77
77	77	79	79	80	81	81	81	82	84
84	84	84	85	85	86	86	87	87	87
88	89	89	89	89	89	91	91	93	93
93	93	93	95	95	95	96	97	99	99

ii. En küçük sayı 22 *ve en büyük sayı* 99 *olduğu için değişim aralığı* $99-22=77$ *dir.*

iii. 50 *den az denildiği için* 50 *dahil değildir. Bu sebeble* 50 *den az puan alan öğrenci sayısı* 18 *dir.*

iv. 70 *veya daha fazla puan denildiği için* 70 *dahil edilir. Dolayısıyla* 70 *veya daha fazla puan alan öğrencilerin sayısı* 52 *dir.*

v. Önce en küçük puanı içine alan ve büyüklüğü 10 *olan birinci sınıf oluşturulur. Daha sonra diğer sınıflar oluşturulur. Buna göre ilk sınıf* $20-29$ *olursa diğer sınıflar şöyle olacaktır:* $30-39$, $40-49$, $50-59$, $60-69$, $70-79$, $80-89$, $90-99$. *Bu sınıfların içine düşen puan sayıları tesbit edilerek dizinin frekansları bulunur. Böylece, veriler aşağıdaki gibi gruplandırılır:*

Sınıf aralığı	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Frekans	1	6	11	12	18	16	22	14

vi. Bir sınıfın alt sınırının bulunması için o sınıfın alt uç değeri ile bir önceki sınıfın üst uç değeri toplanır ve ikiye bölünür. Örneğin $40-49$ *sınıfının alt sınırı* $(40+39)/2=39.5$ *dur. Alt sınıra sınıf aralığının büyüklüğü eklenerek üst sınır bulunur. Dolayısıyla* $40-49$ *sınıfının üst sınırı* $39.5+10=49.5$ *olur. Sınıf değerini veya sınıf orta noktasını bulmak için ise alt ve üst sınıf uçları toplanır ve ikiye bölünür.*

Buna göre, yukarıdaki dizinin her bir sınıfı için sınıf sınırları ve sınıf değerleri aşağıdaki gibi olur.

Sınıflar	*Alt sınır*	*Üst sınır*	*Sınıf değeri*
20-29	*19.5*	*29.5*	*24.5*
30-39	*29.5*	*39.5*	*34.5*
40-49	*39.5*	*49.5*	*44.5*
50-59	*49.5*	*59.5*	*54.5*
60-69	*59.5*	*69.5*	*64.5*
70-79	*69.5*	*79.5*	*74.5*
80-89	*79.5*	*89.5*	*84.5*
90-99	*89.5*	*99.5*	*94.5*

vii. Yukarıdaki gruplandırılmış verilerin frekans histogramı ve frekans poligonu aşağıdaki şekilde gösterilmiştir.

Şekil 6.1:

Dairesel Grafik: Bu grafiğin nasıl çizileceğini kısaca aşağıdaki örnekle açıklayacağız.

Örnek 6.4 *Dünyanın kara kısmının alanı* 133.3 *milyon km karedir. Dünyayı oluşturan büyük kara parçalarının alanı ise:*

Kara Parçaları	Afrika	Asya	K.Amerika	S.S.C.B	G. Amerika	Okyanusya	Avrupa
Alanları (Milyon km^2)	30.3	26.9	24.3	20.5	17.9	8.5	4.9

dir. Bu verilerin dairesel grafiğini çiziniz.

Çözüm 6.4 *Önce kara parçalarının toplam yüzölçümü dairenin merkez açısının ölçüsü olan* 360 *dereceye bölünür,* $(360/133.3) = 2.7$, *daha sonra bulunan* 2.7 *herbir kara parçasının büyüklüğü ile çarpılarak, dairesel grafikte kara parçalarına karşılık gelen açısal bölge bulunur. Bununla ilgili grafik aşağıdadır.*

Kara Parçaları	Afrika	Asya	K.Amerika	S.S.C.B	G. Amerika	Okyanusya	Avrupa
Alanları (Milyon km^2)	30.3	26.9	24.3	20.5	17.9	8.5	4.9
Açılar(derece)	82^o	73^o	66^o	55^o	48^o	23^o	13^o

.

Şekil 6.2:

6.4 Yer Ölçüleri (Ortalamalar)

Ortalamalar, istatistiğin özetleme görevini en ileri seviyede gören istatistik ölçülerdir. Örneğin, 1000 birimlik bir veri, 80 sınıf veya 15 grup halinde özetlenebildiği gibi verinin ortalaması alınmak suretiyle bu 1000 birimlik veri bir tek birimle temsil edilebilmektedir. Ortalamalar başlıca iki gruba ayrılır.

i. Verilerin bütün birimlerini içine alan ortalamalar: aritmetik ortalamalar, geometrik ortamalar, harmonik ortalamalar ve kareli ortalamalardır.

ii. Verilerin bütün birimlerini içine almayan ortalamalar: medyan ve mod.

6.4.1 Aritmetik Ortalama

Tanım 6.8 *i. Bir gözlemde elde edilen $X_1, X_2, ..., X_N$ basit verilerinin aritmetik ortalaması $\bar{X}$ ile gösterilir ve*

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{\sum X_i}{N}, \quad i = 1, 2, ..., N \tag{6.1}$$

formülüyle hesaplanır.

ii. N tane veriden $X_1, X_2, \ldots, X_r$ verilerinin frekansları $f_1, f_2, \ldots, f_r$ (burada $f_1 + f_2 + \cdots + f_r = N$) olmak üzere sınıflandırılmış veriler için *aritmetik ortalama*

$$\bar{X} = \frac{\sum f_i X_i}{\sum f_i}, \quad i = 1, 2, ..., r \tag{6.2}$$

formülüyle hesaplanır.

iii. Gruplandırılmış verilerde ise aritmetik ortalama, $Y_1, Y_2, \ldots, Y_r$ *sınıf ortası ve* $f_1, f_2, \ldots, f_r$ *sınıf aralığı frekansı olmak üzere*

$$\bar{X} = \frac{\sum f_i Y_i}{\sum f_i}, \quad i = 1, 2, ..., r \tag{6.3}$$

formülüyle hesaplanır.

Örnek 6.5 63, 65, 67, 68, 69, 70, 71, 72, 74, 75 *dizisinin aritmetik ortalamasını bulunuz.*

Çözüm 6.5 *Bu dizide* $N = 10$ *dur. Buna göre,* $\sum X = 63 + 65 + 67 + 68 + 69 + 70 + 71 + 72 + 74 + 75 = 694$ *ve böylece, aritmetik ortalama:*

$$\bar{X} = \frac{\sum X}{N} = \frac{694}{10} = 69.4$$

olur.

Örnek 6.6 30 *kişilik bir orkestrada, herbir kişinin kaç tane enstrüman çaldığı aşağıdaki tabloda verilmiştir. Enstrüman çalanların aritmetik ortalamalarını bulunuz.*

Ensrümanların sayısı	*1*	*2*	*3*	*4*	*5*
Frekans	*11*	*10*	*5*	*3*	*1*

Çözüm 6.6 *Veriler sınıflandırılmış olarak verilmiştir. Buna göre,*

X	*f*	*fX*
1	*11*	*11*
2	*10*	*20*
3	*5*	*15*
4	*3*	*12*
5	*1*	*5*
	$\sum f = 30$	$\sum fX = 63$

tablosundan yararlanarak aritmetik ortalama,

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{63}{30} = 2.1$$

olarak bulunur.

Örnek 6.7 *Aşağıda verilen gruplandırılmış veriler için aritmetik ortamayı bulunuz.*

Sınıf aralığı	*2-4*	*5-7*	*8-10*	*11-13*
f	*2*	*13*	*4*	*1*

Çözüm 6.7 *Verilen dizi gruplandırılmış olduğundan,*

Sınıf aralığı	*Y*	*f*	*fY*
2-4	*3*	*2*	*6*
5-7	*6*	*13*	*78*
8-10	*9*	*4*	*36*
11-13	*12*	*1*	*12*
		$\sum f = 20$	$\sum fY = 132$

tablosu yapılabilir (Y nin sınıf ortası olduğuna dikkat ediniz). Buradan istenen aritmetik ortalama,

$$\bar{X} = \frac{\sum fY}{\sum f} = \frac{132}{20} = 6.6$$

olur.

Aritmetik Ortalamanın Özellikleri

i. Verideki birimlerin aritmetik ortalamadan sapmalarının toplamı sıfırdır. Yani,

$$\sum(X - \bar{X}) = 0$$

dır.

ii. Dizideki birimlerin aritmetik ortalamadan sapmalarının kareleri toplamı minimumdur. Yani, A verideki herhangi bir değer ise

$$\sum(X-\bar{X})^2 \leq \sum(X-A)^2$$

olur.

iii. Aritmetik ortalama uç (anormal) değerlerin fazlasıyla etkisinde kalır. Bu da araştırmacıyı yanlış yargıya götürebilir.

iv. Gruplandırılmış bir veride açık sınıf bulunduğu takdirde aritmetik ortalama hesaplanamaz.

6.4.2 Geometrik Ortalama(G)

Geometrik ortalama da aritmetik ortalama gibi verinin bütün birimlerini içine alan bir ortalama çeşididir.

Tanım 6.9 *i. Bir gözlemde elde edilen* $X_1, X_2, ..., X_N$ *verilerinin geometrik ortalaması*

$$G = \sqrt[N]{X_1 X_2 ... X_N}$$

formülüyle hesaplanır.

ii. Sınıflandırılmış verilerde ise, $X_1, X_2, ..., X_r$ *sınıf aralıklarının sınıf ortaları ve* $f_1, f_2, \ldots, f_r$ *aralıkların frekansları (burada* $f_1+f_2+\cdots+f_r = N$*) olmak üzere*

$$G = \sqrt[N]{X_1^{f_1} X_2^{f_2} ... X_r^{f_r}} = \frac{1}{\sum f}(f_1 log X_1 + f_2 log X_2 + \cdots + f_r log X_r)$$

formülüyle hesaplanır.

iii. Gruplandırılmış verilerde ise, $Y_1, Y_2, ..., Y_r$ *sınıf aralıklarının sınıf ortaları ve* $f_1, f_2, \ldots, f_r$ *aralıkların frekansları (burada* $f_1+f_2+\cdots+f_r = N$ *dir) olmak üzere*

$$G = \sqrt[N]{Y_1^{f_1} Y_2^{f_2} ... Y_r^{f_r}} = \frac{1}{\sum f}(f_1 log Y_1 + f_2 log Y_2 + \cdots + f_r log Y_r)$$

formülüyle hesaplanır.

Bu ortalama, geometrik diziler için en ideal ortalama çeşididir. Ancak, veride sıfır veya negatif değer bulunduğu zaman geometrik ortalama anlamsız çıkar.

Örnek 6.8 *Aşağıdaki verilerin geometrik ortalamasını bulunuz.*

i. 3, 9, 27 *ii.* 2, 4, 8, 16, 32 *iii.* 1, 4, 9, 17, 30 *iv.* 4.2, 16.8

Çözüm 6.8

$$
\begin{array}{llll}
i. & G & = & \sqrt[3]{X_1X_2X_3} \qquad = \sqrt[3]{(3)(9)(27)} = 9 \\
ii. & G & = & \sqrt[5]{(2)(4)(8)(16)(32)} = 8 \\
iii. & G & = & \sqrt[5]{(1)(4)(9)(17)(30)} = 11.3 \\
iv. & G & = & \sqrt{(4.2)(16.8)} = 8
\end{array}
$$

Örnek 6.9 *Aşağıdaki verinin geometrik ortalamasını bulunuz.*

X	*1*	*4*	*10*	*20*
f	*2*	*3*	*2*	*1*

Çözüm 6.9 *Önce aşağıdaki tabloyu yapalım.*

X	*1*	*4*	*10*	*20*	
f	*2*	*3*	*2*	*1*	$\sum f = 8$
X^f	*1*	*64*	*100*	*20*	

Tablo yardımıyla geometrik ortalama,

$$
\begin{aligned}
G &= \sqrt[\sum f]{X_1^{f_1} X_2^{f_2} \cdots X_n^{f_n}} \\
G &= \sqrt[8]{(1)(64)(100)(20)} \\
G &= \sqrt[8]{128000}
\end{aligned}
$$

olarak bulunur.

Örnek 6.10 *Aşağıdaki frekans dağılımının geometrik ortalamasını bulunuz.*

Sınıf aralığı	*1-5*	*6-14*	*15-25*	*26-40*
f	*2*	*3*	*4*	*3*

Çözüm 6.10

Sınıf aralığı	*1-5*	*6-14*	*15-25*	*26-40*	
f	*2*	*3*	*4*	*3*	$\sum f = 12$
Y	*3*	*10*	*20*	*33*	

Bu tabloya göre

$$\begin{aligned} G &= \sqrt[\sum f]{Y_1^{f_1} Y_2^{f_2} \cdots Y_r^{f_r}} \\ &= \sqrt[12]{(3^2)(10^3)(20^4)(33^3)} \\ &= 11.5 \end{aligned}$$

bulunur.

6.4.3 Harmonik Ortalama(H)

Tanım 6.10 *i. Bir gözlemde elde edilen $X_1, X_2, ..., X_N$ verilerinin harmonik ortalaması*

$$H = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_N}} = \frac{N}{\sum \frac{1}{X_i}}$$

formülüyle hesaplanır.

ii. Sınıflandırılmış verilerde, $X_1, X_2, ..., X_r$ sınıf aralıklarının sınıf ortaları ve $f_1, f_2, \ldots, f_r$ aralıkların frekansları (burada $f_1 + f_2 + \cdots + f_r = N$ dir) olmak üzere

$$H = \frac{N}{\frac{f_1}{X_1} + \frac{f_2}{X_2} + \cdots + \frac{f_r}{X_r}} = \frac{N}{\sum \frac{f_i}{X_i}}$$

formülüyle hesaplanır.

iii. Gruplandırılmış verilerde ise, $Y_1, Y_2, ..., Y_r$ *sınıf aralıklarının sınıf ortaları ve* $f_1, f_2, \ldots, f_r$ *aralıkların frekansları (burada* $f_1+f_2+\cdots+f_r = N$ *dir) olmak üzere*

$$H = \frac{N}{\frac{f_1}{Y_1} + \frac{f_2}{Y_2} + \cdots + \frac{f_r}{Y_r}} = \frac{N}{\sum \frac{f_i}{Y_i}}$$

formülüyle hesaplanır.

Harmonik ortalama genellikle hız ortalamasının hesabında kullanılır.

Örnek 6.11 $2, 5, 10$ *verilerinin harmonik ortalamasını hesaplayınız.*

Çözüm 6.11 *Veriler bir dizi olarak verildiğinden,*

$$H = \left[\tfrac{1}{3}\left(\tfrac{1}{2} + \tfrac{1}{5} + \tfrac{1}{10}\right)\right]^{-1} = \left(\frac{8}{30}\right)^{-1} = \frac{30}{8} = 3.75$$

bulunur.

Örnek 6.12 *Aşağıdaki frekans dağılımının harmonik ortalamasını bulunuz.*

X	*10*	*15*	*20*	*30*
f	*1*	*3*	*5*	*2*

Çözüm 6.12 *Veriler sınıflandırılmış olarak verildiğinden*

$$H = \sum f \left[\frac{f_1}{X_1} + \frac{f_2}{X_2} + \cdots + \frac{f_n}{X_n}\right]^{-1}$$

formülünden,

$$H = 11 \left[\frac{1}{10} + \frac{3}{15} + \frac{5}{20} + \frac{2}{30}\right]^{-1} = 17.86$$

bulunur.

Örnek 6.13 *Aşağıdaki dağılımın harmonik ortalamasını bulunuz.*

Sınıf aralığı	*10-20*	*20-30*	*30-40*	*40-50*
f	*2*	*4*	*4*	*6*

Çözüm 6.13 *Veriler gruplandırılmış olarak verilmiştir. Yeni bir tablo yaparak sınıf ortalarını belirtelim:*

Sınıf aralığı	*10-20*	*20-30*	*30-40*	*40-50*	
f	*2*	*4*	*4*	*6*	$\sum f = 16$
Y	*15*	*25*	*35*	*45*	

Buradan,

$$H = 16\left[\frac{2}{15} + \frac{4}{25} + \frac{4}{35} + \frac{6}{45}\right]^{-1} = 29.63$$

bulunur.

Örnek 6.14 *A, B ve C şehirleri birbirinden eşit uzaklıktadır. A dan B ye* 30 *km/saat, B den C ye* 40 *km/saat ve C den A ya* 50 *km/saat süratle giden bir sürücünün gezinin tamamı için ortalama hızını bulunuz.*

Çözüm 6.14 *Hız ortalamalarını bulmak için harmonik ortalama kullanılır. Buna göre*

$$H = 3\left[\frac{1}{30} + \frac{1}{40} + \frac{1}{50}\right]^{-1} = 38.46\, km/saat$$

bulunur.

Örnek 6.15 *Bir uçak 2500, 1200, 500 millik uzaklıkları sırasıyla, 500, 400, 250 mil/saat süratle uçmuştur. Bu uçağın ortalama hızını bulunuz.*

Çözüm 6.15 *Hız problemidir:*

$$H = (2500 + 1200 + 500)\left[\frac{2500}{500} + \frac{1200}{400} + \frac{500}{250}\right]^{-1} = 420 \ mil/saat$$

olur.

Örnek 6.16 *Bir işyerinde çalışan 3 işçiden biri bir parçayı 10, ikincisi 15 diğeri ise 30 dakikada yapmaktadır. Bu işçiler bir iş gününde bir parçayı ortalama ne kadar zamanda yaparlar.*

Çözüm 6.16 *Bu tip problemlerin çözümü için harmonik ortalama kullanılır.*

$$H = 3\left[\frac{1}{10} + \frac{1}{15} + \frac{1}{30}\right]^{-1} = 15 \ dakika$$

bulunur.

6.4.4 Kareli Ortalama (K)

Tanım 6.11 *Kareli ortalama, verideki birimlerin karelerinin aritmetik ortalamasının karekökküdür. Basit verilerde kareli ortalama,*

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

formülüyle, sınıflandırılmış dizilerde,

$$K = \sqrt{\frac{\sum fX^2}{\sum f}}$$

ve gruplandırılmış verilerde ise Y sınıf ortaları olmak üzere,

$$K = \sqrt{\frac{\sum fY^2}{\sum f}}$$

formülüyle hesaplanır.

Örnek 6.17 *Aşağıdaki sayıların kareli ortalamalarını bulunuz.*

i. $11, 23, 35$ *ii.* $-3, -1, 0, 4, 6$

Çözüm 6.17 *Basit veriler olduklarından,*

i.

$$K = \sqrt{\frac{\sum X^2}{n}} = \sqrt{\frac{11^2 + 23^2 + 35^2}{3}} = 25$$

ii.

$$K = \sqrt{\frac{(-3)^2 + (-1)^2 + (0)^2 + (4)^2 + (6)^2}{5}} = \sqrt{\frac{6 + 1 + 16 + 36}{5}} = 3.52$$

bulunur.

Örnek 6.18 *Aşağıdaki dağılımın kareli ortalamasını hesaplayınız.*

X	*7*	*10*	*12*	*13*	*15*
f	*3*	*3*	*5*	*2*	*2*

Çözüm 6.18 *Sınıflandırılmış veri olduğundan*

X	*7*	*10*	*12*	*13*	*15*	
f	*3*	*3*	*5*	*2*	*2*	$\sum f = 15$
X^2	*49*	*100*	*144*	*169*	*225*	
fX^2	*147*	*300*	*720*	*338*	*450*	$\sum fx^2 = 1995$

tablosu yapılır. Bu tablodan yararlanarak

$$K = \sqrt{\frac{\sum fX^2}{\sum f}} = \sqrt{\frac{1955}{15}} = 11.42$$

bulunur.

Örnek 6.19 *Aşağıdaki dağılımın kareli ortalamasını bulunuz.*

Sınıf aralığı	$-30-(-20)$	$-20-(-10)$	$-10-0$	0-10	10-20	20-30
f	55	50	45	30	80	100

Çözüm 6.19 *Gruplandırılmış veriler olduğundan sınıf ortalarını bulmak gerekir.*

Sınıf aralığı	-30-(-20)	-20-(-10)	-10-0	0-10	10-20	20-30	
f	55	50	45	30	80	100	$\sum f = 360$
Y	-25	-15	-5	5	15	25	
Y^2	625	225	25	25	225	625	
fY^2	34375	11250	1125	750	18000	62500	$\sum fY^2 = 128000$

Bu tablodan yararlanarak

$$K = \sqrt{\frac{\sum fY^2}{\sum f}} = \sqrt{\frac{128000}{360}} = 18.86$$

bulunur.

6.4.5 Medyan (Ortanca)

Medyan kelime manası itibariyle "orta değer" demektir.

Tanım 6.12 *Bir dizinin elemanlarını küçükten büyüğe veya büyükten küçüğe sıraladığımız takdirde tam ortaya düşen değere (veya ortaya düşen iki değerin ortalamasına) medyan denir.*

Basit ve sınıflandırılmış verilerde medyanı bulmak için şu yol takip edilir: $X_1, X_2, \ldots, X_N$ verileri küçükten büyüğe (veya büyükten küçüğe) doğru sıralandığında;

i. Eğer N tek ise

$$Medyan = X_{\frac{N+1}{2}}$$

olur. Yani medyan, $\frac{N+1}{2}$. terimdir.

ii. Eğer N çift ise

$$Medyan = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2}$$

olur. Yani medyan $\frac{N}{2}$ ve $\frac{N}{2}+1$. terimlerin aritmetik ortalamasıdır.

Gruplandırılmış verilerde ise medyan aşağıda verilen eşitlikle, enterpolasyon yoluyla elde edilir:

$$\text{Medyan} = L_1 + \frac{\frac{N}{2} - \sum f_i}{f_m} c \qquad (6.4)$$

Bu formülde,

L_1: Medyan sınıfının alt sınırını,

N: Toplam frekansı,

$\sum f_i$: Medyan sınıf aralığından önceki sınıf aralıklarının frekansları toplamını,

f_m: Medyan sınıfının frekansını,

c: Medyan sınıfının büyüklüğünü

göstermektedir.

Örnek 6.20 *Aşağıdaki verilerin medyanını bulunuz.*

i. $5, 4, 6, 4, 4, 3, 7, 5, 8$ *ii.* $2/3, 1/3, 5/6, 3/4, 1/2, 1/3, 1/4$
iii. $1, 1, 1, 2, 2, 2, 3, 3, 5, 5, 8, 8$

Çözüm 6.20 *i. Önce verilen veriler*

$$3, 4, 4, 4, 5, 5, 6, 7, 8$$

şeklinde büyüklük sırasına göre yazılır. Buradan

$$\underline{3, 4, 4, 4}, 5, \underline{5, 6, 7, 8}$$

yani, ortadaki terim 5 *olduğundan* $Medyan = 5$ *bulunur.*

Diğer bir düşünüşle, terim sayısı $N = 9$ *(tek) olduğundan* $\frac{N+1}{2} = \frac{9+1}{2} =$ $5.$ *terim olan 5 medyandır denir.*

ii. Büyüklük sırasına koymak için önce paydalar eşitlenir ve küçükten büyüğe doğru aşağıdaki gibi sıralanır.

$$\underline{3/12, 4/12, 4/12}, 6/12, \underline{8/12, 9/12, 10/12}$$

Ortadaki terim $6/12$ *olduğundan* $Medyan = 6/12 = 0.5$ *dir.*

iii. Terim sayısı $N = 12$ *yani çifttir.* $\frac{N}{2} = \frac{12}{2} = 6$ *olduğundan 6. ve 7.terimlerin aritmetik ortalaması medyandır. 6.terim 2 ve 7.terim 3 olduğundan,*

$$Medyan = \frac{2+3}{2} = 2.5$$

olur.

Örnek 6.21 *Aşağıdaki dağılımın medyanını bulunuz.*

X	46	48	49	51	53	55	57	60
f	9	12	11	6	8	10	12	2

Çözüm 6.21 *Önce aşağıdaki tabloyu yapalım,*

X	46	48	49	51	53	55	57	60	
f	9	12	11	6	8	10	12	2	$\sum f = 70$
Kümülatif f	9	21	32	38	46	56	68	70	

$N = 70$ *çift olduğundan medyan,* $N/2$. *ve* $(N/2)+1$. *terimlerin aritmetik ortalamasıdır. Yani, 35. ve 36. terimler olan 51 ve 51 sayılarının aritmetik ortalaması medyandır. Buna göre medyan*=51 *dir.*

Örnek 6.22 *Aşağıdaki tablo* 1972 *yılında Türkiyedeki nüfusun yaş gruplarına göre dağılımını göstermektedir. Medyanı bulunuz.*

Yaşlar	0-4	5-9	10-19	20-29	30-39	40-49	50-59	60-64	65+
Nüfus(100 bin)	43	40	55	46	35	22	19	8	10

Çözüm 6.22

Yaşlar	0-4	5-9	10-19	20-29	30-39	40-49	50-59	60-64	65+	
Nüfus(100 bin)	43	40	55	46	35	22	19	8	10	$N = 278$
Kümülatif f	43	83	138	184	219	241	260	268	278	

$N/2 = 278/2 = 139$ *olduğundan 139. terim ile 140. terimin aritmetik ortalaması medyandır. Bu terimler* $(20-29)$ *sınıfında olduğu için medyan sınıfı* $(20-29)$ *dur.*

$$Medyan \;=\; L_1 + \frac{\frac{N}{2} - \sum f_i}{f_m} c$$

eşitliğinde

$$L_1 = (20+29)/2 = 19.5,\ \sum f_i = 138,\ f_m = 46,\ c = 10$$

olduğundan,

$$Medyan = 19.5 + \frac{\frac{278}{2} - 138}{46}.10 = 19.5 + \frac{10}{46} = 19.72 \approx 20$$

bulunur.

Bazı durumlarda medyan, aritmetik ortalamadan daha tutarlıdır. Örneğin, aylık gelirleri 3 000 000, 3 000 000, 2 000 000, 4 000 000, 3 000 000, 5 000 000, 8 000 000, 10 000 000, 6 000 000, 100 000 000 olan bir topluluğun medyanı 4 500 000; aritmetik ortalaması ise 14 400 000 dir. Görüldüğü gibi, medyan gerçek durumu aritmetik ortalamaya göre daha iyi yansıtmaktadır. Buradan anlaşılacağı gibi, medyan, aritmetik ortalama gibi uç (anormal) değerlerden etkilenmez. Ancak veri sayısından etkilenir.

6.4.6 Mod

Tanım 6.13 *Bir dizide en çok tekrar eden, yani frekansı en yüksek olan değere* **mod** *denir.*

Bir sayı dizisinde elemanların hepsi birbirinden farklı ise bu dizi için mod aranmaz. Örneğin, $5, 7, 12, 13, 28, 40, 35$ sayı dizisinin modu yoktur.

Tanımdan da anlaşıldığı gibi sınıflandırılmış dizide mod en yüksek frekansa karşılık gelen X değeridir. Gruplandırılmış dizilerde ise mod, aşağıdaki eşitlikle hesaplanır:

$$\text{Mod} = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} c \tag{6.5}$$

Bu formülde,

L_1: Mod sınıfının alt sınırını,

Δ_1: Mod sınıfı frekansının bir önceki sınıf frekanslarından fazlalığını,

Δ_2: Mod sınıfı frekansının bir sonraki sınıf frekanslarından fazlalığını,

c: Mod sınıfının büyüklüğünü göstermektedir.

Örnek 6.23 *Aşağıdaki sayı dizilerinin her birinin modunu bulunuz.*

i. $14, 15, 16, 16, 18, 16, 16, 15, 17, 15, 14, 17, 13$

ii. $32, 36, 39, 33, 33, 36, 31, 36, 39, 37, 36, 39, 39$

iii. $57, 41, 65, 50, 58, 34, 42, 38, 46, 59, 33$

Çözüm 6.23 *i.* $14, 15, 16, 16, 18, 16, 16, 15, 17, 15, 14, 17, 13$ *dizisinde en fazla tekraralanan terim* 16 *olduğu için mod* $= 16$ *dır.*

ii. $32, 36, 39, 39, 33, 36, 31, 36, 39, 37, 36, 39, 39$ *dizisinde* 39 *ve* 36 *sayılarının her ikiside* 4 *kere tekrarlanmıştır. Dolayısıyla en çok tekrarlanan terim iki tanedir. Bu sebeble bu dizinin iki modu vardır.*

iii. $57, 41, 65, 50, 58, 34, 42, 38, 46, 59, 33,$ *dizisinin modu yoktur.*

Örnek 6.24 1977 *yılında Türkiye'de cezaevlerine giren hükümlülerin yaş gruplarına göre dağılımı aşağıdaki gibidir. Mod nedir?*

Yaş grupları	12-15	16-18	19-21	22-29	30-39	40-49	50-59	60-64	65+
f	500	1099	3133	12992	9365	6249	2122	380	432

Çözüm 6.24

Sınıf aralığı	12-15	16-18	19-21	22-29	30-39	40-49	50-59	60-64	65+
f	500	1099	3133	12992	9365	6249	2122	380	432

$22 - 29$ *sınıfı mod sınıfıdır. Yani, Mod* 22 *ile* 29 *arasında olacaktır.*

$$\begin{aligned} L_1 &= 21.5 \\ \Delta_1 &= 12992 - 3133 = 9859 \\ \Delta_2 &= 12992 - 9365 = 3627 \\ c &= 29 - 21 = 8 \end{aligned}$$

olduğundan,

$$Mod = 21.5 + \frac{9859}{9859 + 3627}(8) = 27.35$$

olarak bulunur. Yani, Mod 27 *dir.*

Örnek 6.25 *Aşağıdaki frekans dağılımının modunu bulunuz.*

Sınıf aralığı	150-154	155-159	160-164	165-169	170-174	175-179	180-184	185-189
f	3	8	12	9	12	5	2	1

Çözüm 6.25 *Bu dağılımda en büyük frekans iki ayrı sınıfa karşılık geldiği için mod sınıfın hangisi olacağı bilinmez. Bu gibi durumlarda yeniden gruplandırmaya gidilir.*

Sınıf aralığı	*150-159*	*160-169*	*170-179*	*180-189*
f	*11*	*21*	*17*	*3*

Bu yeni dağılımda mod sınıfı $160 - 169$ *sınıfı olacağından,*

$$Mod = 159.5 + \frac{10}{10 + 4}(10) = 166.64$$

olur.

6.4.7 Ortalama, Medyan ve Mod Arasındaki İlişki

Tanım 6.14 *Bir dizinin dağılım frekans eğrisi merkezsel maksimumun sağında daha uzun uzantıya sahipse dağılım için sağa yatık(çarpık), aksi halde sola yatık(çarpık) denir.*

Dizinin durumuna göre, otalama, medyan ve mod arasında:

i. Dizi simetrik ise, $\bar{X} = mod = medyan$,

ii. Dizi sağa çarpık ise, $mod < medyan < \bar{X}$,

iii. Dizi sola çarpık ise, $\bar{X} < medyan < mod$

ilişkileri vardır.

Bu üç yer ölçüsü arasında

$$\bar{X} - mod = 3(\bar{X} - medyan) \tag{6.6}$$

bağıntısı vardır.

Bir eğride mod sınıfından önceki sınıfların frekansları toplamı sonraki sınıfların frekansları toplamından büyükse dizinin sola çarpık olduğu, aksi halde sağa çarpık olduğu söylenebilir.

Şekil 6.3:

6.4.8 Kartiller, Desiller, Pörsentiller

Kartiller, büyüklük sırasına konulmuş bir diziyi dört eşit kısma bölen değerlerdir. Bir dizide üç kartil mevcuttur ve ikinci kartil medyandır.Buna göre n elemanlı bir dizide,

$$\begin{array}{llcl} \text{Birinci kartil} & Q_1 & = & \frac{1}{4}(n+1).\ \text{ değer} \\ \text{Medyan} & Q_2 & = & \frac{1}{2}(n+1).\ \text{ değer} \\ \text{Üçüncü kartil} & Q_3 & = & \frac{3}{4}(n+1).\ \text{ değer} \end{array} \tag{6.7}$$

ve sınıflandırılmış dizide,

$$\begin{array}{llcl} \text{Birinci kartil} & Q_1 & = & \frac{N}{4}+\frac{1}{2}.\ \text{ değer} \\ \text{Medyan} & Q_2 & = & \frac{1}{2}(n+1).\ \text{ değer} \\ \text{Üçüncü kartil} & Q_3 & = & \frac{3N}{4}.\ \text{ değer} \end{array} \tag{6.8}$$

olur.

Gruplandırılmış dizilerde ise,

$$\begin{array}{llcl} \text{Birinci kartil} & Q_1 & = & L_1 + \frac{\frac{N}{4}-\sum f_i}{f_{Q_1}}c.\ \text{ değer} \\ \text{İkinci kartil} & Q_2 & = & L_1 + \frac{\frac{2N}{4}-\sum f_i}{f_{Q_2}}c.\ \text{ değer} \\ \text{Üçüncü kartil} & Q_3 & = & L_1 + \frac{\frac{3N}{4}-\sum f_i}{f_{Q_3}}c.\ \text{ değer} \end{array} \tag{6.9}$$

şeklinde tanımlanır.

Desiller, büyüklük sırasına konulmuş bir diziyi 10 eşit kısma bölen değerlerdir. Beşinci desil medyandır.

Sınıflandırılmış dizilerde,

$$\begin{array}{llcl} \text{Birinci desil} & D_1 & = & \frac{N}{10}+\frac{1}{2}.\ \text{ değer} \\ \text{İkinci desil} & D_2 & = & \frac{2N}{10}+\frac{1}{2}.\ \text{ değer} \\ \cdots & \cdots & & \cdots \\ \text{Dokuzuncu desil} & D_9 & = & \frac{9N}{10}+\frac{1}{2}.\ \text{ değer} \end{array} \tag{6.10}$$

şeklindedir.

Gruplandırılmış dizilerde ise,

$$\begin{array}{llcll} \text{Birinci desil} & D_1 & = & L_1 + \frac{\frac{N}{10} - \sum f_i}{f_{D_1}} c. & \text{değer} \\ \text{İkinci desil} & D_2 & = & L_1 + \frac{\frac{2N}{10} - \sum f_i}{f_{D_2}} c. & \text{değer} \\ \dots & \dots & & \dots & \\ \text{Dokuzuncu desil} & D_9 & = & L_1 + \frac{\frac{9N}{10} - \sum f_i}{f_{D_9}} c. & \text{değer} \end{array} \tag{6.11}$$

olarak tanımlanır.

Pörsentiller, büyüklük sırasına konulmuş bir diziyi 100 eşit kısma bölen değerlerdir. Ellinci pörsentil medyandır.

Sınıflandırılmış dizilerde,

$$\begin{array}{llcll} \text{Birinci pörsentil} & P_1 & = & \frac{N}{100} + \frac{1}{2}. & \text{değer,} \\ \text{İkinci pörsentil} & P_2 & = & \frac{2N}{100} + \frac{1}{2}. & \text{değer,} \\ \dots & \dots & & \dots & \\ \text{Doksan dokuzuncu pörsentil} & P_{99} & = & \frac{99N}{100} + \frac{1}{2}. & \text{değer.} \end{array} \tag{6.12}$$

Gruplandırılmış dizilerde ise,

$$\begin{array}{llcll} \text{Birinci pörsentil} & P_1 & = & L_1 + \frac{\frac{N}{100} - \sum f_i}{f_{P_1}} c. & \text{değer} \\ \text{İkinci pörsentil} & P_2 & = & L_1 + \frac{\frac{2N}{100} - \sum f_i}{f_{P_2}} c. & \text{değer} \\ \dots & \dots & & \dots & \\ \text{Doksan dokuzuncu pörsentil} & P_{99} & = & L_1 + \frac{\frac{99N}{100} - \sum f_i}{f_{P_{99}}} c. & \text{değer} \end{array} \tag{6.13}$$

olarak tanımlanır.

Örnek 6.26 1970 *yılı sayımına göre Türkiye'de nüfusu* 5000 *den fazla olan yerlerdeki işletmelerin işletme büyüklüğüne göre dağılımı aşağıdaki gibidir.*

İş.b.	0.5den az	0.5-0.9	1-1.9	2-2.9	3-3.9	4-4.9	5-9.9	10-19.9	20-49.9	50-99.9
İ s.	18	17	21	12	7	6	12	7	5	1

i. Birinci, ikinci ve üçüncü kartilleri bulunuz.

ii. Üçüncü, beşinci ve yedinci desilleri bulunuz.

iii. Kırkbeşinci pörsentili bulunuz.

Çözüm 6.26 *i. Birinci kartil (Q_1), $N/4$; İkinci kartil(Q_2 = medyan), $2N/4$; Üçüncü kartil (Q_3), $3N/4$ dir. $N/4 = 106/4 = 26.5$ ve birinci kartil sınıfı* $(0.5 - 0.9)$ *dur. Buna göre,*

$$Q_1 = L_1 + \frac{N/4 - \sum f_i}{f_{Q_1}} c$$

formülünde,

$$L_1 = \frac{0.5 + 0.4}{2} = 0.45, \ \sum f_i = 18, \ f_{Q_1} = 17, \ c = 0.9 - 0.4 = 0.5$$

dir. Buna göre birinci kartil,

$$Q_1 = 0.45 + \frac{26.5 - 18}{17} 0.5 = 0.70$$

olarak elde edilir. Bunun demektir ki, işletme büyüklüğü 0.7 *hektar ve daha küçük işletmelerin yüzdesi* %25 *tir.*

Gruplar	0.5den az	0.5-0.9	1-1.9	2-2.9	3-3.9	4-4.9	5-9.9	10-19.9	20-49.9	50-99.9
f	18	17	21	12	7	6	12	7	5	1
Küm. f	18	35	56	68	75	81	93	100	106	106

$N = \sum f = 106$ *İkinci kartil Medyandır.* $N/2 = 106/2 = 53$ *ve medyan sınıfı* $(1 - 1.9)$ *dur.*

$$Medyan = L_1 + \frac{N/2 - \sum f_i}{f_m} c$$

formülünde,

$$L_1 = (1 + 0.9)/2 = 0.95, \ \sum f_i = 35, \ f_m = 21, \ c = 1$$

olduğundan,

$$Q_2 = medyan = 0.95 + \frac{53 - 35}{21}(1) = 1.81$$

olarak elde edilir. Bunun manası, işletme büyüklüğü 1.81 *hektar ve daha küçük işletmelerin yüzdesi* %50 *dir.*

$3N/4 = (3(106))/4 = 79.5$ *ve üçüncü kartil sınıfı* $4 - 4.9$ *dur. Dolayısıyla,*

$$Q_3 = 3.95 + \frac{79.5 - 75}{12}(1) = 4.33$$

olarak elde edilir. Bunun manası, işletme büyüklüğü 4.33 *hektar ve daha küçük işletmelerin yüzdesi* %75 *dir.*

ii. Üçüncü desil (D_3), $3N/10$; Beşinci desil ($D_5 = medyan$), $5N/10$; Yedinci desil (D_7), $7N/10$ dir.

$$D_3 = L_1 + \frac{3N/10 - \sum f_i}{f_{D_3}} c$$

formülünde, $3N/10 = 3(106)/10 = 31.8$ *ve üçüncü desil sınıfı* $0.5 - 0.9$ *dur. Buna göre,*

$$D_3 = 0.45 + ((31.8 - 18)/17)0.5 = 0.86$$

bulunur. Bu da, işletme büyüklüğü 0.86 *hektar ve daha küçük işletme yüzdesinin* %30 *olduğunu belirtir.*

$$D_5 = medyan = 1.81$$

olarak bir önceki şıkta bulundu.

$7N/10 = 7(106)/10 = 74.2$ *olduğundan yedinci desil sınıfı* $3 - 3.9$ *dur. Dolayısıyla,*

$$D_7 = 2.95 + ((74.2 - 68)/7)(1) = 3.84$$

bulunur. Bu demektir ki, İşletme büyüklüğü 3.84 *hektar ve daha küçük işletme yüzdesi* %70 *dir.*

iii. Kırkbeşinci pörsentil, ($45N/100$*) dır.* $45(106)/100 = 47.7$ *ve 45. pörsentil sınıfı* $1 - 1.9$ *dur. Buna göre,*

$$P_{45} = 0.95 + ((47.7 - 35)/21)(1) = 1.55$$

olarak elde edilir. Bu da, işletmelerin %45 *inde işletme büyüklüğü* 1.55 *hektar ve daha küçük olduğuna işaret eder.*

6.5 Dağılım Ölçüleri

Uygulamada bir dizinin belirlenmesinde sadece yer ölçüleri yeterli olmamakta, dizideki birimlerin ortalama etrafındaki dağılım durumunu ortaya koymak gerekmektedir. Örneğin, ortalamaları eşit iki dizi düşünelim. Bu iki dizinin hangisinde, birimlerin ortalama etrafında daha sık, hangisinde daha seyrek biçimde dağıldığını bilmek isteriz. Çünkü, birimlerin ortalama etrafında daha sık dağıldığı dizilerde, ortalamanın diziyi temsil gücü daha yüksektir. Bu durumun tesbiti için dağılım ölçülerine ihtiyaç duyulur. Dağılım ölçülerinin en yaygın olanları **değişim aralığı, ortalama sapma, yarı kartil aralığı, standart sapma** ve **varyans**tır.

6.5.1 Değişim Aralığı

Tanım 6.15 *Değişim aralığı (D.A) bir dizideki en büyük ve en küçük değer arasındaki farktır. Yani,*

$$D.A = X_{max} - X_{min} \tag{6.14}$$

dir.

Değişim aralığının en büyün dezavantajı dizideki bütün birimlerin hesaplamaya girmeyip, sadece iki değerle sonuca ulaşılmasıdır.

Örnek 6.27 *Aşağıdaki dizilerin değişim aralığını bulunuz.*

i. $12, 6, 7, 3, 15, 10, 18, 5$

ii. $9, 3, 8, 8, 9, 8, 9, 18$

iii. 8.772, 6.453, 10.624, 8.628, 9.434, 6.351

Çözüm 6.27 *Bu verilerin değişim aralıkları:*

i. $D.A = X_{max} - X_{min} = 18 - 3 = 15,$

ii. $D.A = 18 - 3 = 15,$

iii. $D.A = 10.624 - 6.351 = 4.273$

olarak bulunur.

Örnek 6.28 *Aşağıdaki frekans dağılımının değişim aralığını bulunuz.*

X	10	12	18	21	25	32	40
f	3	7	6	10	8	4	2

Çözüm 6.28 $D.A = X_{max} - X_{min}$ *formülünden,* $D.A = 40 - 10 = 30$ *bulunur.*

Örnek 6.29 *Aşağıdaki gruplandırılmış dizinin değişim aralığını bulunuz.*

Sınıf aralığı	10-14	15-19	20-24	25-29	30-34
f	5	8	13	10	2

Çözüm 6.29 $D.A = 34 - 10 = 24$ *dür.*

Örnek 6.30 *50 ölçümden en büyüğü* 8.34 *kg ve değişim aralığı* 0.46 *kg olması halinde en küçük ölçüm kaç kg dır.*

Çözüm 6.30 *Değişim aralığı formülü kullanılarak,*

$$D.A = X_{max} - X_{min}, \quad 0.46 = 8.34 - X_{min}, \quad X_{min} = 8.34 - 0.46 = 7.88$$

bulunur.

6.5.2 Ortalama Sapma

Değişim aralığının dezevantajını önlemek amacıyla, bütün birimlerin hesaplamaya katılacağı ölçüler araştırılmış ve ilk olarak ortalama sapma formülü geliştirilmiştir. Basit dizilerde ortalama sapma (O.S)

$$O.S = \frac{\sum |X - \bar{X}|}{n} \tag{6.15}$$

eşitliği ile verilir.Sınıflandırılmış ve gruplandırılmış dizilerde ise ortalama sapmayı hesaplamak için,

$$O.S = \frac{\sum f|X - \bar{X}|}{\sum f} \tag{6.16}$$

formülü kullanılır.

Ortalama sapmanın en önemli özelliği medyan etrafındaki ortalama sapmanın minimum olmasıdır. Yani, $\sum |X - a|$ veya $\sum f|X - a|$ toplamı $a = medyan$ olduğu zaman minimum olur.

Ortalama sapmanın dezevantajı, hesaplamada mutlak değer kullanmak sebebiyle bu ölçümün matematik işlemlere elverişli olmamasıdır.

Örnek 6.31 $\{3, 7, 9, 5\}$ *kümesinin ortalama sapmasını bulunuz.*

Çözüm 6.31 $\bar{X} = \frac{\sum X}{n} = \frac{24}{4} = 6$ *olarak bulunur. Buradan,*

X	*3*	*7*	*9*	*5*	
$X - \bar{X}$	*3-6*	*7-6*	*9-6*	*5-6*	
$\|X - \bar{X}\|$	*3*	*1*	*3*	*1*	$\sum \|X - \bar{X}\| = 8$

tablosu yapılır. Bu tablo yardımıyla,

$$O.S = \frac{\sum |X - \bar{X}|}{n} = \frac{8}{4} = 2$$

bulunur.

Örnek 6.32 *Aşağıdaki frekans dağılımlarının ortalama sapmalarını bulunuz.*

i.

X	*20*	*30*	*40*	*50*	*60*	*70*	*80*
f	*3*	*5*	*10*	*11*	*7*	*3*	*1*

ii.

Sınıf aralığı	*60-62*	*63-65*	*66-68*	*69-71*	*72-74*
f	*5*	*18*	*42*	*27*	*8*

Çözüm 6.32 *i. Önce aşağıdaki tabloyu yapalım.*

X	20	30	40	50	60	70	80	
f	3	5	10	11	7	3	1	$\sum f = 40$
fX	60	150	400	550	420	210	80	$\sum fX = 1870$
$\|X - \bar{X}\|$	26.75	16.75	6.75	3.25	13.25	23.25	33.25	
$f\|X - \bar{X}\|$	80.25	83.75	67.50	35.75	92.75	69.75	33.25	$f\|X - \bar{X}\| = 463.00$

Bu tabloya göre,

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{1870}{40} = 46.75$$
$$O.S = \frac{\sum f|X - \bar{X}|}{\sum f} = \frac{463}{40} = 11.575$$

bulunur.

ii.

Sınıf aralığı	60-62	63-65	66-68	69-71	72-74	
f	5	18	42	27	8	$\sum f = 100$
X	61	64	67	70	73	
fX	305	1152	2814	1890	584	$\sum fX = 6745$
$\|X - \bar{X}\|$	6.45	3.45	0.45	2.55	5.55	
$f\|X - \bar{X}\|$	32.25	62.10	18.90	68.85	44.40	$f\|X - \bar{X}\| = 226.60$

Bu tablo yardımıyla,

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{6745}{100} = 67.45$$
$$O.S = \frac{\sum f|X - \bar{X}|}{\sum f} = \frac{226.5}{100} = 2.265$$

olarak bulunur.

6.5.3 Yarı Kartil Aralığı

Tanım 6.16 *Yarı kartil aralığı, üçüncü kartil ile birinci kartil arasındaki mutlak farkın yarısıdır, yani*

$$Q = \frac{Q_3 - Q_1}{2} \tag{6.17}$$

dir.

Bu dağılım ölçüsü, değişim aralığına göre uç değerlerden daha az etkilenir. Bununla birlikte hesaplamaya bütün birimlerin katılmaması bu dağılım ölçüsü için bir dezevantajdır.

6.5.4 Standart Sapma ve Varyans

Standart sapma, uygulamada en fazla kullanılan dağılım ölçüsüdür. Basit bir dizinin standart sapması,

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} \tag{6.18}$$

formülüyle hesaplanır. Formülde ortalamadan sapmaların karesi alınmak suretiyle mutlak değer alınmasına ihtiyaç kalmamış, ortalama sapmanın "matematik işlemlere elverişli olmama" dezevantajı bu dağılma ölçüsünde böylece ortadan kaldırılmıştır.

Sınıflandırılmış ve gruplandırılmış dizilerde standart sapma,

$$s = \sqrt{\frac{\sum f(X-\bar{X})^2}{\sum f}} = \sqrt{\frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2} \tag{6.19}$$

formülüyle hesaplanır.

Uygulamada genellikle örneklem standart sapması için s, kitle standart sapması için ise σ sembolü kullanılır. Ayrıca küçük örneklemlerde $(n < 30)$,n yerine $(n-1)$, sınıflandırılmış dizilerde $(\sum f)$ yerine $(\sum f - 1)$ kullanılır.

Varyans, standart sapmanın karesidir. Örneklemler için s^2, kitleler için σ^2 sembolü kullanılır. Standart sapma *kg, m, cm, ...* gibi ölçü birimiyle ifade edildiği halde varyans için herhangi bir ölçü birimi kullanılmaz.

6.5.5 Değişim Katsayısı

Gerek standart sapma gerekse diğer dağılım ölçüleri, ölçü biriminden bağımsız değildirler. Bundan dolayı aynı dizi farklı ölçü birimleriyle ifade edildiğinde değişik standart sapma değerleri elde edileceği gibi, farklı ölçü birimleriyle ifade edilmiş iki ayrı dizinin karşılaştırılması da yanıltıcı sonuçlar verecektir. İşte dağılım ölçülerinin bu dezevantajlarını gidermek için **Değişim Katsayısı** (D.K) geliştirilmiştir. Bu katsayıda mutlak dağılım yerine nisbi dağılma esas alınmıştır. Yani, değişim katsayısında standart sapma mutlak olarak değil X in yüzdesi olarak ifade edilmiştir. Değişim katsayısı,

$$D.K = \frac{s}{\bar{X}} \tag{6.20}$$

eşitliği ile hesaplanır. Kolay anlaşılması bakımından bazen bu katsayı 100 ile çarpılarak ifade edilir.

Örnek 6.33 *Aşağıdaki kümelerin standart sapmalarını ve varyanslarını bulunuz.*

i. $\{1, 2, 3, 4, 5, 6, 7\}$

ii. $\{12, 6, 7, 3, 15, 10, 18, 5\}$

Çözüm 6.33 *i. Standart sapma formülünden,*

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

X	*1*	*2*	*3*	*4*	*5*	*6*	*7*	$\sum X = 28$
X^2	*1*	*4*	*9*	*16*	*25*	*36*	*49*	$\sum X^2 = 140$

$$s = \sqrt{\frac{140}{7} - \left(\frac{28}{7}\right)^2} = \sqrt{20-16} = 2$$

bulunur. Varyans ise $s^2 = (2)^2 = 4$ *dür.*

ii.

X	*12*	*6*	*7*	*3*	*15*	*10*	*18*	*5*	$\sum X = 76$
X^2	*144*	*36*	*49*	*9*	*225*	*100*	*324*	*25*	$\sum X^2 = 912$

$$s = \sqrt{\frac{912}{8} - \left(\frac{76}{8}\right)^2} = \sqrt{114 - 90.25} = 4.87$$

dir. Varyans ise $s^2 = (4.87)^2 = 23.4247$ *olur.*

Örnek 6.34 *Aşağıdaki sınıflandırılmış ve greplandırılmış dizilerin standart sapmalarını, varyanslarını ve değişim katsayılarını bulunuz.*

i.

X	*8*	*10*	*16*	*18*	*24*
f	*5*	*3*	*5*	*10*	*5*

ii.

Sınıf aralığı	*10-14*	*15-19*	*20-24*	*25-29*	*30-34*
f	*6*	*6*	*10*	*9*	*9*

Çözüm 6.34 *Çözüm için bazı değerleri hesaplamak için önce aşağıdaki tabloyu yapalım,*

i.

X	*8*	*10*	*16*	*18*	*24*	
f	*5*	*3*	*5*	*10*	*5*	$\sum f = 28$
fX	*40*	*30*	*80*	*180*	*120*	$\sum fX = 450$
fX^2	*320*	*300*	*1280*	*3240*	*2880*	$\sum fX^2 = 8020$

Bulduğumuz toplamları formülde yerlerine koyduğumuzda, standart sapma,

$$s = \sqrt{\frac{\sum f(X-\bar{X})^2}{\sum f}} = \sqrt{\frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2}$$

$$s = \sqrt{\frac{8020}{28} - \left(\frac{450}{28}\right)^2} = \sqrt{286.43 - 258.28} = 5.31$$

olarak elde edilir. Varyans standart sapmanın karesi olduğundan,

$$varyans = s^2 = (5.31)^2 = 28.1961$$

olur. Ortalamanın

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{450}{28} = 16.07$$

olduğu dikkate alınarak, değişim katsayısı,

$$D.K = \frac{5.31}{16.07}(100) = 33.04$$

olarak bulunur. Yani, standart sapma aritmetik ortalamanın %33.04 *üdür. Bu sonuca göre verilen dizideki dağılım ortalama etrafında daha sık bir dağılım göstermiştir.*

ii.

Sınıf aralığı	10-14	15-19	20-24	25-29	30-34	
f	6	6	10	9	9	$\sum f = 40$
X	12	17	22	27	32	
fX	72	102	220	243	288	$\sum fX = 925$
fX^2	864	1734	4840	6561	9216	$\sum fX^2 = 23215$

$$s = \sqrt{\frac{23215}{40} - \left(\frac{925}{40}\right)^2} = \sqrt{580.38 - 534.77} = 6.75$$

Varyans,

$$varyans = s^2 = (6.75)^2 = 45.56$$

olarak bulunur.

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{925}{40} = 23.125$$

ve değişim katsayısı da,

$$D.K = \frac{6.75}{23.125}(100) = 29.19$$

olur.

Örnek 6.35 *Bir okulun A ve B sınıfındaki öğrencilerin boyları ölçülmüş ve boy ortalamaları sırasıyla* 162 *cm ve* 165 *cm ve standart sapmaları* 5 *cm ve* 6 *cm olarak hesaplanmıştır. Bu sonuçlara göre, hangi sınıfta boy ölçüleri ortalama etrafında daha sık dağılmıştır?*

Çözüm 6.35 $(D.K)_A$ *ile A'nın değişim katsayısını gösterelim. Bu durumda*

$$(D.K)_A = \frac{5}{162}(100) = 3.09$$

olur. $(D.K)_B$ *ile B'nın değişim katsayısını gösterelim. Bu durumda da*

$$(D.K)_B = \frac{6}{165}(100) = 3.64$$

olur. $(D.K)_A < (D.K)_B$ *olduğundan A sınıfındaki boy ölçüleri ortalama etrafında daha sık dağılmıştır.*

6.6 Dağılımın Şekli

6.6.1 Momentler

Tanım 6.17 *Bir* X_i *değişkeninin sıfır etrafındaki r.inci momenti,*

$$M_r = \frac{\sum X^r}{n} \tag{6.21}$$

formülüyle bulunur.

Sıfır etrafındaki birinci moment için $r = 1$ dir ve bu moment aritmetik ortalamaya eşittir.

Tanım 6.18 *Bir dizinin aritmetik ortalama etrafındaki r.inci momenti,*

$$M_r = \frac{\sum(X - \bar{X})^r}{n} \tag{6.22}$$

eşitliği ile hesaplanır.

Bir dizide birimlerin aritmetik ortalamadan sapmalarının toplamı sıfır olduğundan, ortalama etrafındaki birinci moment $M_1 = 0$ dır. Ortalama etrafındaki ikinci moment ise varyansı verir, yani $M_2 = s^2$ dir.

Tanım 6.19 *Sınıflandırılmış ve gruplandırılmış dizilerde sıfır etrafındaki r.inci moment,*

$$M_r = \frac{\sum fX^r}{\sum f} \tag{6.23}$$

formülüyle, ortalama etrafındaki r.inci moment ise

$$M_r = \frac{\sum f(fX - \bar{X})^r}{\sum f} \quad (6.24)$$

formülüyle hesaplanır.

Tanım 6.20 *X, $f(X)$ yoğunluk fonksiyonuna sahip sürekli rastlantı değişkeni ise r.inci mertebeden moment,*

$$M_r = \int_{-\infty}^{\infty} X^r f(X)dx \quad (6.25)$$

formülüyle hesaplanır.

Örnek 6.36 $\{4, 7, 5, 9, 8, 3, 6\}$ *kümesi için sıfır etrafındaki birinci, ikinci, üçüncü ve dördüncü momentleri bulunuz.*

Çözüm 6.36 *Birinci moment aynı zamanda aritmetik ortalamadır. Buna göre,*

$$\begin{aligned}
\text{Birinci moment} &= \frac{\sum X}{n} = \frac{4+7+5+9+8+3+6}{7} = 6, \\
\text{İkinci moment} &= \frac{\sum X^2}{n} = \frac{4^2+7^2+5^2+9^2+8^2+3^2+6^2}{7} = 40, \\
\text{Üçüncü moment} &= \frac{\sum X^3}{n} = \frac{4^3+7^3+5^3+9^3+8^3+3^3+6^3}{7} = 288, \\
\text{Dördüncü moment} &= \frac{\sum X^4}{n} = \frac{15360}{7} = 2188
\end{aligned}$$

olarak bulunur.

Örnek 6.37 $\{4, 7, 5, 9, 8, 3, 6\}$ *kümesi için ortalama etrafındaki birinci, ikinci, üçüncü ve dördüncü momentleri bulunuz.*

Çözüm 6.37 *Ortalama etrafındaki birinci moment,*

$$\begin{aligned}
M_1 &= \frac{\sum(X - \bar{X})}{n} \\
&= \frac{(4-6)+(7-6)+(5-6)+(9-6)+(8-6)+(3-6)+(6-6)}{7} = 0,
\end{aligned}$$

ikinci moment,

$$M_2 = \frac{\sum(X - \bar{X})^2}{n} = \frac{4+1+1+9+4+9+0}{7} = \frac{28}{7} = 4,$$

üçüncü moment,

$$M_3 \;=\; \frac{\sum(X-\bar{X})^3}{n} = \frac{-8+1-1+27+8-27+0}{7} = 0,$$

dördüncü moment,

$$M_4 \;=\; \frac{\sum(X-\bar{X})^4}{n} = \frac{16+1+1+81+16+81}{7} = \frac{196}{7} = 28$$

bulunur.

Örnek 6.38 *Aşağıdaki dağılım için ortalama etrafındaki ilk dört momenti bulunuz.*

X	*12*	*14*	*16*	*18*	*20*	*22*
f	*1*	*4*	*6*	*10*	*7*	*2*

Çözüm 6.38

X	12	14	16	18	20	22	
f	1	4	6	10	7	2	$\sum f = 30$
fX	12	56	96	180	140	44	$\sum fX = 528$
$X-\bar{X}$	-5.6	-3.6	-1.6	0.4	2.4	4.4	
$f(X-\bar{X})$	-5.6	-14.4	-9.6	4.0	16.8	8.8	$\sum f(X-\bar{X}) = 0.0$
$f(X-\bar{X})^2$	31.36	51.84	15.36	1.60	40.32	38.72	$\sum f(X-\bar{X})^2 = 179.20$
$f(X-\bar{X})^3$	-175.616	-186.624	-24.576	0.640	96.768	170.368	$\sum f(X-\bar{X})^3 = -119.040$
$f(X-\bar{X})^4$	983.4496	671.8464	39.3216	0.2560	232.2432	749.6192	$\sum f(X-\bar{X})^4 = 2676.7360$

Dağılımın aritmetik ortalaması,

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{528}{30} = 17.6$$

dır. Ortalama etrafındaki birinci moment,

$$M_1 = \frac{\sum f(X-\bar{X})}{\sum f} = 0,$$

ikinci moment,

$$M_2 = \frac{\sum f(X-\bar{X})^2}{\sum f} = \frac{179.2}{30} = 5.97,$$

üçüncü moment,

$$M_3 = \frac{\sum f(X-\bar{X})^3}{\sum f} = \frac{-119.04}{30} = -3.968,$$

dördüncü moment,

$$M_4 = \frac{\sum f(X-\bar{X})^4}{\sum f} = \frac{2676.736}{30} = 89.2245$$

olarak elde edilir.

Örnek 6.39 *Aşağıdaki dağılımın ortalama etrafındaki ilk üç momentini hesaplayınız.*

Sını aralığı	*3-5*	*6-8*	*9-11*	*12-14*	*15-17*
f	*4*	*6*	*10*	*8*	*2*

Çözüm 6.39

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{294}{30} = 9.8$$

Sınıf aralığı	3-5	6-8	9-11	12-14	15-17	
X	4	7	10	13	16	
f	4	6	10	8	2	$\sum f = 30$
fX	16	42	100	104	32	$\sum fX = 294$
$(X-\bar{X})$	-5.8	-2.8	0.2	3.2	6.2	
$f(X-\bar{X})$	-23.2	-16.8	2.0	25.6	12.4	$\sum f(X-\bar{X}) = 0$
$f(X-\bar{X})^2$	134.56	47.04	0.40	81.92	76.88	$\sum f(X-\bar{X})^2 = 340.80$
$f(X-\bar{X})^3$	-780.448	-131.712	0.080	262.144	476.656	$\sum f(X-\bar{X})^3 = -173.280$

Ortalama etrafındaki birinci moment,

$$M_1 = \frac{\sum f(X-\bar{X})}{\sum f} = 0,$$

ikinci moment,

$$M_2 = \frac{\sum f(X-\bar{X})^2}{\sum f} = \frac{340.8}{30} = 11.36,$$

üçüncü moment,

$$M_3 = \frac{\sum f(X-\bar{X})^3}{\sum f} = \frac{-173.28}{30} = -5.776$$

olarak bulunur.

6.6.2 Çarpıklık ve Basıklık

Ortalamaya göre üçüncü momente, simetrik olmayışın ölçüsü denir. Yani, çarpıklık, bir dağılımda simetriden ayrılışın derecesidir.

Dağılımın frekans eğrisi merkezsel maksimumun sağında daha uzun uzantıya sahipse dağılım için sağa yatık ya da pozitif yatıklığa sahiptir denir. Bunun tersi bir durum varsa sola yatık ya da negatif yatıklığa sahiptir denir. Simetrik ve simetrik olmayan dağılımlar aşağıdaki şekillerde gösterilmiştir.

Şekil 6.3:

Pearson Çarpıklık Katsayısı

Simetrik bir dizide X, *mod* ve *medyan* birbirine eşit olmadığında, bu değerler arasından, görülen farklılaşma çarpıklık için bir ölçü oluşturmaktadır. Bu noktadan hareketle Pearson tarafından aşağıdaki iki çarpıklık ölçüsü geliştirilmiştir:

$$\text{Çarpıklık} = \frac{3(\bar{X} - Medyan)}{s} \tag{6.26}$$

$$\text{Çarpıklık} = \frac{\bar{X} - Mod}{s} \tag{6.27}$$

Bu çarpıklık ölçüleri sıfır ise dağılım simetriktir, sıfırdan büyükse dağılım sağa çarpık, küçükse sola çarpıktır.

Moment Çarpıklık Katsayısı

Yukarıdaki iki ölçü yanında uygulamada çoklukla kullanılan bir diğer çarpıklık ölçüsü de "moment çarpıklık ölçüsü" dür. Bu ölçü ortalama etrafındaki üçüncü momentin, standart sapmanın üçüncü kuvvetine bölünmesiyle elde edilir. Yani,

$$\alpha = \frac{M_3}{s^3} \tag{6.28}$$

dür. Simetrik dağılımlarda moment çarpıklık ölçüsü sıfırdır. Eğer $\alpha > 0$ ise dağılım sağa çarpık, $\alpha < 0$ ise dağılım sola çarpıktır.

Kartil Çarpıklık Katsayısı

Herhangi bir dağılıma ait kartil değerleriyle de sözkonusu dağılımın şekli hakkında fikir sahibi olunabilir. Kartillerden yararlanmak suretiyle belli bir dağılımın çarpıklığı,

$$\text{Kartil çarpıklık katsayısı} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} \tag{6.29}$$

formülüyle hesaplanır.

Basıklık

Basıklık bir dağılımın diklik(sivrilik) derecesinin ölçüsüdür. Bu konuda kullanılan en yaygın ölçü "moment basıklık ölçüsü" dür. Moment basıklık ölçüsü ortalama etrafındaki dördüncü momentin standart sapmanın dördüncü kuvvetine bölünmesiyle elde edilir, yani,

$$\alpha_4 = \frac{M_4}{s^4} = \frac{M_4}{(M_2)^2} \tag{6.30}$$

dir. $\alpha_4 = 3$ olduğu takdirde yüksekliğin normal olduğu söylenir. $\alpha_4 > 3$ ise dağılım dik, $\alpha_4 < 3$ ise dağılım basıktır.

Örnek 6.40 *Aşağıdaki dağılımın,*

i. Pearson çarpıklık katsayısını,

ii. Kartil çarpıklık katsayısını,

iii. Moment çarpıklık katsayısını,

iv. Moment basıklık katsayısını hesaplayınız.

X	10	20	30	40	50
f	8	25	10	5	2

Çözüm 6.40 *i.*

$$Çarpıklık = \frac{3(\bar{X} - Medyan)}{s}$$
$$Çarpıklık = \frac{\bar{X} - Mod}{s}$$

X	10	20	30	40	50	
f	8	25	10	5	2	$\sum f = 50$
fX	80	500	300	200	100	$\sum fX = 1180$
Küm.f	8	33	43	48	50	
fX^2	800	10000	9000	8000	5000	$\sum fX^2 = 32800$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{1180}{50} = 23.6$$

$$Mod = 20, \quad Medyan = 20$$

$$s = \sqrt{\frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2} = \sqrt{\frac{32800}{50} - \left(\frac{1180}{50}\right)^2} = 9.925$$

$$Çarpıklık = \frac{3(\bar{X} - Medyan)}{s} = \frac{3(23.6 - 20)}{9.952} = 1.09$$
$$Çarpıklık = \frac{\bar{X} - Mod}{s} = \frac{23.6 - 20}{9.952} = 0.36$$

ii. Kartil çarpıklık katsayısı,

$$\text{Kartil çarpıklık katsayısı} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

X	*10*	*20*	*30*	*40*	*50*	
f	*8*	*25*	*10*	*5*	*2*	$\sum f = 50$
Küm.f	*8*	*33*	*43*	*48*	*50*	

$$\begin{aligned} \text{Birinci kartil,} \quad Q_1 &= \tfrac{N}{4} &= \tfrac{50}{4} = 12.5, \\ \text{İkinci kartil,} \quad Q_2 &= \tfrac{2N}{4} &= \tfrac{(2)(50)}{4} = 25, \\ \text{Üçüncü kartil,} \quad Q_3 &= \tfrac{3N}{4} &= \tfrac{(3)(50)}{4} = 37.5 \end{aligned} \tag{6.31}$$

dir. Buna göre

$$Q_1 = 20, \;\; Q_2 = 20, \;\; Q_3 = 30$$

ve

$$\text{Kartil çarpıklık katsayısı} = \frac{(30 - 20) - (20 - 20)}{30 - 20} = 1$$

olarak bulunur.

iii.

X	10	20	30	40	50	
f	8	25	10	5	2	$\sum f = 50$
$X - \bar{X}$	-13.6	-3.6	6.4	16.4	26.4	
$(X - \bar{X})^3$	-2515.456	-46.656	262.144	4410.944	18399.744	
$f(X - \bar{X})^3$	-20123.648	-1166.400	2621.440	20054.720	36799.488	$\sum f(X - \bar{X})^3 = 40185.600$

Üçüncü moment,

$$M_3 = \frac{\sum f(X - \bar{X})^3}{\sum f} = \frac{40185.6}{50} = 803.712$$

$s = 9.952$ *olduğundan* $s^3 = 985.669$ *olur. Böylece moment çarpıklık katsayısı,*

$$\alpha_3 = \frac{M_3}{s^3} = \frac{803.712}{985.669} = 0.815$$

şeklinde bulunur. Buna göre, dağılımın sağa çarpık olduğu söylenir.

iv.

X	10	20	30	40	50	
f	8	25	10	5	2	$\sum f = 50$
$X - \bar{X}$	-13.6	-3.6	6.4	16.4	26.4	
$(X - \bar{X})^4$	34210.2016	167.9616	1677.7216	72339.4816	485753.2416	
$f(X - \bar{X})^4$	273681.6128	4199.04	16777.216	361697.408	971506.4832	$\sum f(X - \bar{X})^4 = 1627861.76$

Dördüncü moment,

$$M_4 = \frac{\sum f(X-\bar{X})^4}{\sum f} = \frac{1627861.76}{50} = 32557.24$$

ve $s^4 = 9809.38$ *olduğundan, moment basıklık katsayısı,*

$$\alpha_4 = \frac{M_4}{s^4} = \frac{32557.24}{9809.38} = 3.32$$

olarak bulunur. Sözkonusu katsayı dağılımın normale göre dik olduğunu belirtmektedir.

6.7 Problemler

1. Aşağıdaki verileri bir dizi haline getiriniz ve değişim aralığını bulunuz.

$$99, 100, 102, 101, 98, 103, 100, 102, 99, 101$$

2. $6, 8, 5, 7$, dört işçinin günlük çalışma saatleridir:

   i. Aritmetik ortalamasını,

   ii. Geometrik ortalamasını,

   iii. Harmonik ortalamasını,

   iv. Mod ve medyan değerlerini,

   v. Değişim aralığını,

   vi. Standart sapma ve varyansını

   hesaplayınız.

3. Bir öğrencinin sınav sonuçları 83, 92, 69, 65, 89, 74, 70, 82 dir. ortalama değeri bulunuz?

4. 200 öğrencinin sınav sonuçlarının aritmetik ortalaması 55.2 ve 300 öğrencilik diğer bir grubun sınav sonuçlarının aritmetik ortalamasıda 60.5 olarak bulunmuştur. 500 öğrencinin birleştirilmiş kümesi düşünüldüğünde sınav sonuçlarının aritmetik ortalamasını bulunuz?

5. A ve B ile nitelendirilen iki sınıftaki 65 öğrencinin matematik sınav sonuçlarının ortalaması 68 dir. A sınıfındaki 30 öğrenci için ortalama 61 ise B sınıfındaki 27 öğrencinin sınav sonuçlarının ortalaması nedir?

6. 23 deneylik bir kümenin ortalaması 3 ve karelerinin toplamının ortalaması 90 ise bu küme için standart sapmayı bulunuz?

7. Aritmetik ortalaması 5 ve geometrik ortaması 4 olan iki sayıyı bulunuz?

8. Aşağıdaki frekans dağılımının geometrik ortalamasının bulunuz.

Sınıf aralığı	Frekanslar
0-4	8
5-9	10
10-14	20
15-19	35
20-24	15
25-29	10
30-34	3

9. Bir uçak 2300 km. lik uçuşunda yolun ilk 1/3'ünü ve son 1/3'ünü saatte 350 km hızla, ortadaki 1/3'lük yolu 400 km. lik hızla gittiğine göre uçağın hız ortalaması nedir?

10. $3, 5, 2, 6, 5, 9, 5, 2, 8, 6$ ölçümleri için aritmetik, geometrik ve harmonik ortalamayı, medyan ve mod'u bulunuz.

11. $6, 7, 10, 11, 14, 15, 16, 21$ sayıları için aritmetik, geometrik ve harmonik ortalamayı, medyan ve mod'u bulunuz.

12. $217, 011, 217.009, 217.006, 217.012, 217.005$ sayıları için aritmetik, geometrik ve harmonik ortalamayı, medyan ve mod'u bulunuz.

13. Aşağıdaki veriler için aritmetik, ortalamayı, medyan ve mod'u bulunuz.

X	f
93-97	2
98-102	5
103-107	12
108-112	17
113-117	14
118-122	6
123-127	3
128-132	1
	60

14. 40 öğrencinin boy uzunlukları üzerinde yapılan bir araştırma sonunda aşağıdaki ölçümler alınmıştır. Bu ölçümlerin aritmetik, ortalama'sını, medyan'nını ve mod'unu bulunuz.

Uzunluk (X)	f
118-126	3
127-135	5
136-144	9
145-153	12
154-162	5
163-171	4
172-180	2
	40

15. Aşağıdaki veriler için değişim aralığını, ortalama sapma, standart sapma ve varyansı bulunuz.

   i. $3, 5, 6, 9, 10, 13, 15, 16, 18, 20,$

   ii. $3, 8, 8, 8, 8, 9, 9, 9, 15, 18,$

   iii. $61, 63, 64, 65, 66, 66, 68, 72,$

   iv. $119, 120, 123, 125, 126, 127, 127, 132, 134, 135$

16. $2, 3, 6, 7, 8, 11, 13, 14, 15$ sayıları için sıfır ve ortalama civarındaki ilk üç momenti bulunuz.

17. Bir yüksek okuldaki 100 öğrencinin ağırlık dağılımı aşağıdaki tabloda verilmiştir.

Ağırlık (kg)	f
60-62	5
63-65	21
66-68	45
69-71	27
72-74	

i. Sıfır civarındaki,

ii. Ortalama civarındaki,

iii. $a = 65$ noktası civarındaki

momentleri bulunuz.

18. Bir önceki problem için basıklık ve çarpıklık katsayılarını bulunuz?

Chapter 7

ÖRNEKLEM DAĞILIMLARI VE PARAMETRE TAHMİNİ

7.1 Kitle ve Örneklem (Örnek)

İstatistiğin önemli bir görevi kitle parametrelerinin (ortalam, standart sapma, varyans, vs...) örneklem değerleri(örneklem istatistikleri) yardımıyla tahmin edilmesidir. Uygulamada, bütün kitlenin incelenmesi çoğu zaman mümkün olmaz. Bunun yerine söz konusu kitleden alınan bir şans örneğinin incelenmesi yoluna gidilir. Elde edilen örneklem değerlerinin kitle parametresi yerine kullanılabilmesi için iki önemli şart vardır. Birincisi, örneklemin "şans örneği" olması, yani kitledeki her birimin örnekleme girme şansının eşit olması, ikinci şart ise "örneklemin yeterince büyük" olmasıdır. Bu ikinci şarta göre kitle büyüdükçe örnelemin de büyük tutulması gerekecektir.

Örnekleme ya yerine koymalı(iadeli) yada yerine koymamalı(iadesiz) olur. Çekilen birimin kitleye tekrar iade edilmesi halinde "iadeli örnekleme", aksi halde "iadesiz örnekleme" söz konusudur. Herhangi bir kitle birimi, iadeli örneklemede örneğe bir kaç kere girebileceği halde iadesiz örneklemede bir kere girebilir.

Örnekleme ya sınırlı veya sınırsız kitleler için yapılır. Örneklemenin iadeli olarak yapıldığı kitle sınırsız kabul edilir.

7.2 Örneklem Dağılımları

Bir kitleden alınan şans örneklerinin her birisi için örnek istatistikleri hesaplandığında örnek dağılımları ortaya çıkar. Örneğin, her bir örneğin ortalaması hesaplanmışsa elde edilen X_i dağılımı "Ortalamaların Örnek dağılımı" dır. Aynı şekilde, her örnek için p oranları hesaplandığında "Oranların Örnek dağılımı" elde edilmiş olur. İki ayrı kitlenin karşılaştırılması sözkonusu olduğunda ise farklarla ilgili örnek dağılımları ortaya çıkar. her iki kitleden alınan n_1 ve n_2 büyüklüğündeki örneklerin ortalamaları hesaplanmış ve bu X_i ve X_j değerleri arasındaki farklar belirlenmişse elde edilen dağılım "Ortalamalar Arası Farkların Örnek Dağılımı"dır. Aynı şekilde, bu kitlelerden alınan örnekler için oranlar hesaplanmış ve bu oranların kitleler itibariyle gösterdikleri farklılıklar ortaya konulmuşsa, elde edilen dağılım "Oranlar Arası Farkların Örnek Dağılımı"dır.

7.2.1 Ortalamaların Örneklem Dağılımı

Tanım 7.1 *Ortalamaların örneklem dağılımının ortalaması $\mu_{\bar{x}}$ ve standart hatası $\sigma_{\bar{x}}$*

$$\mu_{\bar{x}} = \mu \quad ve \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7.1}$$

eşitlikleriyle elde edilir. Bu dağılımın ortalaması, kitle ortalamasına, varyansı ise kitle varyansının örnekteki birim sayısına bölümüne eşittir. Dolayısıyla, standart hata

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \quad = \quad \frac{\sigma}{\sqrt{n}} \tag{7.2}$$

olur. Örnekleme sınırlı kitleden iadesiz olarak yapılmışsa, standart hata

$$\sqrt{\tfrac{N-n}{N-1}} \tag{7.3}$$

düzeltme faktörü ile çarpılır. Yani, sınırlı N hacminde bir kitleden iadesiz olarak çekilen n hacmindeki örneğe ait standart hata,

$$\sigma_{\bar{x}} \quad = \quad \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \tag{7.4}$$

şeklindedir.

Buna göre, herhangi bir örnek ortalamasının z değeri,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \tag{7.5}$$

eşitliği ile hesaplanır.

Örnek 7.1 500 *tane bilyeli yatağın ortalama ağırlığı* 155 *gr ve standart sapması,* 9.3 *gr bulunmuştur. Bu gruptan alınan* 100 *bilyeli yataklık bir şans örneğinde ortalama ağırlığın,*

i. 154 *gr ile* 155 *gr arasında,*

ii. 158 *gr dan daha fazla olma olasılıklarını bulunuz.*

Çözüm 7.1 *kitle sınırlı olduğu için* $\left(\frac{n}{N} = \frac{100}{500} = 0.20 > 0.05\right)$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

standart hata formülü kullanılır. $N = 500$, $n = 100$, $\mu = 155$, $\sigma = 9.3$ *olarak verilmiştir.*

i.

$$\sigma_{\bar{x}} = \frac{9.3}{10}\sqrt{\frac{400}{499}} = 0.83$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{154 - 155}{0.83} = -1.2$$

.

Şekil 7.1:

$$P(154 < x < 155) = P(-1.2 < z < 0) = 0.3849.$$

ii.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{158 - 155}{0.83} = 3.61$$

Şekil 7.2:

$$P(158 < x) = P(3.61 < z) = 0.5 - 0.4998 = 0.0002$$

bulunur.

Örnek 7.2 *Bir şirketin imal ettiği tüplerin ortalama dayanma süresi* 800 *saat ve standart sapması* 60 *saattir. Bu tüplerden alınan* 36 *tüplük bir şans örneğinde ortalama dayanma müddetinin;*

i. 790 *ve* 810 *saat arasında,*

ii. 785 *saatten az,*

iii. 820 *saatten fazla,*

iv. 770 *ve* 830 *saat arasında olma olasılıklarını bulunuz.*

Çözüm 7.2 $\mu = 800$, $n = 36$, $\sigma = 60$, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = -1$ *olduğundan,*

i.

$$z_1 = \frac{790 - 800}{60/\sqrt{36}} = -1, \quad z_2 = \frac{810 - 800}{60/\sqrt{36}} = 1$$

Şekil 7.3

$$P(790 < x < 800) = 2(0.3413) = 0.6826$$

ii.

$$z = \frac{785 - 800}{60/\sqrt{36}} = -1.5$$

Şekil 7.4

$$P(x < 785) = P(-1.5 < z < 0) = 0.5 - 0.4332 = 0.0668$$

iii.

$$z = \frac{820 - 800}{60/\sqrt{36}} = 2$$

Şekil 7.5

$$P(x > 820) = P(z > 2) = 0.5 - 0.4772 = 0.0228$$

iv.

$$z_1 = \frac{770 - 800}{60/\sqrt{36}} = -3, \quad z_2 = \frac{830 - 800}{60/\sqrt{36}} = 3$$

Şekil 7.6

$$P(770 < x < 830) = 2[P(0 < z < 3)] = 2(0.4987) = 0.9974$$

olarak bulunur.

7.2.2 Oranların Örneklem Dağılımı

Tanım 7.2 *Oranların örneklem dağılımının ortalaması ve standart hatası,*

$$\mu_P = P \quad ve \quad \sigma_P = \sqrt{\frac{PQ}{n}} \tag{7.6}$$

eşitlikleriyle elde edilir. sınırlı kitlelerde yapılan iadesiz örneklemeler için de yukarıdaki

$$\sqrt{\frac{N-n}{N-1}} \tag{7.7}$$

düzeltme faktörü kullanılır. Bir örnek oranının z değeri,

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} \tag{7.8}$$

eşitliğiyle hesaplanır.

Kesikli verilerin sürekli yapılabilmesi için ortalamalarda ± 0.5lik düzeltme gerekir. Buna karşı gelen oranlarda $\pm 1/2n$lik bir düzeltmeye gitmek daha güvenilir sonuçlar sağlar. Örneğin, 4 rakamı, 4 ± 0.5 yani 3.5 ve 4.5 aralığında ifade edilirken; 0.7 oranı, $n = 15$ için

$$0.7 \pm \frac{1}{2(15)}$$

aralığında temsil edilir. Bu aralık 0.667 ve 0.773 aralığıdır.

Örnek 7.3 *Erkek ve kız olma olasılıklarını eşit kabul ederek doğacak* 200 *çocuktan,*

i. %40 *azının erkek olma,*

ii. %43 – %58 *kız olma,*

iii. %54 *fazlasının erkek olma,*

iv. %50 *erkek olma olasılıklarını bulunuz.*

Çözüm 7.3 $n = 200$, $p = 0.50$, $q = 0.50$ *olduğundan:*

i. $z = \frac{p-P}{\sqrt{PQ/n}}$ *formülünü kullanacağız.* %40 *dan az denildiği için,* $P(p < 0.4 - \frac{1}{2n})$ *değeri bulunacaktır.*

$$z = \frac{[0.4 - (1/400)] - 0.5}{\sqrt{(0.5)(0.5)/200}} = \frac{-0.1025}{0.0354} = -2.9$$

Şekil 7.7

$$P(z < -2.9) = 0.5 - 0.4981 = 0.0019$$

ii.

$$z_1 = \frac{0.43 - 0.50}{0.0354} = -1.98, \quad z_2 = \frac{0.58 - 0.50}{0.0354} = 2.26$$

Şekil 7.8

$$P(-1.98 < z < 2.26) = 0.4761 + 0.4881 = 0.9642$$

iii.

$$z = \frac{0.54 + (1/400) - 0.50}{0.0354} = 1.2$$

.

Şekil 7.9

$$P(z > 1.2) = 0.5 - 0.3849 = 0.1151$$

iv.

$$z_1 = \frac{0.50 + (1/400) - 0.50}{0.0354} = -0.07,\ z_2 = \frac{0.50 + (1/400) - 0.50}{0.0354} = 0.07$$

Şekil 7.10

$$P(-0.07 < z < 0.07) = 2(0.0279) = 0.0558$$

bulunur.

7.2.3 Ortalamalar Arası Farkların Örneklem Dağılımı

Tanım 7.3 *Bu dağılımın ortalaması ve standart hatası, aşağıdaki*

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \ ve \ \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{7.9}$$

eşitlikleriyle bulunur. Burada, $\mu_1 - \mu_2$ *kitle ortalamaları arasındaki farktır.* σ_1 *birinci kitlenin standart sapması* σ_2 *ise ikinci kitlenin standart sapmasını;* n_1 *birinci kitle için örnek büyüklüğü,* n_2 *ikinci kitle için örnek büyüklüğünü ifade eder. Ortalamalar arası farklara ait* z *değeri ise,*

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{7.10}$$

şeklinde hesaplanır.

Örnek 7.4 *A ve B firmalarının ürettikleri kabloların ortalama kırılma gücü, sırayla,* 1800 *kg ve* 2000 *kg. Standart sapmaları ise* 135 *kg ve* 90 *kg dır. A marka* 100 *kablo ile B marka* 50 *kablo teste tabi tutulduğunda, B nin ortalama kırılma gücünün A dan*

i. en çok 270 *kg fazla,*

ii. en az 180 *kg fazla, olma olasılıkları nedir ?*

Çözüm 7.4

$$\begin{array}{ll} \mu_1 = 1800\ kg & \mu_2 = 2000\ kg \\ \sigma_1 = 135\ kg & \sigma_2 = 90\ kg \\ n_1 = 100 & n_2 = 50 \end{array}$$

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

formülünü kullanacağız.

i.

$$\bar{X}_2 - \bar{X}_1 \leq 270 \text{ ve } \mu_1 - \mu_2 = 200$$

şeklindedir.

$$z = \frac{270 - 200}{\sqrt{\frac{135^2}{100} + \frac{90^2}{50}}} = 3.77$$

Şekil 7.11

$$P(z < 3.77) = 0.5 + 0.4999 = 0.9999$$

ii.

$$\bar{X}_2 - \bar{X}_1 \geq 180$$

$$z = \frac{180 - 200}{18.55} = -1.08$$

Şekil 7.12

$$P(z > -1.08) = 0.5 + 0.3599 = 0.8599$$

7.2.4 Oranlar Arası Farkların Örneklem Dağılımı

Tanım 7.4 *Bu dağılımın ortalaması ve standart hatası,*

$$\mu_{p_1-p_2} = P_1 - P_2 \text{ ve } \sigma_{p_1-p_2} = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}} \tag{7.11}$$

eşitlikleriyle hesaplanır. Burada, $P_1 - P_2$ kitle oranları arasındaki fark, $Q_1 = 1 - P_1$ ve $Q_2 = 1 - P_2$ dir. Formüldeki n_1 birinci kitleden alınan örneğin ve n_2 ikinci kitleden alınan örneğin büyüklüğüdür. İki örnek oranı arasındaki farkın z değeri:

$$z = \frac{(P_1 - P_2) - (Q_1 - Q_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

eşitliğiyle hesaplanır.

Örnek 7.5 *Seçim istatistiklerine göre, belirli bir aday oyların %65 ini almıştır. Herbiri 200 seçmenden ibaret olan iki şans örneğinde, sözkonusu adaya ait oy oranlarının %10 dan daha fazla bir farklılık gösterme olasılığını bulunuz.*

Çözüm 7.5 $P_1 = 0.65$, $P_2 = 0.35$ *ve* $P_1 - P_2 = 0.30$

$$z = \frac{(P_1 - P_2) - (Q_1 - Q_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} = \frac{(0.10 - 0.30)}{\sqrt{\frac{(0.65)(0.35)}{200} + \frac{(0.35)(0.65)}{200}}} = \frac{-0.20}{0.0477} = -4.192$$

$$P(P_1 - P_2 > 0.10) = 1 \ (\textit{yani yaklaşık } \%100 \textit{ olasılıkla})$$

7.3 Tahmin Teorisi

Tahmin teorisinin konusu, kitle parametrelerinin örnek istatistikleri yardımıyla tahmin edilmesidir. Eğer kitle parametresi bir tek değer olarak tahmin ediliyorsa bu tahmine "nokta tahmini" denilir. Nokta tahminlerinde örnekten elde edilen değer, doğrudan, kitle parametresinin bir tahmini olarak işlem görür. Uygulamada daha çok, kitle parametresinin bir tek değer - yerine belli bir önem seviyesinde ve belli bir aralıkta tahmin edilmesi yoluna gidilmektedir. Şöyle ki, bir ornek ortalamasının 15 olması halinde kitle parametresini $\mu = X = 15$ olarak tahmin etmek bir "nokta tahmini" dir. Bu tahmine, belli bir ihtimal seviyesine ait, hata payının (örneğin, 2 birimlik bir hatanın) eklenip, çıkartılmasıyla, kitle ortalamasının 15 ± 2 yani 13 ile 17 arasında tahmin etmek bir "aralık tahmini" dir. Yukarıda da işaret ettiğimiz gibi aralık tahmininde, yapılan tahmin belli bir ihtimal seviyesi ile ifade edilir. Örneğin, yaptığımız bu aralık tahmininde kitle ortalamasının %95 veya %99 ihtimalle 13 ile 17 arasında olduğu söylenir.

7.3.1 Güven Sınırları

Örnek değerlerine belli bir hata payının eklenip çıkarılmasıyla kitle parametresinin güven sınırları elde edilir. Bu güven sınırlarından küçük olanına "alt güven sınırı" büyüğüne ise "üst güven sınırı" denilir.

7.3.2 Ortalamalar İçin Güven Aralığı

Tanım 7.5 *Ortalamalar için güven aralığı,*

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} \tag{7.12}$$

şeklinde hesaplanır. Örnek ortalamasına eklenip çıkartılan

$$z\frac{\sigma}{\sqrt{n}} \tag{7.13}$$

değeri "tahmin hatası" dır.

Güven sınırlarının belirleneceği olasılık durumuna göre z değeri değişir. Örneğin, güven sınırları %95 olasılıkla belirlenecekse z cetvelinde ortalamanın sağ ve solunda simetrik olarak, %95 lik sahayı sınırlayan z değerine bakılır. Bu değerler ± 1.96 dır. Aynı şekilde %99 güven seviyesi için z değerleri ± 2.58 bulunur. Örneklemenin sınırlı bir kitleden iadesiz olarak yapılması durumunda tahmin hatası,

$$\sqrt{\frac{N-n}{N-1}} \tag{7.14}$$

değeriyle çarpılır, yani bu kitleler için güven sınırları

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \tag{7.15}$$

formülüyle hesaplanır.

Uygulamada kitle ortalaması gibi, kitle standart sapması da genellikle bilinmez. Zaten bu parametrelerin bilinmesi halinde bunların tahmininden söz edilemez. Bu sebeple güven sınırları için takdim ettiğimiz yukarıdaki ifadelerde σ bilinmeyeceğinden örnek standart sapması s, kitle standart sapmasının (σ nın) tahmini olarak işleme girer. Büyük örneklerden ($n > 30$), σ'nın bu şekilde tahmini gerçeğe yakın sonuçlar verir.

Örnek 7.6 *Bir makinada, bir hafta içerisinde yapılan* 200 *bilyeli yatağın çapları ölçülmüş ve ortalama* $2.09cm$ *standart sapma ise* $0.11cm$ *bulunmuştur. Bütün bilyeli yataklara ait çap ortalaması için;*

i. %95 *ve*

ii. %99 *güven sınırlarını bulunuz.*

Çözüm 7.6

$$n = 200, \ \ \bar{x} = 2.09cm, \ \ s = 0.11cm.$$

i. %95 *güven sınırları:*

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = 2.09 \pm 1.96\frac{0.11}{\sqrt{200}} = 2.09 \pm 0.015$$

olarak bulunur. Bu ifadeyi daha açık olarak:

$$\begin{aligned} \textit{Alt güven sınırı} &= 2.09 - 0.015 = 2.075 \\ \textit{Üst güven sınırı} &= 2.09 + 0.015 = 2.105 \end{aligned}$$

şeklinde de yazmak mümkündür.

ii. %99 *güven sınırları:*

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = 2.09 \pm 2.58\frac{0.11}{\sqrt{200}} = 2.09 \pm 0.02$$

dir. Daha açık bir ifadeyle,

$$\begin{aligned} \textit{Alt güven sınırı} &= 2.09 - 0.02 = 2.07 \\ \textit{Üst güven sınırı} &= 2.09 + 0.02 = 2.11 \end{aligned}$$

dir.

7.3.3 Oranlar için Güven Aralığı

Tanım 7.6 *Oranlar için güven aralığı,*

$$p \pm z\sqrt{\frac{PQ}{n}} \tag{7.16}$$

şeklinde hesaplanır. Örnekleme sınırlı bir kitleden iadesiz olarak yapılıyorsa, güven sınırları

$$p \pm z\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}} \tag{7.17}$$

formülüyle hesaplanır.

Örnek 7.7 *Bir belediye seçimi için tahmin yapmak üzere seçmenlerden şansa bağlı olarak* 100 *kişilik bir örnek alınmış ve bunların* %55*'inin belli bir adayın lehinde oldukları tesbit edilmiştir. Bu adayın lehinde olan bütün seçmenlerin oranı için* %99 *güven sınırlarını bulunuz.*

Çözüm 7.7

$$n = 100, \qquad p = 0.55, \qquad q = 1 - p = 0.45$$

$$p \pm z\sqrt{\frac{PQ}{n}} = 0.55 \pm 2.58\sqrt{\frac{(0.55)(0.45)}{100}} = 0.55 \pm 2.58(0.0497)$$

şeklinde hesaplanır. Daha açık bir ifadeyle,

$$\begin{aligned} \textit{Alt güven sınırı} &= 0.43 \\ \textit{Üst güven sınırı} &= 0.663 \end{aligned}$$

dir.

7.3.4 Ortalamalar Arası Farklar İçin Güven Aralığı

Tanım 7.7 *Ortalamalar arası farklar için güven aralığı,*

$$(\bar{X}_1 - \bar{X}_2) \pm z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{7.18}$$

formülüyle hesaplanır.

Örnek 7.8 *Bir üniversitenin A fakültesinden alınan* 200 *büyüklüğündeki bir şans örneğinde not ortalaması* 7 *ve standart sapma* 1.2*; B fakültesinden alınan* 150 *büyüklüğündeki bir şans örneğinde ise not ortalaması* 5 *ve standart sapma* 0.9 *olarak tesbit ediliyor. Örneklerin verdiği bu bilgilere göre söz konusu iki fakültenin not ortalamaları arasındaki farkın* %90 *güven sınırlarını bulunuz.*

Çözüm 7.8

$$\begin{aligned} n_1 &= 200 & n_2 &= 150 \\ x_1 &= 7 & x_2 &= 5 \\ s_1 &= 1.2 & s_2 &= 0.9 \end{aligned}$$

Ortalamalar arası farklar için %90 *güven sınırları*

$$\begin{aligned} (\bar{X}_1 - \bar{X}_2) \pm z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= \\ (7-2) \pm 1.645\sqrt{\frac{(1.2)^2}{200} + \frac{(0.9)^2}{150}} &= 5 \pm 1.645(0.112) \end{aligned}$$

şeklinde hesaplanır. Daha açık bir ifadeyle,

$$\begin{aligned} \textit{Alt güven sınırı} &= 5 - 0.18 = 4.82 \\ \textit{Üst güven sınırı} &= 5 + 0.18 = 5.18 \end{aligned}$$

dir.

7.3.5 Oranlar Arası Farklar İçin Güven Aralığı

Tanım 7.8 *Oranlar arası farklar için güven aralığı,*

$$(p_1 - p_2) \pm z\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}} \tag{7.19}$$

formülüyle hesaplanır.

Örnek 7.9 *Bir kitabevi yayınladığı bir eserin hangi yaş grupları arasında daha fazla tutulduğunu tesbit etmek üzere bir anket yapılıyor. Anket sonunda, 80 çoçuktan 20 sinin eseri beğendiği, 120 gencin arasında da 70 inin bu eseri beğendiği tesbit ediliyor. Bu bilgiye göre gençlerin ve çoçukların eseri beğenme oranları arasındaki fark için* %95 *güven sınırlarını bulunuz.*

Çözüm 7.9

$$P_1 = \frac{20}{80} \;=\; 0.25 \;\; P_2 = \frac{70}{120} = 0.58$$

%95 *için oranlar arası farkların güven sınırları,*

$$\begin{aligned} (p_1 - p_2) \pm z\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}} &= \\ (0.58 - 0.25) \pm 1.96\sqrt{\frac{(0.25)(0.75)}{80} + \frac{(0.58)(0.42)}{120}} &= 0.33 \pm 1.96(0.066) \end{aligned}$$

şeklinde bulunur. Daha açık bir ifadeyle,

$$\begin{aligned} \textit{Alt güven sınırı} &= 0.33 - 0.13 = 0.20 \\ \textit{Üst güven sınırı} &= 0.33 + 0.13 = 0.46 \end{aligned}$$

olarak bulunur.

7.3.6 Standart Sapmalar İçin Güven Aralığı

Tanım 7.9 *Standart sapmalar için güven aralığı,*

$$s \pm z\frac{\sigma}{\sqrt{2n}} \tag{7.20}$$

formülüyle hesaplanır. σ bilinmediği zaman örnek standart sapması s, σ'nın bir tahmini olarak kullanılır.

Örnek 7.10 *Bir psikolog reaksiyon zamanı ölçümünde standart sapmayı 0.05 saniye olarak tahmin ediyor. Psikoloğun, yaptığı bu tahmindeki hatanın 0.01 saniyeyi geçmeyeceğinden %95 güvenli olabilmesi için örnek büyüklüğünü ne kadar alması gerekmektedir.*

Çözüm 7.10 $s = 0.05$ *dir.*

$$\textit{Tahmin hatası} \;=\; 1.96\frac{\sigma}{\sqrt{2n}} = 0.01 \implies 1.96\frac{\sigma}{\sqrt{2n}} = 0.01$$

$$\begin{aligned} 196(0.05) &\leq \sqrt{2n} \implies [196(0.05)]^2 \leq 2n \\ n &\geq \frac{196(0.05)^2}{2} = 48 \end{aligned}$$

7.3.7 Problemler

1. Bir şirket tarafından yapılan paketlerin ortalama ağırlığı $30kg$ ve standart sapma $5kg$ dır. Emniyet sınırları $820kg$ olarak belirlenmiş bulunan bir asansöre, şansa bağlı olarak yüklenen 25 paketin emniyetle götürülme olasılığını bulunuz.

2. Bir torbada %60'ı kırmızı, %40'ı beyaz olmak üzere 80 tane bilye bulunmaktadır. Torbadan tekrar yerine koymak şartıyla (iadeli) çekilen ve herbiri 20 bilyeden meydana gelen 50 örneğin kaç tanesinde,

   i. Eşit sayıda kırmızı ve beyaz bilye,

   ii. 12 kırmızı 8 beyaz bilye,

   iii. 10 veya daha fazla beyaz bilye bulunacağını tahmin edersiniz.

3. Bir üretici size herbiri 100 elektrik ampülünden meydana gelen 1000 takım ampül göderiyor. Normal olarak, ampüllerin %5'inin kusurlu çıkması halinde takımlardan kaç tanesinde,

   i. 90 dan az sağlam ampül,

   ii. 98 veya daha fazla sağlam ampül çıkacağını tahmin edersiniz?

4. Bir zeka testinde öğrencilerin not ortalaması 72, standart sapma ise 8 bulunmuştur. 28 ve 436 öğrenciden meydana gelen iki öğrenci grubunda not ortalamasının birbirinden,

   i. 3 veya daha fazla puan,

   ii. $2-5$ puan kadar farklılık gösterme olasılıkları nedir?

5. Sigara içenlerin oranının p olduğu bir kitleden alınan $N = 100$ olan örnekte 35 kişinin sigara içtiği görülüyor. p için %95 lik güven aralığını hesaplayınız?

6. Deneyde kullanılan farelerin bir kitlesinden bir test uygulamak amacıyla 200 fare seçilmiştir. 67'sinin testte başarılı, geri kalanının başarısız olduğunu kabul edelim. Kitle içinde testte başarılı olanların gerçek oranı için %95 lik güven aralığını bulunuz?

7. 4000 kişilik bir lisedeki öğrenciler arasından seçilen 300 öğrencinin boy uzunluklarının ortalaması 165 cm ve standart sapması 21 cm dir. 4000 öğrencinin ortalama boy uzunluğu için %90 lik ve %99 lık güven aralıklarını bulunuz?

8. Bir üniversitedeki rastgele seçilen 201 öğrencinin yetenek puanlarının varyansı 125 olarak belirlenmiştir. Üniversitedeki tüm öğrencilerin yetenek puanlarının varyansı için %95 lik güven aralığı bulunuz?

9. Bir kutuda bilinmeyen oranda yeşil ve siyah top vardır. iadeli olarak çekilen 75 topun %60 ının siyah olduğu görülüyor. Kutudaki siyah topların gerçek oranının %90 ve %95 lik güven sınırlarını bulunuz?

10. Bir depodaki piller arasından seçilen 50 pilin voltajının standart sapması 0.5 volt olarak bulunmuştur. Böyle pillerin voltajının varyansı için %905 ve %99 luk güven sınırlarını bulunuz.

11. Yurdumuzun Karadeniz Bölgesinde Nisan ayına ait yağış ortalaması son 35 yıllık verilere göre 20 cm ve standart sapma 3 cm dir. Marmara Bölgesinde aynı ay içinde son 20 yıllık verilere göre ortalama yağış 14 cm ve standart sapma 4 cm dir. Gözlemlerin farklı varyanslı normal

dağılımlı kitlelerden geldiğini varsayarak, bu iki bölgedeki gerçek ortalama yağışların farkı için %95 lik güven aralığı bulunuz.

12. Benzer iki A ve B gruplarında sırasıyla 85 ve 135 koyun vardır. A grubuna yeni cins bir yiyecek, B grubuna ise kontrol yiyeceği verilmiştir. A grubunda belli bir zaman aralığında kazanılan ortalama ağırlık 32 kg ve standart sapma 4 kg dır. B grubunda aynı zaman aralığında kazanılan ortalama ağırlık 27 kg ve standart sapma 3 kg dır. İki eşit yiyeceğin sebep olduğu ortalama ağırlık artışlarındaki fark için %95 ve %99 lık güven aralıklarını bulunuz.

Chapter 8

HİPOTEZ TESTLERİ

8.1 Sıfır Hipotezi ve Alternatif Hipotez

Kitle parametreleri hakkında tahmin yürütürken bir takım kabullerden hareket edilir. Doğru veya yanlış olması muhtemel olan bu kabullere "hipotez" denilir. Uygulamada genellikle hipotez kurulurken bu hipotezin reddedilmesi, yani yanlışlığının ispatı esas alınır. Bu şekilde kurulan hipotezlere "sıfır hipotez" denilir. Örneğin, kitle ortalamasıyla ilgili bir hipotez testinde sıfır hipotezi,

$$H_0 : \mu = 300$$

şeklinde kurulmuşsa, bu hipotezi reddetmek üzere, karşı yani alternatif hipotez

$$H_1 : \mu \neq 300, \text{ veya } H_1 : \mu < 300, \text{ veyahut } H_1 : \mu > 300$$

şeklinde kurulur. Bazı problemlerde örnek istatistiğinin kitle parametresine eşit olması, yani, bu parametreden önemli ölçüde sapma göstermemesi istenir. Bu gibi durumlarda çift yönlü test sözkonusudur. Çift yönlü testte doğru bir hipotezin reddedilme olasılığını gösteren α önem seviyesi normal eğrinin sağ ve sol uçlarında eşit oranlarda yer alır. Örneğin, hipotezi %95 olasılıkla yani %5 önem seviyesinde test ettiğimiz takdirde normal eğrinin her iki ucuna

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \tag{8.1}$$

birimlik alan düşer. Hipotezin testinde esas alacağımız örnek istatistiğine ait z değeri, bu bölgelere düştüğü takdirde hipotez red, aksi halde kabul edilir.

Bazı problemlerde ise örnek istatistiğinin belli bir değerden büyük olmaması üzerinde durulur. Yani, örnek istatistiğine ait z değerinin belli bir değerden büyük olması halinde hipotezin reddi yoluna gidilir. Böyle durumlarda alternatif hipotez şöyle kurulu:

$$H_1 : \text{Parametre} > A \;\; \text{(A: Belli bir değer)}$$

Alternatifin bu şekilde kurulması halinde α önem seviyesine ait bölge normal eğrinin sağ ucunda yer alır.

Örnek istatistiğinin belli bir değerden küçük olmamasının istendiği problemde ise alternatif hipotez,

$$H_1 : \text{Parametre} < A \;\; \text{(A: Belli bir değer)}$$

şeklinde kurulur. Bu problemde örnek istatistiğine ait z değerinin belli bir z değerinden küçük çıkması halinde hipotez reddedilir. Buna göre böyle problemlerde α önem seviyesine ait bölge, normal eğrinin sol ucunda yer alır.

Yukarıda ifade edilen tek ve çift yönlü hipotez testlerine ait kabul ve red bölgeleri aşağıdaki gibidir.

0.05 önem seviyesi için z değerleri(kritik değerleri), çift yönlü testte ± 1.96 , tek yönlü testlerde ya -1.645 veya 1.645 olur.

0.01 önem seviyesi için ise z değerleri, çift yönlü testte ± 2.58 , tek yönlü testlerde ya -2.33 veya 2.33 olur.

Yukarıda da ifade edildiği gibi gözlem sonuçlarına dayanarak bir hipotez ya reddedilir ya da kabul edilir. Böyle bir incelemede aşağıdaki durumlardan biri ortaya çıkar:

i.	H_0	doğrudur ve kabul edilir.
ii.	H_0	doğrudur ve reddedilir.
iii.	H_0	yanlıştır ve kabul edilir.
iv.	H_0	yanlıştır ve reddedilir.

Bu dört durum sonucunda iki çeşit hata yapılabilir. H_0 doğru olduğu halde reddedilirse *I.tip hata*, yanlış iken kabul edilirse *II.tip hata* yapılır. Bu tür hatalar aşağıdaki tabloda özetlenmiştir.

	Karar	
Hipotez	H_0 kabul	H_0 red
H_0 doğru	Doğru karar	I.tip hata
H_0 yanlış	II.tip hata	Doğru karar

H_0 hipotezinin reddedildiği bir deneyin sonuçlarının kümesine *kritik bölge* ya da *red bölgesi* denir. H_0 hipotezinin kabul edildiği sonuçların kümesine de *kabul bölgesi* denir.

Yukarıda da değindiğimiz gibi I.tip hata yapma olasılığına bir deneyin *önem* seviyesi denir ve α ile göterilir. II.tip hata yapma olasılığı β ile gösterilir. Yani,

$$\begin{aligned} \alpha &= P(H_0 \text{ red}/H_0 \text{ doğru}) = P(I.\text{tip hata}) \\ \beta &= P(H_0 \text{ kabul}/H_0 \text{ yanlış}) = P(II.\text{tip hata}) \end{aligned}$$

dir.

8.1.1 Ortalamalarla İlgili Hipotez Testleri

Bu testlerde hipotez ve alternatifler aşağıdaki gibi belirlenir:

Çift yönlü test	**Tekyönlü test**		
$H_0 : \mu = A$	$H_0 : \mu = A$	veya	$H_0 : \mu = A$
$H_1 : \mu \neq A$	$H_1 : \mu > A$		$H_1 : \mu < A$

Burada A hipotezde iddia edilen değerdir. Bu tür hipotez testlerinde esas olan z değeri,

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{8.2}$$

eşitliği ile tanımlanır. Bu z değeri kabul bölgesine düşerse hipotez kabul, aksi halde reddedilir.

Ortalamalarla ilgili test istatistigi ve kritik bölgeler aşağıdaki tabloda verilmiştir.

Hipotez	Karşıt Hipotez	Test İstatistiği	Kritik Bölge
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$	$Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n-1})$,	$Z > z_{1-\alpha/2}$, $Z < -z_{1-\alpha/2}$
$H_0 : \mu = \mu_0$	$H_1 : \mu > \mu_0$	$t = (\bar{X} - \mu_0)/(\sigma/\sqrt{n-1})$,	$Z > z_{1-\alpha}$
$H_0 : \mu = \mu_0$	$H_1 : \mu < \mu_0$	$t = (\bar{X} - \mu_0)/(\sigma/\sqrt{n-1})$,	$Z < -z_{1-\alpha}$

Örnek 8.1 *Bir pil fabrikasında ürünlerin ortalama dayanma süresinin* 200 *saat olduğu iddia ediliyor. Sözkonusu iddianın testi için* $n = 36$ *büyüklüğünde bir örnek alınıyor ve ortalama dayanma süresi* 198 *saat, standart sapma ise* 6 *saat olarak tesbit ediliyor. Örnekten elde edilen bu bilgilere göre söz konusu hipotezi*

i. 0.05 *önem seviyesinde,*

ii. 0.01 *önem seviyesinde*

test ediniz.

Çözüm 8.1 *i. Tek yönlü test yapılacaktır. Buna göre pillerin ortalama dayanma süresi* 200 *saatten az olursa hipotez reddedilecektir.*

$H_0 : \mu = 200$ *(Ortalama ömür* 200 *saattir.)*
$H_1 : \mu < 200$ *(Ortalama ömür* 200 *saatten azdır.)*
$\alpha = 0.05$

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{198 - 200}{6/\sqrt{36}} = -2$$

$-2 < -1.645$ *olduğu için z değeri red bölgesine düşmüştür. Yani hipotez reddedilir. Dolayısıyla pillerin ortalama dayanma süresinin* 200 *saatten az olduğuna* 0.05 *önem seviyesinde karar verilir.*

ii. $-2 > -2.33$ *olduğundan* 0.01 *önem seviyesinde hipotez kabul edilir. Pillerin ortalama dayanma sürelerinin* 200 *saat olduğuna* 0.01 *önem seviyesinde karar verilir.*

8.1.2 Oranlarla İlgili Hipotez Testleri

Oranlarla ilgili testlerde hipotez ve alternatif hipotezler aşağıdaki gibi belirlenir:

Çift yönlü test	**Tekyönlü test**	
$H_0 : P = a$	$H_0 : P = a$ veya	$H_0 : P = a$
$H_1 : P \neq a$	$H_1 : P > a$	$H_1 : P < a$

Burada a hipotezde iddia edilen değerdir. Bu testler için z değeri,

$$z = \frac{p - P}{PQ/n} \tag{8.3}$$

eşitliği ile hesaplanır. Bu z değeri kabul bölgesine düşerse hipotez kabul, aksi halde reddedilir.

Örnek 8.2 *Bir üretici ürettiği mallarda kusurlu oranının en fazla* %3 *olduğunu iddia ediyor. Bu ürünlerden alınan* 40 *büyüklüğündeki bir şans örneğinde kusurlu oranı* %4 *olarak belirleniyor. Bu bilgilere göre söz konusu hipotezi* 0.01 *önem seviyesinde test ediniz.*

Çözüm 8.2

$$P = 0.03, \quad n = 40, \quad p = 0.04, \quad \alpha = 0.01$$

$$\begin{aligned} H_0 &: P = 0.03 \quad \text{(Kusurlu oranı 0.03 tür.)} \\ H_1 &: P > 0.03 \quad \text{(Kusurlu oranı 0.03 ten fazladır.)} \end{aligned}$$

$$z = \frac{p - P}{PQ/n} = \frac{0.04 - 0.03}{(0.03)(0.97)/40} = \frac{0.01}{0.03} = 0.33$$

$0.33 < 2.33$ *olduğundan* H_0 *hipotezi kabul edilir. Yani, kusurlu oranının* 0.03*'ten fazla olmadığına* 0.01 *önem seviyesinde karar verilir.*

8.1.3 Ortalamalar Arası Farklarla İlgili Hipotez Testleri

Bu testlerde hipotez ve alternatifler aşağıdaki gibi belirlenir:

Çift yönlü test	**Tekyönlü test**	
$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2$ veya	$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$

Bu test için kullanılacak z değeri ise,

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{8.4}$$

formülüyle hesaplanır.

Hipotez $\mu_1 = \mu_2$ şeklinde kurulduğunda kesrin payındaki $\mu_1 - \mu_2$ farkı sıfırdı. Bu durumda yine hipotez, "örneklerin aynı kitleden alındığı" şeklinde kurulduğu için σ_1 ve σ_2 değerleri yerine tek bir σ^2 değeri kullanmak yerinde olur. Bu ortak varyans, her iki varyansın ağırlıklı aritmetik ortalaması alınarak hesaplanır. Yani,

$$\sigma^2 = \frac{n_1\sigma^2 + n_2\sigma^2}{n_1 + n_2} \tag{8.5}$$

dir. Buna göre, testte esas alınacak z değeri,

$$z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{8.6}$$

şeklini alır.

Örnek 8.3 *A ve B sınıflarında başarı durumunun eşit olduğu iddia ediliyor. Bu iddiayı test için A sınıfından alınan* 25 *kişilik bir şans örneğinde not ortalaması* 6 *ve standart sapma* 0.9*; B sınıfından alınan* 20 *kişilik bir şans örneğinde ise not ortalaması* 7 *ve standart sapma* 1.2 *bulunuyor. Bu bilgilere göre not ortalamasının her iki sınıfta aynı olduğu hipotezini* 0.05 *önem seviyesinde test ediniz.*

Çözüm 8.3 *Test çift yönlüdür. Çünkü her iki sınıfın başarı oranlarının eşit olup olmaması sözkonusudur.*

$n_1 = 25 \quad x_1 = 6 \quad s_1 = 0.9$
$n_2 = 20 \quad x_2 = 7 \quad s_2 = 1.2$

$H_0 : \mu_1 = \mu_2$ *(Başarı oranları eşittir.)*
$H_1 : \mu_1 \neq \mu_2$ *(Başarı oranları farklıdır.)*
$\alpha = 0.05$

$$\sigma = \sqrt{\frac{n_1\sigma^2 + n_2\sigma^2}{n_1 + n_2}} = \sqrt{\frac{25(0.9^2) + 20(1.2^2)}{25 + 20}} = 1.02$$

Buradan da

$$z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(6-7)}{1.02\sqrt{\frac{1}{25} + \frac{1}{20}}} = -\frac{1}{0.306} = -3.26$$

bulunur. $-3.26 < -1.96$ *olduğundan* H_0 *hipotezi reddedilir. Yani, A ve B sınıflarındaki başarı oranlarının farklı olduğuna* 0.05 *önem seviyesinde karar verilebilir.*

8.1.4 Oranlar Arası Farklarla İlgili Hipotez Testleri

Bu testlerde hipotez ve alternatifler aşağıdaki gibi belirlenir:

Çift yönlü test	**Tekyönlü test**		
$H_0 : P_1 = P_2$	$H_0 : P_1 = P_2$	veya	$H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$	$H_1 : P_1 > P_2$		$H_1 : P_1 < P_2$

Bu test için kullanılacak z değeri ise,

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \tag{8.7}$$

eşitliği ile hesaplanır.

Hipotez $P_1 = P_2$ şeklinde kurulduğunda kesrin payındaki $P_1 - P_2$ farkı sıfırdır ve yine hipotez, "örneklerin aynı kitleden alındığı" şeklinde kurulduğu için P_1 ve P_2 değerleri yerine tek bir P değeri kullanılacaktır. Bu değer her iki oranın ağırlıklı aritmetik ortalaması alınmak suretiyle yani,

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \tag{8.8}$$

şeklinde hesaplanır. Buna göre, testte esas alınacak z değeri

$$z = \frac{(p_1 - p_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{8.9}$$

şeklini alır.

Örnek 8.4 *Herbiri* 100 *kişiden meydana gelen iki grup insan belli bir hastalığa tutuluyor. Hastalığın tedavisinde A grubuna serum verildiği halde B grubuna verilmiyor. Sonunda A grubundan* 75*, B grubundan* 65 *kişi hastalıktan kurtulduğuna göre verilen serumun hastalığın tedavisine yardım edip etmediğini* 0.01 *önem seviyesinde test ediniz.*

Çözüm 8.4 $p = 65/100 = 0.75 \quad p_2 = 75/100 = 0.65$

$H_0 : P_1 = P_2$ *(Görülen farklılık şanstan ileri gelmiştir. Serum tesirsizdir.)*

$H_1 : P_1 > P_2$ *(Serum tesirlidir.)*

$\alpha = 0.01$

$$\begin{aligned} P &= \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{100(0.75) + 100(0.65)}{100 + 100} = 0.70 \\ Q &= 1 - P = 1 - 0.70 = 0.30 \end{aligned}$$

$$z = \frac{(p_1 - p_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.75 - 0.65)}{\sqrt{(0.7)(0.3)\left(\frac{1}{100} + \frac{1}{100}\right)}} = \frac{0.10}{0.064} = 1.54$$

bulunur. $1.54 < 2.33$ *olduğundan* 0.01 *önem seviyesinde* H_0 *kabul edilir. Yani, görülen farklılık şanstan ileri gelmiştir. Serum tesirsizdir.*

8.2 t Dağılımı ve t Testi

Normal eğri alanı tablosu (z cetveli) ancak büyük örnekler için (yani $n \geq 30$ için) gerçekci sonuçlar verebilir. Örnek büyüklüğü 30'dan daha küçük olan problemlerde, örnek değerleri normal dağılımdan uzaklaşırlar. Eğri daha basık ve yaygın bir şekil alır. n küçüldükçe normalden uzaklaşma da o nisbette artar. Aşağıdaki grafiklerde bunu açıkca görmek mümkündür.

Şekil 8.1:

Örneğin, $n = 2$ için, X'in X_1 ile X_2 arasında bulunma olasılığı, gerçekte, b grafiğindeki taralı alan kadar olduğu halde, z cetveli bu olasılığı olduğundan daha büyük olarak verecektir. Bu durum dikkate alınarak z cetveli yerine, çeşitli örnek büyüklükleri ve olasılık seviyeleri için ayrı ayrı hesaplanmış, t cetvelleri kullanılır. Bu cetvellerde, çeşitli önem seviyeleri ve $(n-1)$ serbestlikdereceleri için ayrı ayrı t değerleri verilmiştir. $n = 30$ için, t değeri, z değerine çok yaklaşır. Bu sebeble, $n \geq 30$ olan örneklerde, t cetveli yerine z cetveli kullanılabilir.

Uygulamada genellikle σ bilinmediğinden, yerine onun bir tahmini olan s kullanılır. s ile yapılan σ tahmini, gerçeğin altında olma eğilimi gösterdiğinden,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{8.10}$$

eşitliğinde, kesrin paydasının da küçültülmesi yoluna gidilir ve n yerine $n-1$ kullanılır. Böylece ortalamaların örnek dağılımının standart hatası,

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n-1}} \tag{8.11}$$

şeklinde tahmin edilmiş olur. Bu tahmin, gerçeği dah iyi yansıtır. Buna göre,

$n < 30$ olan örneklerde,

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{8.12}$$

yerine,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} \tag{8.13}$$

değeri kullanılır. Dolayısıyla, küçük örneklerde, kitle ortalamasının güven sınırları,

$$\text{K.O.G.S} = \bar{X} \pm t\frac{s}{\sqrt{n-1}} \tag{8.14}$$

Kitle oranının sınırları ise,

$$\text{K.O.S} = p \pm t\sigma_p \tag{8.15}$$

şeklinde hesaplanır. t testi ile ilgili test istatistiği ve kritik bölgeler aşağıdaki tabloda verilmiştir.

Hipotez	Karşıt Hipotez	Test İstatistiği	Kritik Bölge
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$	$t = (\bar{X} - \mu_0)/(s/\sqrt{n-1})$,	$t > t_{1-\alpha/2}$, $t < -t_{1-\alpha/2}$
$H_0 : \mu = \mu_0$	$H_1 : \mu > \mu_0$	$t = (\bar{X} - \mu_0)/(s/\sqrt{n-1})$,	$t > t_{1-\alpha}$
$H_0 : \mu = \mu_0$	$H_1 : \mu < \mu_0$	$t = (\bar{X} - \mu_0)/(s/\sqrt{n-1})$,	$t < -t_{1-\alpha}$

Örnek 8.5 *Bir makina,* $1.27cm$ *kalınlığında vida pulları üretmektedir. Makinanın normal çalışma düzeninde olup olmadığını belirlemek için* 10 *vida pulundan meydana gelen bir örnek seçilmiş ve bu örnekte ortalama kalınlık* $1.35cm$ *ve standart sapma* $0.08cm$ *bulunmuştur. Makinanın normal çalışma düzeninde olduğu hipotezini,*

i. 0.05 *önem seviyesinde,*

ii. 0.01 *önem seiyesinde*

test ediniz.

Çözüm 8.5 *Önce hipotezleri ifade edelim.*

$H_0 : \mu = 1.27,$ *makina normal çalışma düzenindedir.*
$H_1 : \mu \neq 1.27,$ *makina normal çalışmıyor.*

Alternatif hipotezden anlaşılacağı gibi, çift yönlü test uygulanacaktır. Örnek büyüklüğü $n = 10 < 30$ *olduğundan* t *testine göre karar verilmesi gerekir.*

i. 0.05 *önem seviyesi için karar modeli şöyle oluşturulur.* $n-1 = 10-1 = 9$ *serbestlik derecesi için kritik değerler:* $t_{1-\frac{\alpha}{2}} = t_{0.975;9} = \pm 2.26$ *dır.*

Şekil 8.2:

Bulunacak t *değeri için,* $-2.26 < t < 2.26$ *olması halinde* H_0 *hipotezi kabul edilecek, aksi halde reddedilecektir.*

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} = \frac{1.35 - 1.27}{0.08/\sqrt{10-1}} = 3$$

$3 < 2.26$ *olduğundan* H_0 *reddedilir. Yani, makinanın normal çalışma düzeninde olmadığına* 0.05 *önem seviyesinde karar verilir.*

ii. 0.01 *önem seviyesi için kritik değerler,* $t_{1-\frac{0.01}{2}} = t_{0.995;9} = \pm 3.25$ *tir.*

Şekil 8.3:

$-3.25 < 3 < 3.25$ *olduğundan* H_0 *kabul edilir. Yani, makinanın normal çalışma düzeninde olduğuna* 0.01 *önem seviyesinde karar verilir.*

Örnek 8.6 *Bir şirket tarafından üretilen ampüllerin ortalama dayanma süresi,* 1120 *saat ve standart sapma* 125 *saat olarak belirlenmiştir. Yeni üretilen ampüllerden,* 8 *ampüllük bir örnek seçilmiş ve bu örnekte ortalama dayanma süresi* 1070 *saat bulunmuştur. Ampüllerin ortalama dayanma sürelerinin değişmediği hipotezini:*

i. 0.05 *önem seviyesinde,*

ii. 0.01 *önem seviyesinde*

test ediniz.

Çözüm 8.6

$H_0 : \mu = 1120$

$H_1 : \mu \neq 1120$ *(Çift yönlü test)*

$n = 8 < 30$ *olduğundan t dağılımı esas alınacaktır.*

i. 0.05 *önem seviyesi için, Serbestlik derecesi* $= n - 1 = 8 - 1 = 7$, *Kritik değerler,* $t_{0.975;7} = \pm 2.36$ *olur.*

Şekil 8.4:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} = \frac{1070 - 1120}{125/\sqrt{8-1}} = 1.06$$

$-2.36 < 1.06 < 2.36$ *olduğundan sıfır hipotezi kabul edilir. Yani, ampüllerin ortalama dayanma süresinin değişmediğine* 0.05 *önem seviyesinde karar verilir.*

ii. 0.01 *önem seviyesi için,* $t_{0.995;7} = 3.50$ *bulunur (t cetvelinden.)* $-3.50 < 1.06 < 3.50$ *olduğundan, bu önem seviyesi içinde aynı karar verilir.*

8.3 χ^2 (Ki-Kare) Testi

Örnekleme yoluyla elde edilen rakamların, kitle rakamlarına uygun olup olmadığı, bir başka ifadeyle, gerçek değerlerin, teorik değerlerle uygunluk gösterip göstermediği, ki-kare testi ile belirlenir.

χ^2 testi ilk defa, 1900 yılında Karl Pearson tarafından bulunmuş ve uygulamaya konulmuştur.

χ^2 dağılımı sağa çarpıktır ve normalden daha diktir. n büyüdükçe diklik ve asimetri azalır ve dağılım normale yaklaşır.

Şekil 8.5:

χ^2 değerleri için alt sınır 0 dır. Bu ise, bütün gerçek değerlerin, beklenen değerlerle aynı olması, %100 uyuşmsı demektir. Bu sebeple, eğri sıfır noktasından başlar ve sonsuza uzanır.

o_i gerçek (gözlenen frekanslar) e_i teorik (beklenen) değerler olmak üzere, χ^2 değerleri:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_k - e_k)^2}{e_k} \tag{8.16}$$

eşitliği ile hesaplanır. $\sum o_i = \sum e_i = N$ olması halinde yukarıdaki eşitlikten,

$$\chi^2 = \sum \frac{o_i^2}{e_1} - N \tag{8.17}$$

formülü elde edilir. Her iki eşitlikte aynı sonucu verir.

Aynı miktar bir ayrılışın, küçük sayılarda büyük sayılardan daha önemli olduğu dikkate alınarak, (8.16) nolu eşitlikte $(o_i - e_i)$ farkının karesi e_i ye bölünmüş, yani, e_i nin yüzdesi olarak ifade edilmiştir. Örneğin, teorik olarak 20 olması gereken bir sayının 10 çıkması ve yine teoride 1000 olması gereken bir diğer sayınında 990 çıkması halinde, her iki durumda da beklenen ve gerçek değer arasındaki fark 10 olmakla beraber, bu farkın birinci halde çok daha önem kazandığı açıktır.

Uygulamada, gerçek frekanslarla teorik frekanslar arasında az çok bir farklılık görülür. Diğer taraftan, değişik örnek büyüklükleri ile elde edilen ki-kare değerleri de birbirinden farklı çıkarlar. Bu değerlerin ana kütle değerlerinden önemli ölçüde sapma gösterip göstermediğini, yani bu sapmaların şansa bırakılıp bırakılmayacağını ortaya koymak için ki-kare tabloları hazırlanmıştır. 0.01 ve 0.05 önem seviyelerinde her örnek büyüklüğü için ayrı ayrı ki-kare değerleri hesaplanmış ve bu tablolarda gösterilmiştir. Tablo değerleri, "son tolerans sınırı" nı ifade eder. Yani bu değerlerden daha büyük çıkan ki-kare değerleri için, gerçek değerlerle teorik değerler arasındaki farklılığın, istatistik olarak önemli bulunduğu ve şansa bırakılmasının mümkün olmayacağı ifade edilir.

Yates Düzeltmesi

Ki-kare testinin kesikli dağılımlara uygulanması halinde $o_i - e_i$ farkından ayrıca 0.5'i çıkarmak, kesikli verileri böylece sürekli yapmak gerekir. Bununla beraber, serbestlik derecesinin 1 den büyük olduğu problemlerde, düzeltilmiş ve düzeltilmemiş veriler, yaklaşık sonuçlar verdiğinden, düzeltmeğe lüzum kalmaz. Serbestlik derecesi, $\nu = 1$ olduğunda ise aşağıdaki düzeltme yapılır:

$$\chi^2 = \sum \frac{(|o_i - e_i| - 0.5)^2}{e_i} \tag{8.18}$$

Bu düzeltmeye "Yates düzeltmesi" denilir.

Uygunluk Tabloları

Gerçek değerlerin sadece bir sıra (veya sütun) olmakla kalmayıp, yeni bir tablo oluşturduğu durumlarda, bunun zaruri bir sonucu olarak, beklenen değerler de bir tablo halinde oluşturulur. Bu tablonun birbirine uygunluğunu, ki-kare ile test etmek istediğimizde, esas olarak yine,

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \tag{8.19}$$

formülü kullanılır. Bu tablolarda beklenen değerlerin nasıl hesaplandığı, 2×3 tablosu için aşağıda gösterilmiştir.

	I	II	III	Toplam
A	a_1	a_2	a_3	N_A
B	b_1	b_2	b_3	N_B
Toplam	N_1	N_2	N_3	N

	I	II	III	Toplam
A	$N_1 \frac{N_A}{N}$	$N_2 \frac{N_A}{N}$	$N_3 \frac{N_A}{N}$	N_A
B	$N_1 \frac{N_B}{N}$	$N_2 \frac{N_B}{N}$	$N_3 \frac{N_B}{N}$	N_B
	N_1	N_2	N_3	N

Diğer tablolar için de benzer işlemler yapılır.

Uygunluk tabloları için uygulamada hesap kolaylığı sağlayan bazı özel formüller geliştirilmiştir.

2×2 tablosu için ki-kare değeri şu eşitlikle de hesaplanabilmektedir.

$$\chi^2 = \frac{N(a_1 b_2 - a_2 b_1)^2}{(a_1 + b_1)(a_2 + b_2)(a_1 + a_2)(b_1 + b_2)} = \frac{N\delta^2}{N_1 N_2 N_A N_B} \tag{8.20}$$

2×3 tabloları için ise

$$\chi^2 = \frac{N}{N_A}\left(\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \frac{a_3^2}{N_3}\right) + \frac{N}{N_B}\left(\frac{b_1^2}{N_1} + \frac{b_2^2}{N_2} + \frac{b_3^2}{N_3}\right) - N \tag{8.21}$$

eşitliği geliştirilmiştir.

Uygunluk tablolarında serbestlik dereceleri şöyle hesaplanır: Sıra sayısı r, kolon sayısı k olmak üzere, serbestlik derecesi,

$$\nu = (r-1)(k-1)$$

dir. Örneğin, normal yoğunluk testlerinde örnek ortalaması ($\bar{X}$) ve standart sapması (s); μ ve σ nın birer tahmini olarak işleme koyduğumuzda, bu testlerde serbestlik derecesi,

$$\nu = (r-1)(k-1) - 2$$

olur.

Uygunluk Katsayısı

Ki-kare testi ile sadece gerçek ve teorik değerler arasında bir uygunluğun bulunup bulunmadığını anlayabiliriz, eğer varsa, bu uygunluğun oranını belirleyemeyiz. Bu oran "uygunluk katsayısı" ile ortaya konulur. Uygunluk katsayısı:

$$c = \sqrt{\chi^2/\chi^2 + N} \tag{8.22}$$

ile verilir.

Ki-karenin Toplanabilirlik Özelliği

Serbestlik dereceleri ν_1, ν_2, ν_3, ... olan çeşitli ki-kare değerlerinin toplanması ile bulunacak değerler de ki-kare dağılımı gösterirler. Bu dağılımın serbestlik derecesi, ki-karelere ait serbestlik derecelerinin toplamına eşittir. Yani,

$$\nu = \nu_1 + \nu_2 + \nu_3 \cdots$$

yazılabilir.

Örnek 8.7 *Dört ilde yapılan bir seçim sonunda, A, B ve C adaylarının aldıkları oy sayıları, aşağıdaki tabloda belirtilmiştir. Buna göre adayların bu dört şehirden herbirinde, aynı oranda oy aldıkları hipotezini* 0.01 *önem seviyesinde test ediniz.*

İller$\rightarrow$	*I*	*II*	*III*	*IV*	*Toplam*
A Adayı	*54*	*93*	*56*	*47*	*250*
B Adayı	*41*	*62*	*59*	*38*	*200*
C Adayı	*25*	*20*	*30*	*25*	*100*
Toplam	*120*	*175*	*145*	*110*	*550*

Çözüm 8.7 *Yukarıdaki tabloda gösterilen veriler, gözlem sonucu elde edilen değerlerdir. Beklenen değerler ise şöyle elde edilir:*

I nolu ilde adayların alması beklenen oy sayılarını bulalım.

Genelde, 550 *seçmenden* 250*'si A adayına*
200*'ü B adayına*
100*'ü C adayına oy vermiştir.*
O halde, 120 *seçmenden* $120(250/550) = 54.5$*'inin A adayına*
$120(200/550) = 43.6$*'sının B adayına*
$120(100/550) = 21.9$*'unun C adayına*

oy vermesi beklenir. Diğer iller için de aynı yol takip edilerek aşağıdaki tablo düzenlenir.

	I	II	III	IV	Topl.
A	$\frac{(120)(250)}{550} = 54.5$	$\frac{(175)(250)}{550} = 79.5$	$\frac{(145)(250)}{550} = 65.9$	$\frac{(110)(250)}{550} = 50$	250
B	$\frac{(120)(200)}{550} = 43.6$	$\frac{(175)(200)}{550} = 63.6$	$\frac{(145)(200)}{550} = 52.7$	$\frac{(110)(200)}{550} = 40$	200
C	$\frac{(120)(100)}{550} = 21.9$	$\frac{(175)(100)}{550} = 31.9$	$\frac{(145)(100)}{550} = 26.4$	$\frac{(110)(100)}{550} = 20$	100
Topl.	120	175	145	110	550

Aday sayısı ondalık sayılarla gösterilmez. Ancak, bulduğumuz bu değerler teorik değerler olduğundan, ondalık sayı kullanılabilir. Ayrıca, beklenen değerler tablosu düzenlenirken, her il için A ve B'nin oyları hesaplanarak il toplamından çıkarılıp, C'nin oy sayısı bulunmuştur. Yaklaşık değerler alındığı için, adayların beklenen toplam oy sayıları son sütundaki toplam değeri tam olarak vermeyebilir.

o_i*, gözlenen frekanslar ve* e_i*, beklenen frekanslar olmak üzere ki-kare istatistiğini aşağıdaki gibi hesaplayabiliriz:*

$$\chi^2 = \sum \frac{(o_i - e_i)}{e_i}$$

o_i	e_i	$o_i - e_i$	$(o_i - e_i)^2$	$(o_i - e_i)^2/e_i$
54	54.5	-0.5	0.25	0.005
93	79.5	13.5	185.25	2.292
56	65.9	-9.9	98.01	1.487
47	50.0	-3.0	9.00	0.180
41	43.6	-2.6	6.76	0.155
62	63.6	-1.6	2.56	0.040
59	52.7	6.3	39.69	0.753
38	40.0	-2.0	4.00	0.100
25	21.9	3.1	9.61	0.439
20	31.9	-11.9	141.61	4.439
30	26.4	3.6	12.96	0.491
25	20.0	5.0	25.00	1.250
				$\chi^2 = 11.631$

Serbestlik derecesi = (Satır sayısı -1)(Sütun sayısı-1) $= (3-1)(4-1) = 6$, $\alpha = 0.01$ *önem seviyesine göre, kritik değerler* $\chi^2_{0.99;6} = 16.8$ *dir. Bu değer tablodan* 6 *serbestlik derecesine göre bulunmuştur. Bu test için karar modeli aşağıdaki gibi olur.*

Şekil 8.6

$11.631 < 16.8$ *olduğundan, gerçek değerlerle teorik değerler arasındaki farklılığın,* 0.01 *önem seviyesinde, önemsiz olduğuna karar verilir. Yani, adayların oy oranlarının bölgelere göre farklılık göstermediği,* 0.01 *önem seviyesinde, kabul edilebilir.*

Örnek 8.8 *Herbiri* 100 *kişiden meydana gelen iki grup (A ve B) belli bir hastalığa tutuluyor. Hastalığın tedavisinde A grubundan* 75, *B grubundan* 65 *kişi hastalıktan kurtulduğuna göre, verilen serumun hastalığın tedavisinde yardım edip etmediğini* 0.05 *önem seviyesinde test ediniz.*

Çözüm 8.8 *Verileri bir tablo olarak gösterelim:*

	İyileşen hastalar	*İyileşmeyen hastalar*	*Toplam*
A grubu serum kullanan	*75*	*25*	*100*
B grubu serum kullanmaz	*65*	*35*	*100*
Toplam	*140*	*60*	*200*

Bu tablo gözlenen frekansları göstermektedir. Beklenen frekanslar ise aşağıdaki gibi olur.

	İyileşen hastalar	*İyileşmeyen hastalar*	*Toplam*
A	$\frac{(140)(100)}{200} = 70$	$\frac{(60)(100)}{200} = 30$	*100*
B	$\frac{(140)(100)}{200} = 70$	$\frac{(60)(100)}{200} = 30$	*100*
Toplam	*140*	*60*	*200*

Sebestlik derecesi $= (2-1)(2-1) = 1$, $\chi^2_{0.95;1} = 3.84$ *olur. Serbestlik derecesi* 1 *olduğu için Yates düzeltmesi yapılmalıdır. Yates düzeltmesi için formül,*

$$\chi^2 = \sum \frac{(|o_i - e_i| - 0.5)^2}{e_i}$$

şeklindeydi. Bu formül kullanılarak χ^2 *değeri,*

$$\chi^2 = \frac{(|75-70|-0.5)^2}{70} + \frac{(|65-70|-0.5)^2}{70} + \frac{(|25-30|-0.5)^2}{30} + \frac{(|35-30|-0.5)^2}{30} = 1.93$$

şeklinde elde edilir.

$1.93 < 3.84$ *olduğu için kullanılan serumun hastalığın tedavisine yardım etmediği söylenebilir.*

8.3.1 Problemler

1. *A* fabrikasında üretilen pillerin ortalama ömrünün 210 saat, *B* fabrikasının pillerinde ise bu süre 204 saat olduğu iddia ediliyor. Bu iki

marka pil arasında ömür süresi yönünden bir farklılık olup olmadığını test etmek gayesiyle A fabrikasından 30 birimlik bir şans örneği alınıyor ve ortalama ömrü 200 saat, standart sapması 7 saat bulunuyor. B fabrikasından ise 50 birimlik bir şans örneği alınıyor ve ortalama ömrü 212 saat, standart sapma ise 9 saat olarak tesbit ediliyor. Bu bilgilere göre, iddiayı 0.01 önem seviyesinde destekleyebilir miyiz?

2. Bir basketbol oyuncusu hata atışlarının %60'ında sayı elde ettiğini söylemektedir. Bu oyuncu bir mevsim boyunca yaptığı 100 atıştan 70'inde sayı elde etmiştir. Oyuncunun hata atışlarında $\alpha = 0.05$ önem seviyesinde gelişme olup olmadığını test ediniz?

3. Tıbbi bir denemede fareler özel bir gıda rejimi ile beslenerek 2 aylık bir zaman sonunda 40 gramlık ortalama ağırlık artışı görüldüğü bilinmektedir. Yeni bir diet 16 farelik bir örnek üzerinde denenmiş ve ortalama ağırlık artışının 43 gram ve varyansının 16 gram olduğu belirlenmiştir. $\alpha = 0.01$ alınarak yeni dietin ağırlığı arttırdığı hipotezini test ediniz.

4. Kabul edilen yeni bir öğretim metodu aynı yaşlarda 25 çocuk üzerinde denenip, bunların verilen bir metni ortalama olarak 12 dakşkada öğrendikleri belirleniyor. Ayrıca halen kullanılmakta olan metodla aynı yaşlardaki çocukların aynı metni ortalama olarak 13.4 dakikada öğrendikleri ve standart sapmanın 2.5 dakika olduğu bilinmektedir. $\alpha = 0.01$ alınarak yeni metodun, kullanılmakta olan metoddan daha etkin olduğunu söyleyebilir miyiz?

5. Bir tarım istasyonunda, belli bir gübrenin buğday üretimi üzerindeki tesiri incelenmek istenmiştir. Bu amaçla eşit büyüklükte 24 parsel seçilerek, bu tarlaların yarısı gübrelenmiş, yarısı ise gübrelenmemiştir. Diğer şartlar sabittir. Gübrelenen parselde ortalama verim 5.1 kile ve standart sapma 0.36 kile; gübrelenmeyen parsellerde ise ortalama verim 4.8 kile ve standart sapma 0.40 kile olarak belirlenmiştir. Gübreleme dolayısıyla buğday üretiminin arttığına 0.01 önem seviyesinde karar verebilir miyiz?

6. Bir fabrikada belirli çaplarda demir boru üretilmektedir. Kontrol etmek amacıyla rastgele seçilen 400 borudan 12 sinde iç çapların önceden

belirlenmiş sınırların içinde olmadığı belirlenmiştir. Örnek sonucuna dayanarak, üretim süresinin %2 den daha çok kusurlu boru verdiği sonucu çıkarılabilir mi?($\alpha = 0.05$ alınız.)

7. Bir zarın 120 kez atılmasıyla aşağıdaki sonuçlar elde edilmiştir.

Üste gelen yüzler	1	2	3	4	5	6
Frekanlar	25	15	12	26	15	27

$\alpha = 0.05$ önem seviyesinde, zarın hileli olup olmadığını test ediniz.

8. Bir sekreterin daktilo ettiği 100 sayfalık yazıda, sayfalarda yaptığı hataların sıklık dağılımı şöyle bulunmuştur.

Bu sayfadaki hata sayısı(x_i)	0	1	2	3	4	5
Gözlenen sayfa sayısı	42	33	14	6	4	1

dağılımın normale uygunluğunu 0.05 önem seviyesinde test ediniz.

9. Bir şehirde çeşitli meslek gruplarına ait 500 kişi üzerinde yapılan bir anket sonunda aşağıdaki bilgiler elde edilmiştir.

	Sanayiciler	Tüccarlar	Çiftçiler	Toplam
Faaliyeti artanlar	65	95	90	250
Faaliyeti düşenler	35	45	55	135
F.değişimi farkedilmeyen	30	40	45	115
Toplam	130	180	190	500

Bu üç sınıfa karşılık gelen oranların dağılımının, üç meslek grubu için de aynı olduğu hipotezini 0.05 önem seviyesinde test ediniz.

10. 60 kişilik bir sınıfta öğrencilerin boy dağılımı aşağıdaki gibi belirlenmiştir. Dağılımın normale uygunluğunu 0.01 önem seviyesinde test

ediniz. ($\bar{X} = 163$ ve $s = 7.91$ dir.)

Öğrencilerin boyları	Frekanslar
145-150	3
150-155	8
155-160	10
160-165	12
165-170	17
170-175	5
175-180	5

11. Bir fabrikanın ürettiği malların %10'unun iyi, %70'inin orta, %20'sinin de kötü kaliteli olduğu belirtilmiştir. Bu mallardan şansa bağlı olarak alınan 30 birimlik bir örnekte kalite dağılımı aşağıdaki gibi elde edilmiştir. Buna göre bu firma için belirtilen kalite oranlarını 0.05 önem seviyesinde destekleyebilir miyiz?

Kalite	İyi	Orta	Kötü
Mal sayısı	2	21	7

12. Bir fabrikanın ürettiği pillerin%40'ının bozuk olduğu iddia edilmiştir. Bir kontrol heyetince 4'er birimlik 50 örnek teste tabi tutulmuş ve aşağıdaki dağılım elde edilmiştir.

Kusurlu sayısı	0	1	2	3	4
Örnek sayısı	10	12	15	7	6

Bu dağılımın binoma uygunluğunu, 0.01 önem seviyesinde test ediniz ve iddianın desteklenip destenlenmeyeceğine karar veriniz.

13. $E5$ karayolunda, saat $8 - 9$ arası ortalama 2 trafik kazası olduğu belirlenmiştir. Kazaların Poisson dağılımı gösterdiğini belirlemek amacı ile, birbirini takip eden 100 gün içerisinde söz konusu saatlerde meydana gelen kaza sayısı tesbit edilerek aşağıdaki tablo düzenlenmiştir. Buna göre dağılımın Poissona uygun olup olmadığını 0.05 önem seviyesinde test ediniz.

Kaza sayısı	0	1	2	3	3'den çok
Gün sayısı	10	30	35	15	10

Chapter 9

VARYANS ANALİZİ

9.1 Temel Kavramlar ve Kitle Ortalamalarının Eşitliğinin Testi

Varyans analizinde temel amaç, ikiden fazla örnek için sapmaların kareler toplamını, bu sapmalara sebep olan nedenler itibariyle, kısımlara ayırmak ve analiz etmektir.

Bu analiz sonunda, örnekler arasında uygunluk olup olmadığı, yani söz konusu örneklerin aynı kitleye ait birer şans örneği olup olmadıkları da ortaya konulmuş olur.

İki örnek ortalaması arasındaki farkın önem kontrolü, örneğin büyüklüklerine göre z veya t testlerinden biriyle yapılır. Bu testlerle ikiden fazla örnek ortalamasını birlikte test etmek ve aralarındaki farkın önem kontrolünü yapmak mümkün değildir. Bu durumda varyans analizine baş vurulur.

$$\sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2 \tag{9.1}$$

değerinin, yani örneklerdeki bütün X_{ij} değerlerinin genel ortalamadan gösterdikleri sapmaların kareler toplmının iki kaynağı vardır. Birincisi, örnek ortalamalarının genel ortalamadan gösterdikleri sapmalar, diğeri ise herbir örnekteki değerlerin kendi örnek ortalamalarından gösterdiği farklılıklardır. Bunlardan birincisi "gruplar arası kareler toplamıdır (G.A.K.T)" ve

$$n \sum (\bar{X}_i - \bar{\bar{X}})^2 \tag{9.2}$$

şeklinde hesaplanır. Diğer ise, "gruplar içi kareler toplamı (G.İ.K.T)"

$$\sum_i \sum_j (X_{ij} - \bar{X}_i)^2 \tag{9.3}$$

olarak ifade edilir. Buna göre,

$$\text{Genel Kareler Toplamı} = \text{G.A.K.T} + \text{G.İ.K.T}$$

veya

$$\sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2 \; = \; n \sum (\bar{X}_i - \bar{\bar{X}})^2 + \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 \tag{9.4}$$

yazılabilir. Bu eşitlikle üç varyasyon kaynağının herbirinin uygun bir serbestlik derecesine bölünmesiyle birer varyans elde edilir.

$$\frac{\sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2}{kn - 1}, \; \frac{n \sum (\bar{X}_i - \bar{\bar{X}})^2}{k - 1}, \; \frac{\sum_i \sum_j (X_{ij} - \bar{X}_i)^2}{k(n - 1)} \tag{9.5}$$

Burada, k örnek sayısı, n ise örnek büyüklüğü olmak üzere, genel kareler toplamı için serbestlik derecesi, $kn - 1$ dir. Gruplar arası kareler toplamının serbestlik derecesi örnek sayısının bir noksanı, yani, $k - 1$ olarak ifade edilir. Gruplar içi kareler toplamının serbestlik derecesi ise, $(kn - 1) - (k - 1) = k(n - 1)$ olur, yani her örneğin serbestlik derecesi $(n - 1)$ örnek sayısıyla çarpılmıştır. Bunlardan birincisi genel kareler ortalaması, ikincisi gruplar arası kareler ortalaması, üçüncüsü ise gruplar içi kareler ortalaması veya hata kareler ortalamasıdır.

Varyans analizinde gruplar arası kareler ortalaması olan S_1^2, gruplar içi kareler ortalaması olan S_2^2'ye bölünerek bir F değeri elde edilir. S_1^2 örnek ortalamaları arasındaki farklılığın göstergesidir. S_2^2 ise şansa bağlı farklılaşmayı ifade eder. Eğer analize tabi tutulan örnekler aynı kitleye ait şans örnekleri olmuş olsalar, $S_1^2/S_2^2 = 1$ olması gerekir. S_1^2/S_2^2 1'den ne kadar büyük çıkarsa, örnek ortalamaları arasındaki farklılıkta o derece önemlidir.

Çeşitli önem seviyeleri ve örnek büyüklükleri için S_1^2/S_2^2 nin hangi noktaya kadar şansa verilebileceği, hangi noktadan sonra önemli kabul edilerek örneklerin farklı kitlelere ait olduklarına karar verilebileceği, F cetvelleriyle tesbit edilmiştir. Yani, bizim bulduğumuz F değeri cetveldeki değerden küçükse örnek ortalamaları arasındaki farklılık şansa bağlıdır ve örnekler, aynı kitleye aittir. Bulduğumuz değer cetvel değerinden büyükse örnek ortalamaları arasındaki farklılığın önemli olduğuna ve bu örneklerin farklı kitlelere ait olduklarına karar verilir.

F dağılımı sağa çarpıktır. F, negatif değer alamayacağından, test tek yönlüdür. Red bölgesi, eğrinin sadece sağ ucunda yer alır.

Şekil 9.1:

Buraya kadar olan açıklamalarımızı, $F = S_1^2/S_2^2$ olmak üzere, bir tablo halinde aşağıdaki şekilde özetleyebiliriz;

Değişme	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması
Gruplar arası	$k-1$	G.A.K.T	S_1^2
Gruplar içi	$k(n-1)$	G.İ.K.T	S_2^2
Toplam	$nk-1$	G.K.T	

Kareler Toplamlarının Hesaplanmasında Kısa Formüller

Gruplar arası kareler toplamı;

$$\text{G.A.K.T.} = \sum_{i=1}^{k} \frac{(\sum X_i)^2}{n} - \frac{(\sum\sum X_{ij})^2}{kn}$$

$$(\sum X_i)^2 = \text{Grupların toplamlarının kareler toplamı} \quad (9.6)$$

$$(\sum\sum X_{ij})^2 = \text{Bütün birimlerin toplamının karesi}$$

Gruplar içi kareler toplamı;

$$\text{G.İ.K.T.} \quad = \quad \sum\sum X_{ij}^2 - \sum_{i=1}^{k} \frac{(\sum X_i)^2}{n} \tag{9.7}$$

olarak bulunur. Burada,

$\sum\sum X_{ij}^2$ bütün birimlerin karelerinin toplamıdır.

Genel kareler toplamı;

$$\text{G.K.T.} \quad = \quad \sum\sum X_{ij}^2 - \sum_{i=1}^{k} \frac{(\sum\sum X_{ij})^2}{nk} \tag{9.8}$$

olur.

Daha ileri analizlerde gruplar arası kareler toplamı da kendisini belirleyen faktörler ve bu faktörlerin birlikte bulunmalarından ortaya çıkan interaksiyonlar yönünden ayrı ayrı analiz edilir.

Varyans analizi, regresyon önem kontrollerinde de kullanılır. Bu konu regresyon bölümünde tekrar ele alınacaktır. Yeri gelmişken şu önemli hususu da belirtmekte fayda vardır. Varyans analizi örnek ortalamaları arasında herhangi bir farklılık bulunup bulunmadığını ortaya koymakla beraber bu farklılığa hangi örnek ortalamasının(veya ortalamalarının) sebep olduğu hususunda bir fikir vermez. Farklılık gösteren grupların belirlenmesi için ayrı metodlar geliştirilmiştir. Bunlara burada değinmeyeceğiz.

Örnek 9.1 *Bir fakültenin İktisat, İşletme, Pazarlama ve Yönetim bölümlerinden mezun olan üçer kişi, müfettişlik sınavına giriyor ve aşağıdaki notları alıyorlar. Başarı durumunun bölümlere göre farklılık götermediğini* 0.05 *önem seviyesinde test ediniz.*

İktisat	*İşletme*	*Pazarlama*	*Yönetim*
7	*5*	*7*	*3*
8	*6*	*7*	*6*
4	*6*	*4*	*5*

Çözüm 9.1 *Bu problemde, ikiden fazla ortalamaların karşılaştırılması söz konusu olduğundan, varyans analizi yapılacaktır. Kullanılacak formüller şunlardır:*

$$\begin{aligned} G.K.T &= \sum^k \sum^n X_{ij}^2 - \frac{(\sum\sum X_{ij})^2}{nk} \\ G.A.K.T &= \sum^k \frac{(\sum X_i)^2}{n} - \frac{(\sum\sum X_{ij})^2}{nk} \\ G.\dot{I}.K.T &= G.K.T - G.A.K.T \end{aligned} \tag{9.9}$$

$\sum^k \sum^n X_{ij}$:	*Bütün birimlerin toplamı (genel toplam)*
$\sum^k \sum^n X^2$:	*Herbir birimin karelerinin toplamı*
$(\sum X_i)^2$:	*Grup toplamlarının kareleri*
n :	*Her gruptaki birim sayısı (örnek büyüklüğü)*
k :	*Grup sayısı*
nk :	*Toplam birim sayısıdır.*

Gruplar; iktisat, işletme, pazarlama ve yönetim bölümleridir. Bunları sırasıyla: I, II, III, IV rakamları ile gösterelim.

	I	II	III	IV
	7	5	7	3
	8	6	7	6
	4	6	4	5
$\sum X_i \rightarrow$	19	17	18	14
$(\sum X_i)^2 \rightarrow$	19^2	17^2	18^2	14^2

$\sum\sum X_{ij} = 19 + 17 + 18 + 14 = 68,$

$$\sum \frac{(\sum X_i)^2}{n} = \frac{19^2 + 17^2 + 18^2 + 14^2}{3} = \frac{1170}{3} = 390$$

$n = 3;\ k = 4$ *dir.*

$(I)^2$	$(II)^2$	$(III)^2$	$(IV)^2$
49	25	49	9
64	36	49	36
16	36	16	25
129	97	114	70

Tabloya göre,

$$\sum\sum X_{ij}^2 = 129 + 97 + 114 + 70 = 410$$

bulunur. Bulduğumuz değerleri formüllerde yerlerine yazdığımızda,

$$\begin{aligned} G.K.T &= 410 - \frac{(68)^2}{3.4} = 24.67 \\ G.A.K.T &= 390 - \frac{(68)^2}{3.4} = 4.67 \\ G.\dot{I}.K.T &= 24.67 - 4.67 = 20 \end{aligned}$$

elde edilir. Şimdi varyans tablosunu düzenleyelim.

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması
Gruplar Arası	$k-1=4-1=3$	G.A.K.T. = 4.67	G.A.K.O. =4.67/3 = 1.56
Gruplar İçi	$k(n-1)=4(3-1)=8$	G.İ.K.T. = 20	G.İ.K.O. =20/8 = 2.5

$$F = \frac{G.A.K.O}{G.\dot{I}.K.O} = \frac{1.56}{2.5} = 0.62$$

F tablosundan 0.05 önem seviyesi ve 3 ile 8 serbestlik derecesine göre F değeri, $F_{0.05;3-8} = 4.07$ bulunur. H_0, hipotezi ortalamaların eşit olduğunu ifade ettiğinden, 0.05 önem seviyesinde, ortalamalar arasındaki farklılığın önemli olmadığına karar verilir. Yani, başarı durumunun bölümlere göre farklılık göstermediği söylenebilir.

Örnek 9.2 *Civata üreten bir fabrikada civata iç çaplarının kontrolü amacıyla, üretimden rastgele 5'er birimlik 4 örnek alınmış ve gözlemler aşağıdaki tabloda kaydedilmiştir. İç çaplarda gözlenen farklılıkların önemli olup olmadığın, %1 önem seviyesinde karar veriniz.*

1.örnek	*2.örnek*	*3.örnek*	*4.örnek*
0.54	0.54	0.59	0.44
0.44	0.55	0.55	0.44
0.44	0.49	0.60	0.59
0.52	0.56	0.51	0.59
0.58	0.49	0.56	0.54

Çözüm 9.2

1.örnek	*2.örnek*	*3.örnek*	*4.örnek*
0.54	0.54	0.59	0.44
0.44	0.55	0.55	0.44
0.44	0.49	0.60	0.59
0.52	0.56	0.51	0.59
0.58	0.49	0.56	0.54
2.52	2.63	2.81	2.47

$$\sum\sum X_{ij} = 2.52 + 2.63 + 2.81 + 2.47 = 10.43$$

ve

$$\sum \frac{(\sum X_i)^2}{n} = \frac{6.3504 + 6.9196 + 7.8961 + 6.1009}{5} = 5.4529$$

olur. Şimdi varyans tablosunu düzenleyelim.

$(1.örnek)^2$	$(2.örnek)^2$	$(3.örnek)^2$	$(4.örnek)^2$
0.2916	0.2916	0.3481	0.1936
0.1936	0.3025	0.3025	0.1936
0.1936	0.2401	0.3600	0.3481
0.2707	0.3136	0.2601	0.2112
0.3364	0.2401	0.3136	0.2916
1.2859	1.3879	1.5843	1.2385

$$\sum\sum X_{ij}^2 = 1.2859 + 1.3879 + 1.5843 + 1.2385 = 5.4966$$

$$G.K.T = \sum\sum X_{ij}^2 - \frac{(\sum\sum X_{ij})^2}{kn} = 5.4966 - \frac{(10.43)}{5(4)} = 0.0574$$

$$G.A.K.T = \sum^{k} \frac{(\sum X_i)^2}{n} - \frac{(\sum\sum X_{ij})^2}{kn} = 5.4529 - 5.4392 = 0.0137$$

$$G.\dot{I}.K.T = G.K.T - G.A.K.T = 0.0574 - 0.0137 = 0.0437$$

olarak elde edilir. Şimdi varyans analizi tablosunu düzenleyelim.

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması
Gruplar Arası	$k-1=4-1=3$	G.A.K.T. = 0.0137	G.A.K.O. = 0.0137/3 = 0.0046
Gruplar İçi	$k(n-1)=4(5-1)=16$	G.İ.K.T. = 0.0437	G.İ.K.O. = 0.0437/16 = 0.0027

$$F = \frac{G.A.K.O.}{G.\dot{I}.K.O.} = \frac{0.0046}{0.0027} = 1.7$$

F tablosunda, 0.01 önem seviyesi ve 3 ile 16 serbestlik derecesine ait F değeri, $F_{0.01;3;16} = 5.29$ dur. $1.70 < 5.29$ olduğundan hipotez kabul edilir. Yani, üretilen civataların iç çaplarında görülen farklılıklar, 0.01 önem seviyesinde, önemli değildir. Dolayısıyla, fabrikanın yaptığı üretimin kontrol altında olduğuna, bu önem seviyesinde, karar verilir.

9.2 Varyansların Eşitliğinin Testi

Bundan önceki kesimde ifade ettiğimiz gibi k tane kitle aynı varyanslı normal dağılıma sahipse, G.A.K.O/G.İ.K.O oranı F dağılımına sahiptir. Bu oranın kullanılmasıyla kitle ortalamalarının eşitliği hipotezini test ederiz. Bazan çeşitli kitlelerin varyanslarının karşılaştırılması da ilginç olabilir. Bu kesimde Bartlett testi olarak bilinen bir metodla

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 \qquad (9.10) \\ H_A &: \text{Varyansların tümü eşit değildir} \end{aligned}$$

hipotezini test edeceğiz. Bu amaçla birbirinden bağımsız normal dağılımlı k tane kitleden $n_1, n_2, \cdots, n_k$ elemanlı örnek seçiyoruz. Sırasıyla $S_1^2, S_2^2, \cdots, S_k^2$ örnek varyanslarını, sonra S_p^2 ile göstereceğimiz birleştirilmiş varyansı hesaplarız. $\sum_{i=1}^{k} n_i = n$ alınırsa:

$$S_i^2 = \frac{1}{n_i - 1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_i)^2, \ \ i = 1, 2, 3, \cdots, k \qquad (9.11)$$

ve

$$S_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{n - k} \tag{9.12}$$

olur.

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{n-k}\right] \tag{9.13}$$

olmak üzere, test istatistiği,

$$B = \frac{1}{C}[(n-k)ln(S_p^2) - \sum_{i=1}^{k}(n_i - 1)ln(S_i^2)] \tag{9.14}$$

olur. Burada C'ye düzeltme çarpanı denir. Bartlett testinin dayandığı test istatistiğinin örnek dağılımı yaklaşık olarak χ^2 dağılımıdır. Burada serbestlik derecesi $k-1$ dir. Böylece B'nin değerini $k-1$ serbestlik derecesinde χ^2 tablosundan bulunan χ^2 değeri ile karşılaştırırız. Örnek büyüklükleri küçük olduğunda C çarpanı χ^2'ye yaklaşımı kuvvetlendirir. Daima $C > 1$ olduğundan C'nin kullanılmasıyla test istatistiğinin hesaplanan değeri küçülecektir. Örnek varyansları birbirinden çok farklı ise B'nin değeri büyük olur ve H_0 reddedilir.

Örnek 9.3 *Aşağıdaki tablodaki değerler üç ayrı kitleden alınmış örnekleri gösteriyor. kitle varyanslarının eşitliği hipotezini* 0.05 *önem seviyesi için test ediniz.*

A	*4*	*7*	*6*	*6*		
B	*5*	*1*	*3*	*5*	*3*	*4*
C	*8*	*6*	*8*	*9*	*5*	

Çözüm 9.3 *Ortalamalar:*

$\bar{X}_1 = 23/4$, $\bar{X}_2 = 21/6$, $\bar{X}_3 = 36/5$.

Hipotez:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \\ H_A &: \textit{Varyansların tümü eşit değildir} \end{aligned}$$

Önem seviyesi: $\alpha = 0.05$ *olarak veriliyor.*

Serbestlik derecesi $= k - 1 = 3 - 1 = 2.$

$B > \chi^2_{1-\alpha} = 5.991$ *olarak bulunur.*

Verilen tablodan, $n_1 = 4,\ n_2 = 6,\ n_3 = 5,\ n = 15$ *yazılır.*

Örneklerin varyansları:

$$S_1^2 = 1.583,\ \ S_2^2 = 2.300,\ \ S_3^2 = 2.700$$

ve birleştirilmiş varyans:

$$S_p^2 = \frac{3(1.583) + 5(2.300) + 4(2.700)}{12} = 2.254$$

olur. Ayrıca,

$$C = 1 + \frac{1}{3(3-1)}[1/3 + 1/5 + 1/4 - 1/12] = 1.1167$$

olduğundan test istatistiğinin değeri:

$$B = \frac{1}{1.1167}[(15-3)ln(2.254) - (3ln(1.583) + 5ln(2.300) + 4ln(2.700))] = 0.213$$

dir. $0.213 < 5.991$ *olduğundan varyansların eşitliği hipotezi reddedilemez.*

9.3 Problemler

1. Üç yeni ilaç belli bir rahatsızlığı önleme bakımından öneriliyor. Bu ilaçları, test ediyoruz. İyileşmeye başlamadan önce geçen saatlerin sayısı ölçülen değişkendir. Başlangıç çalışmasında aşağıdaki sonuçlar bulunmuştur.

1.İlaç	2.İlaç	3.İlaç
0.6	1.2	1.5
1.1	1.0	1.2
1.0	0.9	1.3
0.7	0.9	1.2

%5 önem seviyesinde ilaçların etkinliklerinin önemli derecede farklı olup olmadığını test ediniz.

2. Üç cins buğday parsellenmiş 12 arazi üzerine rastgele ekilmiştir. Dekar başına elde edilen buğday teneke olarak aşağıdaki tabloda verilmiştir.

1.cins	2.cins	3.cins
9	14	8
10	12	9
12	13	12
11	10	11

Varyans analizi ile %5 önem seviyesinde buğday cinsleri arasında önemli bir fark olmadığına karar veriniz.

3. Aşağıda verilen tablodaki verilerin:

   i. Varyans analizi tablosunu yapınız.

   ii. Grup ortalamalarını tahmin ediniz.

   iii. Grup ortalamaları arasındaki fark önemlimidir.

1.grup	2.grup	3.grup
25	22	8
15	21	16
10	8	4
12	15	
	11	

4. 16 kuzu herbirinde 4 kuzu olan 4 gruba rastgele bölünmüştür. Her gruba farklı yem verilmiştir. Her kuzunun belli bir zaman sonunda kazandığı ağırlık kilogram olarak aşağıdaki tabloda verilmiştir.

1.yem	2.yem	3.yem	4.yem
65	77	100	92
67	80	102	87
64	72	95	86
70	78	105	95

Varyans analizi tablosunu kurunuz ve %5 önem seviyesinde yemler arasında önemli bir fark olup olmadığını kontrol ediniz.

5. Aşağıdaki tablodaki değerler üç ayrı kitleden alınmış örnekleri gösteriyor. Kitle varyanslarının eşitliği hipotezini 0.05 önem seviyesi için test ediniz.

A	B	C
4	5	8
7	1	6
6	3	8
6	5	9
	3	5
	7	

Chapter 10

REGRESYON VE KORELASYON TEORİSİ

10.1 Temel Kavramlar

Uygulamada genellikle çeşitli değişkenler arasında bir ilginin bulunduğu görülür. Bu ilginin derecesi "korelasyon"un, matematik olarak ifade edilmesi ise "regresyonun"un konusudur.

Regresyonda, bir bağımlı değişken bir de ona etki eden bağımsız değişken veya değişkenler söz konusudur. Korelasyonda ise, değişkenlerin birinin bağımlı diğerlerinin bağımsız olma şartı yoktur. Başka etkilerle değişme gösteren iki değişken arasında da korelasyon hesaplanır. İleride göreceğimiz gibi, korelasyon katsayısının formülü de iki yönlü bir ilişkiyi göstermektedir.

Dağılma Diyagramları

Tanım 10.1
Koordinat sisteminde (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_n, Y_n) *çiftlerinin kümesinin grafiğine dağılım diyagramı denir.*

Bu diyagramdaki noktaların durumuna göre, regresyon denkleminin şekline karar verilir. Daha sonra bu denkleme ait katsayılar hesaplanır ve nihayet önem kontrolü yapılarak, seçimin iyi yapılıp yapılmadığı ortaya konur. Eğer önem kontrolü sonunda, regresyon denkleminin dağılımı iyi temsil edemediği ortaya çıkarsa, bir başka denklem üzerinde çalışılır.

Uygulamada, en fazla lineer regresyon ve parabol kullanılır. Bunların yanında çok değişkenli problemler, yani, bir bağımlı değişkeni birden fazla bağımsız değişkenin belirlediği problemler de oldukça yaygındır.

Örnek 10.1 *Aşağıdaki tablo* 10 *öğrencinin matematik ve fizik derslerinden aldığı notları göstermektedir. Dağılım diyagramını çiziniz.*

Öğrenci	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Matematik notu(X)	*1*	*8*	*15*	*18*	*23*	*28*	*33*	*39*	*45*	*45*
Fizik notu(Y)	*3*	*14*	*8*	*20*	*19*	*17*	*36*	*26*	*14*	*29*

Çözüm 10.1 .

Şekil 10.1:

10.1.1 En Küçük Kareler Metodu

Regresyon denklemlerinin kuruluşunda en yaygın ve en gerçekçi metod, "en küçük kareler metodu"dur. Bu metodun esası, gerçek değerlerin regresyon doğrusundan (veya eğrisinden) uzaklaşmalarını minimum yapan denklemin bulunmasıdır.

Gerçek değerler Y_i, regresyon değerleri (tahmini değerler) $\hat{Y}_i$ ve bu değerler arasındaki fark m_i olmak üzere,

$$\sum m_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \text{minimum} \tag{10.1}$$

olan denklem, dağılımı, en iyi temsil eden denklemdir. Buradaki m_i aşağıdaki şekilde gösterilmiştir.

Şekil 10.2:

$(X_0, Y_0), (X_1, Y_1), \cdots, (X_n, Y_n)$ noktaları arasından sonsuz sayıda doğru geçebilir. Her doğru için Y_i değerleri ile $\hat{Y}_i$ değerleri arasında değişik farklar çıkacaktır. Bunlar içerisinde herhangi bir doğru için, bu farkların kareler toplamı minimum ise, işte o doğru dağılımı en iyi temsil eden doğrudur.

Genel olarak, $(X_0, Y_0), (X_1, Y_1), \cdots, (X_n, Y_n)$ noktalarından geçen doğru denklemi,

$$\hat{Y}_i = aX_i + b \tag{10.2}$$

olarak yazılabilir. Bu Y_i değeri (10.1) de yerine yazılırsa:

$$\sum m_i^2 = \sum (aX_i + b - Y_i)^2, \quad i = 0, 1, 1, \cdots, n \tag{10.3}$$

olur. Biz $\sum m_i^2$ değeri minimum olacak şekilde a ve b değerlerini bulmak istiyoruz. Bunun için (10.3) eşitliğinin sağ tarafını sıfıra eşitleyip hem a'ya hemde b'ye göre kısmi türev alırsak, aşağıdaki iki denklemi elde ederiz:

$$\begin{aligned} \sum Y &= a \sum X + nb \\ \sum XY &= a \sum X^2 + b \sum X \end{aligned} \tag{10.4}$$

Bu denklemlere "normal denklem" denir. Normal denklemler yardımıyla a ve b değerleri bulunur ve $y = ax + b$ denkleminde yerine yazılarak istenen regresyon denklemi bulunmuş olur.

Parabol için de benzer yolla

$$\begin{aligned} \sum Y &= nc + b\sum X + a\sum X^2 \\ \sum XY &= c\sum x + b\sum X^2 + a\sum X^3 \\ \sum X^2 Y &= c\sum X^2 + b\sum X^3 + a\sum X^4 \end{aligned} \tag{10.5}$$

denklemler elde edilir.

Zaman serisi problemlerinde, X_i zamanı gösterir. X_i değerlerini istediğimiz gibi kodlama imkanına sahip olduğumuzdan, bu değerleri $\sum X = 0$ olacak şekilde kodladığımızda X ve X^3 toplamları sıfır olacağından yukarıdaki normal denklemler

$$\begin{aligned} \sum Y &= nc + a\sum X^2 \\ \sum XY &= b\sum X^2 \\ \sum X^2 Y &= c\sum X^2 + a\sum X^4 \end{aligned} \tag{10.6}$$

şekline dönüşür. Buradan,

$$b = \frac{\sum XY}{\sum X^2} \tag{10.7}$$

olarak bulunur. Burada b'ye regresyon katsayısı denir.

Örnek 10.2 *Aşağıdaki tabloda verilen X, Y değişkenleri için en küçük kareler metodunu kullanarak doğru denklemini bulunuz.*

X	0	2	4	6	8	10
Y	1.0	5.1	9.0	13.0	17.0	21.0

Çözüm 10.2 *İstenen doğrunun denklemi*

$$Y = aX + b$$

şeklindedir. Bu denklemdeki a ve b sabitlerini bulmamız için gerekli regresyon normal denklemleri şunlardır:

$$\begin{aligned} \sum Y &= a\sum X + nb \\ \sum XY &= a\sum X^2 + b\sum X \end{aligned}$$

Bu denklem sisteminde, ihtiyaç duyulan toplamlar aşağıda hesaplanmıştır.

X	Y	X^2	XY
0	1	0	0
2	5.1	4	10.2
4	9.0	16.0	36.0
6	13.0	36.0	78.0
8	17.0	64.0	136.0
10	21.0	100.0	210
30	66.1	220.0	470.2

Bu toplamları normal denklemlerde yerlerine yazarsak

$$\begin{aligned} 66.2 &= 30a + 6b \\ 470.2 &= 220.0a + 30b \end{aligned}$$

lineer denklem sistemi elde edilir. Bu sistem çözülürse $a = 1.9885,\ b = 1.090$ *olarak bulunur. Buna göre regresyon doğru denklemi,*

$$Y = 1.9885X + 1.090$$

olarak ifade edilir.

Örnek 10.3 *Aşağıdaki tablo, bir fakültenin öğrencileri arasından rastgele seçilen* 10 *öğrencinin Matematik ve Fizik notlarını göstermektedir:*

Matematik	75	80	93	65	87	71	98	68	84	77
Fizik	82	78	86	72	91	80	95	72	89	74

Buna göre:

i. *Matematik notlarını bağımsız değişken kabul ederek, verilere için en uygun doğrusuyu bulunuz.*

ii. *Matematik dersinden* 75 *alan bir öğrencinin Fizik dersinden kaç alması beklenir.*

iii. *Fizik dersinden* 95 *almış bir öğrencinin Matematik'ten kaç alması beklenir.*

Çözüm 10.3 *i. İstenen doğru, en küçük kareler metodu kullanılarak bulunacak doğrudur. Regresyon normal denklemlerinin*

$$\begin{aligned} \sum Y &= a\sum X + nb \\ \sum XY &= a\sum X^2 + b\sum X \end{aligned}$$

şeklinde olduğu bilinmektedir. Buradaki ifadeleri bulmak için aşağıdaki tabloyu kullanırız:

X	Y	XY	X^2
75	82	6150	5625
80	78	6240	6400
93	86	7998	8649
65	72	4680	4225
87	91	7917	7569
71	80	5680	5041
98	95	9310	9604
68	72	4896	4624
84	89	7476	7056
77	74	5698	5929
798	819	66045	64722

Bu toplamları regresyon denklemlerinde yerlerine yazdığımızda

$$\begin{aligned} 00819 &= 798a + 10b \\ 66045 &= 64722a + 798b \end{aligned}$$

denklem sistemi elde edilir. Bu denklem sistemi çözülerek $a = 0.66,\ b = 29.23$ *bulunur. a ve b değerleri doğru denkleminde yerlerine yazılırsa, istenen doğrunun denklemi,*

$$Y = 0.66X + 29.23$$

olarak bulunur.

ii. Bağımsız değişken X, matematik notlarını temsil ettiği için, doğru denkleminde $X = 75$ yazılırsa,

$$\hat{Y} = 0.66(75) + 29.23 \approx 79$$

bulunur.

iii. Bu defa doğru denkleminde $Y = 95$ yazılırsa,

$$\begin{aligned} 95 &= 0.66X + 29.23 \\ X &= (95 - 29.23)/0.66 = 99.65 \approx 100 \end{aligned}$$

bulunur.

10.1.2 Tahminin Standart Hatası

Standart hata, gerçek değerlerin, tahmini değerlerden ayrılışının bir ölçüsüdür. Temelde standart sapmaya benzer. Standart sapmada, verilerin ortalama etrafındaki dağılışları belirleniyor ve hangi seride bu sapma eksik çıkarsa, onun daha düzgün olduğuna, pratik olarak, karar veriliyordu. Standart hatada ise, $\hat{Y}$ değerlerinin Y değerlerinden sapmaları ölçülmekte ve standart hatası daha düşük olan denklemlerin, dağılımı daha iyi temsil edebileceklerine, pratik olarak, karar verilebilmektedir. Bununla beraber, bu kararın regresyon analiziyle verilmesi en sağlam yoldur. Bu analizin esası yine standart hataya dayanmaktadır.

Standart hata için genel formül

$$S_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{s.d}}, \quad s.d = \text{serbestlik derecesi} \tag{10.8}$$

şeklindedir. Serbestlik derecesi doğru için $n-2$, parabol için $n-3$, kübik eğri için $n-4$ ve iki bağımsız parabol için $n-3$ dür.

Standart sapmayı hesaplamak için kullanılan sade formüllerin benzeri standart hata için de belirlenmiştir. Bunlar

$$S_{YX} = \sqrt{\frac{\sum y^2 - a\sum xy}{n-2}}, \quad \text{burada } (x = X - \bar{X},\ y = Y - \bar{Y}) \tag{10.9}$$

veya

$$S^2_{YX} = \frac{\sum y^2 - a\sum xy}{n-2} \tag{10.10}$$

olarak yazılır. Parabol için,

$$Y = aX^2 + bX + c \tag{10.11}$$

olmak üzere,

$$S^2_{YX} = \frac{\sum y^2 - b\sum xy - a\sum x^2 y}{n-2} \tag{10.12}$$

şeklindedir.

10.2 Lineer Regresyon Katsayısının Önem Kontrolü

10.2.1 t Testi İle Önem Kontrolü

Regresyon katsayısı (lineer regresyon için $a = \sum XY / \sum X^2$, parabol için $b = \sum XY / \sum X^2$ dir) bir istatistiksel parametresi olup, ait olduğu kitlenin regresyon katsayısı β dır.

Bu testte hipotez, "β'nın önemli olmadığı" şeklinde kurulur. Alternatif hiotez ise, "β'nın önemli olduğu" şeklindedir. Yani,

$$\begin{aligned} H_0 &: \beta = 0 \\ H_1 &: \beta \neq 0 \end{aligned} \tag{10.13}$$

dır. Buna göre, test sonunda hipotez kabul edilirse, regresyon un önemli olmadığına, reddedilirse önemli olduğuna karar verilecektir. Teste esas olacak t değeri ise lineer regresyon için

$$t = \frac{a - \beta}{S_a} = \frac{a}{S_a},$$

parabol için,

$$t \quad = \quad \frac{b-\beta}{S_b} = \frac{b}{S_b} \tag{10.14}$$

formülüyle hesaplanır. Burada S_a veya S_b, regresyon katsayısının standart sapmasıdır ve

$$S_a \quad = \quad S_b = \frac{S_{YX}}{\sqrt{\sum x^2}}, \quad (\sum x^2 = \sum (X - \bar{X})^2 = \sum X - (\sum X)^2/n) \tag{10.15}$$

dir.

Şekil 10.3:

10.2.2 Varyans Analiziyle Önem Kontrolü

Varyans analizinde, genel kareler toplamı $\sum(Y-\bar{Y})^2$, regresyon $\sum(\hat{Y}-\bar{Y})^2$ ve hata kareler toplamı $\sum(Y-\hat{Y})^2$ olmak üzere iki kısma ayrılarak analize tabi tutulur. Bunlar arasında

$$\sum(Y-\bar{Y})^2 = \sum(\hat{Y}-\bar{Y})^2 + (Y-\hat{Y})^2 \tag{10.16}$$

bağıntı vardır. Hipotez, regresyonun önemli olmadığı şeklinde kurulur. Lineer regresyon için regresyon kareler toplamı $a\sum xy$, hata kareler toplamı ise $\sum y^2 - a\sum xy$ dir.

.

Şekil 10.4:

Bu toplamların kendi serbestlik derecelerine bölünmesiyle birer varyans tahmini elde edilir. Daha sonra, regresyon kareler ortalaması, hata kareler ortalamasına bölünerek F değeri bulunur. Bulunana F değeri, F cetvelindeki kritik değerlerden büyükse, hipotez reddedilerek "regresyon denkleminin dağılımı temsil edemediğine" karar verilir. Lineer regresyon için regresyon analizi tablosu (s.d ile serbestlik derecesini, K.T ile kareler toplamını ve K.O ile de kareler ortalamasını gösterirsek) aşağıdaki gibidir:

Değişim Kaynağı	s.d	K.T	K.O
Regresyon	1	$a\sum xy$	$a\sum xy$
Hata	$n-1$	$\sum y^2 - a\sum xy$	$(\sum y^2 - a\sum xy)/(n-2)$

Burada,

$$\frac{\sum y^2 - a\sum xy}{n-2} = S_{yx}^2 \tag{10.17}$$

ve dolayısıyla,

$$F = \frac{a\sum xy}{S_{yx}^2} \tag{10.18}$$

olur. Burada,

$$\sum xy = \sum XY - \frac{(\sum X)(\sum Y)}{n}, \quad \sum y^2 = \sum Y^2 - \frac{\sum Y^2}{n}$$

şeklindedir.

10.3 Parabol İçin Regresyon Önem Kontrolü

Parabol için regresyon analiz tablosu aşağıdaki gibi oluşturulur. Analizin diğer bölümleri lineer regresyon gibidir.

Değişim Kaynağı	s.d	K.T.	K.O
Regresyon	2	$b\sum xy + a\sum x^2y$	R.K.T/2
Hata	$n-3$	$\sum y^2 - (b\sum xy + a\sum x^2y)$	H.K.T/(n-3)

$$\sum x^2y = \sum x_i^2 y_i - \frac{(\sum x_i^2)(\sum y_i)}{n} \tag{10.19}$$

olarak yazılır. Burada $\sum x^2y = \sum X^2Y - ((\sum X^2)(\sum Y))/n$ dir.

10.4 Çok Değişkenli regresyon

Bağımsız değişkenin birden fazla olduğu problemlerde, normal denklemler, en küçük kareler metoduyla ve daha önce açıkladığımız yolla bulunur. Bununla beraber pratik olarak, regresyon denkleminin her iki tarafı, söz konusu değişkenlerle ayrı ayrı çarpılarak taraf tarafa toplanır. Örneğin, iki bağımsız değişkenli problemlerde regresyon denklemi,

$$Y = a + b_1X_1 + b_2X_2 \tag{10.20}$$

şeklindedir. Normal denklemlerin bulunması için, eşitliğin her iki yanı taraf tarafa n defa toplanarak birinci denklem elde edilir. Sonra bu denklemin her iki tarafı X_1 ile çarpılarak ikinci denklem bulunur. Daha sonra ise birinci denklemin her iki tarafı X_2 ile çarpılarak üçüncü denklem yazılır. Böylece

$$\begin{aligned}
\sum Y &= na + b_1\sum X_1 + b_2\sum X_2 \\
\sum X_1Y &= a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \\
\sum X_2Y &= a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2
\end{aligned} \tag{10.21}$$

normal denklemleri elde edilir.

İki bağımsız değişkenli problemlerde standart hata

$$S_{yx} = \sqrt{\frac{\sum y^2 - b_1 \sum x_1 y - b_2 \sum x_2 y}{n-3}} \tag{10.22}$$

ile hesaplanır.

10.4.1 İki Bağımsız Değişkenli Problemlerde Regresyon Analizi

Bu problemlerde, aşağıdaki analiz tablosu kullanılır. Analizin diğer kısımları lineer durumda olduğu gibi yürütülür.

Değişim Kaynağı	s.d	K.T.	K.O
Regresyon	2	$b_1 \sum x_1 y + b_2 \sum x_2 y$	R.K.T/2
Hata	$n-3$	$\sum y^2 - (b_1 \sum x_1 y + b_2 \sum x_2 y)$	H.K.T/(n-2)

İkiden fazla bağımsız değişken ihtiva eden problemlerde de benzeri tablolar oluşturulur. Örneğin, üç değişkenli problemlerde tablodaki regresyon kareler toplamına $b_3 \sum x_3 y$ ilave edilir. Hata hesaplamasında da $\sum y^2$ den çıkarılan kısım içine yine bu $b_3 \sum x_3 y$ eklenir. regresyon için $s.d$ ise 3 olur.

Örnek 10.4 *Tasarrufun milli gelire bağlı olarak değiştiğini kabul ederek aşağıdaki veriler için:*

i. En küçük kareler metodunu kullanarak regresyon doğrusunu bulunuz.

ii. Kişi başına milli gelirin ayda 2 milyon lira olması halinde, tasarrufun ne kadar olacağını hesaplayınız.

iii. Tahmini standart hatayı bulunuz.

iv. Regresyon katsayısının %1 seviyesinde önem kontrolünü yapınız.

Kişi başına milli gelir (1 000 000 TL)	*Tasarruf (100 000)*
6.0	*5*
7.2	*8*
8.6	*9*
12.0	*15*
15.6	*17*

Çözüm 10.4 *i. Regresyon normal denklemleri şunlardır:*

$$\begin{aligned} \sum Y &= a\sum X + nb \\ \sum XY &= a\sum X^2 + b\sum X \end{aligned}$$

Tasarruf, milli gelire bağlı olarak değiştiği için, bağımlı değişkendir. Yani, tasarruf Y, milli gelir ise X olarak alınır.

X	Y	XY	X^2	Y^2
6.0	*5*	30.0	36.00	25
7.2	*8*	57.6	51.84	64
8.6	*9*	77.4	73.96	81
12.0	*15*	180.0	144.00	225
15.6	*17*	265.2	243.36	289
49.4	*54*	610.2	549.16	684

Burada $n = 5$ dir. Bu toplamlar regresyon normal denklemlerinde yerlerine yazılırsa,

$$\begin{aligned} 54 &= 49.4a + 5b \\ 610.2 &= 549.16a + 49.4b \end{aligned}$$

olur. Bu denklem sistemi çözülerek $a = 1.25$ ve $b = -1.55$ bulunur. Regresyon doğru denklemi,

$$Y = 1.25X - 1.55$$

olur.

ii. $Y = 1.25X - 1.55$ *denkleminde* $X = 2$ *yazılırsa,* $Y = 1.25(2) - 1.55 = 0.95$ *milyon TL olur.*

iii. Tahmini standart hatayı aşağıdaki iki yol ile bulabiliriz:

1. **yol:**

$$S_{YX} = \sqrt{\frac{\sum y^2 - b\sum y - a\sum xy}{n-2}}$$

formülünü kullanalım. a ve b değerleri yukarıda bulundu. Buna göre,

$$S_{YX} = \sqrt{\frac{684 - (-1.55)(54) - (1.25)(610.2)}{5-2}} = 1.28$$

olarak elde edilir.

2. **yol :** *Şimdi de*

$$S_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}}$$

formülünü kullanalım. Bunun için $\hat{Y}_i$ *değerlerini tek tek bulmak gerekir. Bilindiği gibi her bir* X *değeri regresyon denkleminde yerine yazılırsa buna karşılık gelen değer* $\hat{Y}$ *dir.*

X	Y	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
6.0	*5*	$-1.55 + 1.25(6.00) = 5.950$	-0.95	0.9025
7.2	*8*	$-1.55 + 1.25(7.20) = 7.450$	0.550	0.3025
8.6	*9*	$-1.55 + 1.25(8.60) = 9.200$	-0.20	0.0400
12.0	*15*	$-1.55 + 1.25(12.0) = 13.45$	1.550	2.4025
15.6	17	$-1.55 + 1.25(15.6) = 17.95$	-0.95	0.9025
				4.5500

Bu tabloya göre

$$S_{YX} = \sqrt{\frac{4.55}{5-2}} = 1.23$$

olarak bulunur.

Görüldüğü gibi 1.yol ile $S_{YX} = 1.28$*; 2.yol ile* $S_{YX} = 1.23$ *bulunmuştur. Aralarındaki bu farklılık, a ve b sabitlerinin hesabında, yaklaşık değerlerin kullanılmasından doğmuştur.*

iv. 1. **yol (t testi ile önem kontrolü):** *Bilindiği gibi*

$$t = \frac{a-\beta}{S_a} = \frac{a}{S_a}, \quad (S_a = \frac{S_{YX}}{\sum x^2})$$

idi. Yukarıda $a = 1.25$; $S_{YX} = 1.28$ *olarak bulunmuştu. O halde*

$$\sum x^2 = \sum (X - \bar{X})^2 = \sum X^2 - (\sum X)^2/n = 549.16 - (49.4)^2/5 = 61.09$$

olur. Bu değerleri yerine yazarsak,

$$t = \frac{1.25}{1.28/61.09} = 59.65$$

olarak bulunur. %1 *önem seviyesi ve* $n - 2 = 3$ *serbestlik derecesi için karar modeli aşağıdaki gibi olur.*

Şekil 10.5:

$59.65 > 5.84$ *olduğundan, regresyon katsayısının önemli olduğuna,* 0.01 *önem seviyesinde karar verilir. Yani, bu regresyon denklemi ile tahmin yapılabilir.*

2. **yol (Varyans analizi ile önem kontrolü):** *Regresyon denklemi lineer olduğu için regresyon analizi tablosu aşağıdaki gibi olur.*

Değişim Kaynağı	*s.d*	*K.T.*	*K.O*
Regresyon	*1*	$a\sum xy$	$a\sum xy$
Hata	$n-1$	$\sum y^2 - a\sum xy$	$(\sum y^2 - a\sum xy)/(n-2)$

$$\frac{\sum y^2 - a\sum xy}{n-2} = S_{yx}^2$$

ve

$$F = \frac{a\sum xy}{S_{yx}^2}$$

dir. $\sum xy = \sum XY - [(\sum X)(\sum Y)]/n$ *ve* $\sum y^2 = \sum Y^2 - [(\sum Y)^2]/n$ *şeklindedir. Tablodan gerekli değerler yerine yazılırsa;*

$\sum xy = 610.2 - [(49.4)(54)]/5 = 76.68$ *ve* $\sum y^2 = 684 - [(54)^2]/5 = 100.8$ *olur. Böylece*

$$F = \frac{(1.25)(76.68)}{(100.8 - (1.25)(76.68))/(5-2)} = 58.09$$

bulunur. Tablo değeri, 0.01 *önem seviyesi için* $F_{0.01;1;3} = 34.1$ *dir.* $58.09 > 34.1$ *olduğundan regresyon katsayısının önemli olduğuna,* 0.01 *önem seviyesinde karar verilir.*

Örnek 10.5 *Aşağıdaki tablo bir mal için fiat-talep ilişkisini göstermektedir.*

Fiat	*1*	1.5	2.2	*3*	*4*
Talep	*18*	*12*	*8*	*2*	*0*

Buna göre,

i. Dağılımın diyagramını çiziniz,

ii. Verilerin parabolik eğri gösterdiğini kabul ederek, regresyon denklemini bulunuz,

iii. Belirtilen fiatlar için tahmini talep değerlerini bulunuz,

iv. Tahmini standart hatayı bulunuz,

v. %1 *önem seviyesinde, regresyon önem kontrolünü yapınız.*

Çözüm 10.5 *i.*

Şekil 10.6:

ii. Parabol için regresyon denklemi $Y = aX^2+bX+c$ *şeklindedir. Regresyon normal denklemleri ise aşağıdaki gibi olur.*

$$\begin{aligned}\sum Y &= nc + b\sum X + a\sum X^2\\ \sum XY &= c\sum X + b\sum X^2 + a\sum X^3\\ \sum X^2y &= c\sum X^2 + b\sum X^3 + a\sum X^4\end{aligned}$$

a, b ve c değerlerini bulmak için normal denklemlerdeki toplamları bulmamız gerekir. Bu toplamlar aşağıdaki tabloda gösterilmiştir. Talebin fiata göre değiştiği düşünülürse, talebi Y*, fiatı* X *ile göstermek gerekir.*

X	Y	XY	X^2	Y^2	X^2Y	X^3	X^4
1.0	*18*	18.0	1.00	*324*	18.00	1.000	1.0000
1.5	*12*	18.0	1.25	*144*	27.00	3.375	5.0623
2.2	*8*	17.6	4.84	*64*	38.72	10.648	23.4256
3.0	*2*	0.0	9.00	*4*	18.00	27.000	81.0000
4.0	0	0.0	16.00	0	0.00	64.000	256.0000
11.7	*40*	59.6	33.09	*534*	101.72	106.023	366.4881

$$\begin{aligned}40 &= 5c + 11.7b + 33.09a\\ 59.6 &= 11.7c + 33.09b + 106.023a\\ 101.72 &= 33.09c + 106.023b + 366.4881a\end{aligned}$$

Bu denklem sistemi çözülürse, $a = 1.31$*,* $b = -12.49$ *ve* $c = 28.54$ *bulunur. O halde regresyon denklemi,*

$$Y = 1.31X^2 - 12.49X + 28.54$$

olur.

iii.

Fiat (X)	*Tahmini talep* ($\hat{Y}$)
1.0	$28.54 - (12.49)(1.0) + (1.31)(1.0)^2 = 17.36$
1.5	$28.54 - (12.49)(1.5) + (1.31)(1.5)^2 = 12.75$
2.2	$28.54 - (12.49)(2.2) + (1.31)(2.2)^2 = 7.400$
3.0	$28.54 - (12.49)(3.0) + (1.31)(3.0)^2 = 2.860$
4.0	$28.54 - (12.49)(4.0) + (1.31)(4.0)^2 = -0.46$

iv. Parabol için tahmini standart hata şu üç formül ile de elde edilebilir. Fakat sonuçta bazı küçük hatalar olabilir bunun da sebebi hesaplamada bazı değerlerin yaklaşık olarak alınmasıdır.

$$
\begin{aligned}
S_{YX} &= \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-3}} \\
&= \sqrt{\frac{\sum y^2 - b\sum xy - a\sum x^2 y}{n-3}} \\
&= \sqrt{\frac{\sum Y^2 - c\sum Y - b\sum XY - a\sum X^2 Y}{n-3}}
\end{aligned}
$$

Bu formüllerin her biriyle tahminin standart hatasını bulalım.

y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
18	17.36	0.64	0.4096
12	12.75	−0.75	0.5625
8	7.40	0.60	0.3600
2	2.86	0.86	0.7396
0	−0.46	0.46	0.2116
			2.2833

$$
S_{YX} = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-3}} = \sqrt{2.28335 - 3} = 1.07
$$

$$
\begin{aligned}
S_{YX} &= \sqrt{\frac{\sum y^2 - b\sum xy - a\sum x^2 y}{n-3}} \\
&= \sqrt{\frac{214-(-12.49)(-34)-(1.31)(-1.63)}{5-3}} = 1.20
\end{aligned}
$$

$$
\begin{aligned}
S_{YX} &= \sqrt{\frac{\sum Y^2 - c\sum Y - b\sum XY - a\sum X^2 Y}{n-3}} \\
&= \sqrt{\frac{534-(28.54)40-(-12.49)(59.6)-(1.31)(101.72)}{5-3}} = 1.33
\end{aligned}
$$

olur. Eğer işlemler daha hassas olarak yapılırsa yukarıdaki üç değerde birbirine eşit olur.

v. Regresyon için varyans analizi tablosu aşağıdaki gibidir.

Değişim Kaynağı	*s.d*	*K.T.*	*K.O*
Regresyon	*2*	$b\sum xy + a\sum x^2y$	*R.K.T/2*
Hata	$n-3$	$\sum y^2 - (b\sum xy + a\sum x^2y)$	*H.K.T/(n-3)*

Buradan

$$\begin{aligned} F &= \frac{\textit{Regresyon kareler ortalaması}}{\textit{Hata kareler ortalaması}} \\ &= \frac{(b\sum xy + a\sum x^2y)/2}{(\sum y^2 - (b\sum xy + a\sum x^2y))/(n-3)} \\ &= \frac{(-12.49)(-34) + (1.31)(-163)/2}{1.435} \\ &= 73.56 \end{aligned}$$

olarak bulunur. Payın serbestlik derecesi 2, paydanın serbestlik derecesi $n-3=2$ *ve önem seviyesi* 0.01 *olduğundan,* $F_{0.01;2;2} = 99$ *dur.*

Şekil 10.7:

$73.56 < 99$ *olduğundan, regresyon katsayısının önemli olmadığı hipotezi, %1 önem seviyesinde kabul edilir. Yani, bu regresyon denklemini kullanarak tahmin yapmak,* 0.01 *önem seviyesinde uygun değildir.*

10.5 Korelasyon

İki değişken arasındaki ilgi, ya bunlardan birinin diğerine bağlı olarak değişmesinden veya her iki değişkenin başka faktörlere bağlı olarak birlikte değişmelerinden kaynaklanır. Genelde bunlardan birincisi regresyon, ikincisi ise korelasyonun konusudur. Bununla beraber, korelasyon katsayısı, regresyon değerinin gerçek değere uygunluğunu ölçmede de kullanılır.

Korelasyon katsayısı +1 ile -1 arasında değişir. Yani, $-1 \leq r \leq +1$ dir. r nin -1 çıkması ters yönde, +1 çıkması ise aynı yönde, %100'lük bir korelasyonu ifade eder.

Korelasyon katsayısı için genel formül,

$$r = \pm\sqrt{\frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y-\bar{Y})^2}} \tag{10.23}$$

şeklindedir. Kök içerisindeki kesrin payı, "sebebi bilinen değişim" olarak adlandırılır. Yani, genel kareler toplamı içerisinde regresyonla belirlenebilen kısmı ifade eder. Kesrin paydası ise genel kareler toplamıdır. Buna göre, kök içerisindeki kısım, toplam değişimin yüzde kaçının, regresyonla belirlenebileceğini gösterir. Bu oranın, örneğin %90 çıkması halinde toplam değişimin %90'ı regresyonla, yani bağımsız değişkenler tarafından belirlenmiştir, %10'u ise ya tesadüfen meydana gelmiş veya dikkate alınmayan başka değişkenlerce belirlenmiştir denilir. Bu orana determinasyon katsayısı adı verilir ve

$$r^2 = \frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y-\bar{Y})^2} = \frac{\sum(Y-\bar{Y})^2 - \sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2} = 1 - \frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2} \tag{10.24}$$

veya

$$r^2 = 1 - \frac{S_{YX}^2}{S_Y^2} \tag{10.25}$$

olarak ifade edilir.

10.5.1 Lineer Korelasyon Katsayısı

Korelasyon problemlerinde, genellikle, X ile Y arasında lineer bir ilgi olduğu kabul edilir. Bu kabule göre, yukarıdaki genel denklem yerine, lineer korelasyon için şu eşitlik geliştirilmiştir.

$$r = \frac{n\sum XY - (\sum X)(\sum X)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \tag{10.26}$$

Gerçek X ve Y değerleri yerine, bunların ortalamadan farkları esas alındığında, korelasyon katsayısı formülü

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}, \quad (x = X - \bar{X}),\ (y = Y - \bar{Y}) \tag{10.27}$$

şekline dönüşür. Formüllerden de anlaşılacağı gibi, korelasyon katsayısında iki değişken arasındaki karşılıklı ilişki üzerinde durulur. Değişkenlerden hangisini x veya y olarak alırsanız alınız, sonuçta bir değişme olmaz.

Yukarıdaki eşitlikte her iki tarafın karesi alındığında,

$$r^2 = \frac{(\sum xy)^2}{(\sum x^2)(\sum y^2)} = \left(\frac{\sum xy}{\sum x^2}\right)\left(\frac{\sum xy}{\sum y^2}\right) = (b_{xy})(b_{yx}) \qquad (10.28)$$

elde edilir. Yani determinasyon katsayısı, her iki regresyon katsayısının çarpımı olarak bulunmuştur. Dolayısıyla, korelasyon katsayısı da bu iki katsayının geometrik ortalaması olarak

$$r = \sqrt{(b_{xy})(b_{yx})} \qquad (10.29)$$

şeklinde yazılır.

Örnek 10.6 *Aşağıdaki tabloda verilen X ve Y değişkenleri arasındaki lineer korelasyon katsayısını bulunuz.*

X	2	4	5	7	9	10	12	15
Y	1	1	3	3	4	5	7	8

Çözüm 10.6 *Çözümü iki yoldan yapacağız.*

1.yol: *Problemdeki verilerden $\bar{X} = 64/8 = 8$ ve $\bar{Y} = 32/8 = 4$ olarak bulunur. Bunları kullanarak aşağıdaki tabloyu yapalım:*

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	xy	x^2	y^2
2	1	-6	-3	18	36	9
4	1	-4	-3	12	16	9
5	3	-3	-1	3	9	1
7	3	-1	-1	1	1	1
9	4	1	0	0	1	0
10	5	2	1	2	4	1
12	7	4	3	12	16	9
15	8	7	4	28	49	16
64	32			76	132	46

Buna göre

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{76}{\sqrt{(132)(46)}} = 0.97$$

bulunur.

2.yol: *İkinci yol olarak korelasyon katsayısını bulmak için önce aşağıdaki tablo düzenlenir.*

X	Y	XY	X^2	Y^2
2	1	2	4	1
4	1	4	16	1
5	3	15	25	9
7	3	21	49	9
9	4	36	81	16
10	5	50	100	25
12	7	84	144	49
15	8	120	225	64
64	32	332	644	174

Buradan da,

$$\begin{aligned} r &= \frac{n\sum XY - (\sum X)(\sum X)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \\ &= \frac{8(332) - (64)(32)}{\sqrt{[8(644) - (64)^2][8(174) - (32)^2]}} = 0.97 \end{aligned}$$

bulunur.

$\bar{X}$ ve $\bar{Y}$ çoğu zaman tamsayı olmadığından, korelasyon katsayısını ikinci yolla hesaplamak daha kolay olacaktır.

10.5.2 Korelasyon Katsayısının Önem Kontrolü

r korelasyon katsayısı bir istatistiksel parametre olup, ona karşılık gelen kitle parametresi ρ dur. Önem kontrolünde hipotez "ρ'nun önemsiz olduğu" şeklinde kurulur. Yani,

$$\text{Hipotez,} \quad H_0 : \rho = 0$$

$$\text{Alternatif}, \quad H_1 : \rho \neq 0 \tag{10.30}$$

dır. Testte esas olan t dağılımı,

$$t = \frac{r-\rho}{S_r} = \frac{r}{S_r} \tag{10.31}$$

formülüyle hesaplanır. Burada S_r korelasyon katsayısının standart hatasıdır ve,

$$S_r = \sqrt{\frac{1-r^2}{n-2}} \tag{10.32}$$

eşitliği ile bulunur.

Örnek büyüklüğünün $n > 30$ olduğu durumlarda, t yerine z değeri kullanılır ve testte esas olacak z değeri,

$$z = \frac{r}{\sigma_r} \tag{10.33}$$

eşitliği ile hesaplanır. Burada σ_r ise

$$\sigma_r = \frac{1}{\sqrt{n-1}} \tag{10.34}$$

dir. Her iki test çift yönlüdür.

Burada bir noktanın belirtilmesinde fayda vardır. Bulunan korelasyon katsayısı önemli çıktığı zaman, x ile y arasında önemli bir ilişki olduğunu söyleyebiliriz. Fakat, öncelikle bu değişkenler arasında mantıki bir ilginin bulunması şarttır. Bazen hiç alakası olmayan değişkenler arasında da yüksek bir korelasyon çıkabilmektedir. Bu tip korelasyonlara "sahte korelasyon" adı verilir.

Diğer önemli bir husus da şudur: Örnek büyüklüğü arttıkça, daha düşük yüzdelerin bile testte önemli çıkması mümkün olabilmektedir. Örneğin, bir problemde %90'lık bir korelasyon, test sonunda önemsiz çıkarken, diğer bir problemde %80'lik korelasyon önemli çıkabilmektedir. Böyle durumlarda ikinci testte, örneğin büyük olmasının bu sonucu doğurduğu bilinecektir.

Örnek 10.7 *Bir işletme aynı üretim faktörlerini kullanarak A ve B gibi iki tip mal üretmektedir. Bu malların üretim miktarları aşağıdaki tabloda belirtilmiştir.*

A	2	4	6	8	10	12	14
B	99	96	91	84	75	64	51

Üretim miktarları için,

i. *Lineer korelasyon katsayısını bulunuz,*

ii. *Parabolik korelasyon katsayısını bulunuz,*

iii. *Her iki korelasyon katsayısının* 0.05 *önem seviyesinde önem kontrolünü yaparak sonucu yorumlayınız.*

Çözüm 10.7 i. *Lineer korelasyon katsayısını bulmak için aşağıdaki tablo kullanılır:*

X	Y	XY	X^2	Y^2	X^2Y
2	99	198	4	9801	396
4	96	384	16	9216	1536
6	91	546	36	8281	3276
8	84	672	64	7056	5376
10	750	75	100	5625	7500
12	64	768	144	4096	9216
14	51	714	196	2601	9996
56	560	4032	560	46676	37296

Buradan

$$\begin{aligned} r &= \frac{n\sum xy - (\sum x)(\sum x)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \\ &= \frac{7(4032) - (56)(560)}{\sqrt{[7(560) - (56)^2][7(46676) - (560)^2]}} \\ &= \frac{-3136}{3208} \\ &= -0.97 \end{aligned}$$

bulunur. A ve B arasında negatif bir ilişki vardır.

ii. *Parabolik korelasyon katsayısı için, a ve b regresyon katsayılarını hesaplamamız gerekir. A serisindeki veriler sabit artış gösterdiği için kodlama metodu kullanılır.*

A	B	X	XY	X^2Y	X^2	X^4
2	99	-3	-297	891	9	81
4	96	-2	-192	384	4	16
6	91	-1	-91	91	1	1
8	84	0	0	0	0	0
10	750	1	75	75	1	1
12	64	2	128	256	4	16
14	51	3	153	459	9	81
56	560	0	-224	2156	28	196

olur. Tabloda hesaplanan toplam değerler,

$$\begin{aligned} \sum y &= nc + b\sum x + a\sum x^2 \\ \sum xy &= c\sum x + b\sum x^2 + a\sum x^3 \\ \sum x^2y &= c\sum x^2 + b\sum x^3 + a\sum x^4 \end{aligned}$$

parabolik regresyon normal denklemlerinde yerlerine yazılırsa,

$$\begin{aligned} 560 &= 7c + 0b + 28a \\ -224 &= 0c + 28b + 0a \\ 2156 &= 28c + 0b + 196a \end{aligned}$$

lineer denklem sistemi elde edilir. Bu denklem sisteminin çözülmesiyle, $a = -1$, $b = -8$ *ve* $c = 84$ *bulunur. Ayrıca ilgili formüller kullanılarak,* $\sum xy = -224$, $\sum x^2y = -84$, $\sum y^2 = 1876$ *değerleri de elde edilir. Bulunan bu değerler formülde yerine yazılırsa,*

$$r = \frac{b\sum xy + a\sum x^2y}{\sum y^2} = \frac{-8(-224) + (-1)(-84)}{1876} = -1$$

değeri elde edilir. Bu sonuca göre A ve B arasında $\%100$*'lük ters bir ilişki olduğu ortaya çıkar. Yani,* (X, Y) *ikililerinin tamamı regresyon* *eğrisi üzerindedir.*

$$Y = -X^2 - 8X + 84$$

denklemi, değişen X *değerleri için gerçek* Y *değerlerini verir. Gerçekten,* $A = 12$ *için* $X = 2$ *yazıldığında,*

$$\hat{Y} \;=\; -(2)^2 - 8(2) + 84 = 64$$

bulunur.

iii. Lineer korelasyon katsayısının önem kontrolü:

$$t \;=\; \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.97}{\sqrt{\frac{1-0.94}{5}}} = -8.85$$

ve $t_{0.05;5} = \pm 2.57$ *dir.* $-8.85 < -2.57$ *olduğundan, lineer korelasyon katsayısının* %5 *önem seviyesinde önemli olduğuna karar verilir.*

Parabolik korelasyon katsayısı için $t = \infty$ *çıkar. Bu ise korelasyon katsayısının, serbestlik derecesi ve önem seviyesi ne olursa olsun, önemli olması demektir.*

Her iki durumda da korelasyon katsayısı önemli olmakla beraber, ikinci durumda tam bir korelasyon olduğu görülmektedir.

Örnek 10.8 *Her biri* 65 *birimden meydana gelen* X *ve* Y *değşkenleri arasındaki korelasyon katsayısı* 0.75 *olarak hesaplanmıştır. Bu korelasyon katsayısının önemli olup olmadığını* 0.01 *önem seviyesinde test ediniz.*

Çözüm 10.8 $n = 65 > 30$ *olduğundan önem testi için* z *istatistiği kullanılır.*

$$z \;=\; \frac{r}{\sqrt{\frac{1}{n-1}}} = \frac{-0.75}{\sqrt{\frac{1}{65-1}}} = 6$$

0.01 *önem seviyesi için kritik değerler* $z = \pm 2.58$ *olduğundan, korelasyon katsayısının önemli olduğuna,* 0.01 *önem seviyesinde karar verilir.*

10.5.3 Sıra Korelasyon Katsayısı

Korelasyon katsayısı ekstrem (uç veya anormal) değerlerden fazlasıyla etkilenir. Seride uç değerlerin bulunduğu durumlarda, ya bu değerler elimine edilecek veya sıra korelasyon katsayısı kullanma yoluna gidilecektir. Sıra korelasyon katsayısı x_i, y_i değerlerinin, kendilerinin değil, sıralarının daha önemli olduğu problemlerde başarı ile uygulanır.

Sıra korelasyon katsayısı

$$r_{sıra} = 1 - \frac{6\sum D^2}{n(n^2-1)} \tag{10.35}$$

formülü ile hesaplanır. Burada, D, sıralar arasındaki fark; n ise örnek büyüklüğüdür.

Katsayı hesaplanırken, x ve y serilerindeki rakamlar, kendi içlerinde büyüklük sırasına konulurlar. Daha sonra, bunlara sıra numaraları verilir. Korelasyon hesaplanırken, gerçek değerler yerine bu sıra numaraları esas alınır. Bir seride, örneğin, dördüncü ve beşinci sayılar aynı ise, bunların her ikisinede sıra numarası olarak 4.5 verilir.

Katlı (Çoklu) Korelasyon Katsayısı

Katlı korelasyon katsayısı ikiden fazla değişkenin birlikte değişmelerinin ölçüsüdür. Bağlı değişkenin iki bağımsız değişken tarafından belirlendiği durumlarda regresyon denklemimizi, $Y = a + b_1X_1 + b_2X_2$ şeklinde yazdığımız takdirde, determinasyon katsayısı,

$$R^2 = \frac{\sum(\hat{Y}-Y)^2}{\sum(Y\bar{Y})^2} = \frac{b_1\sum X_1Y + b_2\sum X_2Y}{\sum Y^2} \tag{10.36}$$

olur. Örneğin, $R^2 = 0.80$ çıktığında Y deki toplam değişimin %80'inin X_1 ve X_2 değişkenlerince belirlendiği anlaşılır. Bu katsayının karekökü ise katlı korelasyon katsayısıdır.

Yukardaki determinasyon katsayısını,

$$R^2_{yx_1x_2} = 1 - \frac{S^2_{yx_1x_2}}{S^2_y} \tag{10.37}$$

şeklinde ifade edebiliriz. Burada

$$S^2_{yx_1x_2} = \frac{\sum y^2 - b_1 \sum x_1 y - b_2 \sum x_2 y}{n} \tag{10.38}$$

ve

$$S^2_y = \frac{\sum y^2}{n} \tag{10.39}$$

dir.

Y değişkeninin üç değişkenle belirlenmesi halinde, R^2 denkleminde kesrin payına $b_3 \sum x_3 y$ değeri ilave edilecektir.

Örnek 10.9 *Aşağıdaki X ve Y değişkenleri için sıra korelasyon katsayısını bulunuz.*

X	22	47	38	204	45	164	55
Y	354	602	105	410	98	380	175

Çözüm 10.9 *Önce gerekli değerleri hesaplamak için aşağıdaki tabloyu yapalım.*

X	Y	X sıra	Y sıra	D	D^2
22	354	1	4	-3	9
47	602	4	7	-3	9
38	105	2	2	0	0
204	410	7	6	1	1
45	98	3	1	2	4
164	380	6	5	1	1
55	175	5	3	2	4
					28

Buna göre

$$r_{sıra} = 1 - \frac{6 \sum D^2}{n(n^2-1)} = 1 - \frac{6(28)}{7(49-1)} = 0.5 \tag{10.40}$$

bulunur.

10.5.4 Kısmi Korelasyon Katsayısı

Y bağımlı değişkenine etki eden değişkenlerden bir kısmının etkilerinin sabit kabul edilmesi halinde, diğeriyle Y arasındaki ilişkinin ölçüsü, kısmi korelasyondur.

Örneğin, bağımlı değişkenin X_1, bağımsız değişkenlerin ise X_2 ve X_3 olduğunu kabul edelim. X_3'ün sabit tutulması halinde $r_{12,3}$, X_1 ile X_2 arasındaki kısmi korelasyon katsayısıdır ve şu eşitlikle hesaplanır:

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \tag{10.41}$$

r_{12}; X_1 ile X_2 arasındaki lineer korelasyon katsayısıdır. Üç bağımsız değişkenli bir problemde, X_3 ile X_4'ün sabit tutulmaları halinde, X_1 ile X_2 arasındaki kısmi korelasyon katsayısı ise aşağıdaki eşitliklerin biriyle hesaplanabilir:

$$r_{12,34} = \frac{r_{12,4} - r_{13,4}r_{23,4}}{\sqrt{(1-r_{13,4}^2)(1-r_{23,4}^2)}} = \frac{r_{12,3} - r_{14,3}r_{24,3}}{\sqrt{(1-r_{14,3}^2)(1-r_{24,3}^2)}} \tag{10.42}$$

Örnek 10.10 *Bir tarlada yapılan araştırmaya göre, son beş yıl için, verim, hava sıcaklığı ve gübre miktarına ait veriler aşağıdaki gibi elde edilmiştir.*

Ortalama verim (ton) X_1	*Ortalama hava sıcaklığı (C^o)* X_2	*Ortalama gübre miktarı (100 kg)* X_3
5.0	*16*	*10*
4.5	*14*	*9*
4.0	*15*	*8*
3.0	*13*	*7*
2.5	*14*	*6*

Buna göre aşağıdaki kısmi korelasyonları hesaplayınız.

i. $r_{12,3}$

ii. $r_{13,2}$

iii. $r_{23,1}$

Çözüm 10.10 *Önce kısmi korelasyon formüllerinde kullanılacak değerleri hesaplayalım.*

X_1	5.0	4.5	4.0	3.0	2.5	19.0
X_2	16	14	15	13	14	72
X_3	10	9	8	7	6	40
X_1X_2	80	63	60	39	35	277
X_1X_3	50.0	40.5	32.0	21.0	15.0	158.5
X_2X_3	160	126	120	91	84	581
X_1^2	25.00	20.25	16.25	9.25	6.25	76.50
X_2^2	256	196	225	169	196	1042
X_3^2	100	81	64	49	36	330

Buradan,

$$\begin{aligned} r_{12} &= \frac{n\sum X_1X_2 - (\sum X_1)(\sum X-2)}{\sqrt{[n\sum X_1^2-(\sum X_1)^2][n\sum X_2^2-(\sum X_2)^2]}} \\ &= \frac{5(277)-19(72)}{\sqrt{[5(76.5)-19^2][5(1042)-72^2]}} \\ &= \frac{17}{\sqrt{(21.5)(26)}} \\ &= 0.72 \end{aligned}$$

$$\begin{aligned} r_{13} &= \frac{n\sum X_1X_3 - (\sum X_1)(\sum X-3)}{\sqrt{[n\sum X_1^2-(\sum X_1)^2][n\sum X_3^2-(\sum X_3)^2]}} \\ &= \frac{5(158.5)-19(40)}{\sqrt{(21.5)[5(330)-40^2]}} \\ &= \frac{32.5}{\sqrt{(21.5)(50)}} \\ &= 0.99 \end{aligned}$$

$$r_{23} = \frac{n\sum X_2X_3 - (\sum X_2)(\sum X-3)}{\sqrt{[n\sum X_2^2-(\sum X_2)^2][n\sum X_3^2-(\sum X_3)^2]}}$$

$$
\begin{aligned}
&= \frac{5(581) - 72(40)}{\sqrt{(26)(50)}} \\
&= \frac{25.5}{\sqrt{(26)(50)}} \\
&= 0.69
\end{aligned}
$$

olarak bulunur. Bulunan bu korelasyon katsayıları kısmi korelasyon formüllerinde yerlerine yazılırsa

i.

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0.72 - (0.99)(0.69)}{\sqrt{(1 - 0.98)(1 - 0.48)}} = 0.36$$

ii.

$$r_{13,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} = \frac{0.99 - (0.72)(0.69)}{\sqrt{(1 - 0.52)(1 - 0.48)}} = 0.99$$

iii.

$$r_{23,1} = \frac{r_{23} - r_{13}r_{12}}{\sqrt{(1 - r_{13}^2)(1 - r_{12}^2)}} = \frac{0.69 - (0.99)(0.72)}{\sqrt{(1 - 0.98)(1 - 0.52)}} = -0.23$$

elde edilir.

10.6 Problemler

1. Aşağıda verilen bilgileri kullanarak ikinci dereceden bir regresyon denklemi bulunuz.

X	0	1	2	3	4	5	6	7	8	9
$f(X)$	1.3	1.4	1.6	1.65	1.70	1.85	2.0	2.8	3.5	5.0

2. Bir firmanın 1970 – 1978 yıllarına ait kar dağılımı aşağıdaki tabloda verilmiştir.

Yıllar	1970	1971	1972	1973	1974	1975	1976	1977	1978
Kar(milyar TL)	12	14	18	23	28	30	36	42	46

Kar değişiminin lineer olduğunu kabul ederek, *i*. regresyon denklemini bulunuz,

ii. 1980 yılı karını tahmin ediniz,

iii. $1970 - 1978$ yıllarına ait tahmini karları bulunuz,

iv. Tahminin standart hatasını bulunuz,

v. %5 önem seviyesinde, regresyon katsayısının önem kontrolünü yapınız.

3. İkinci problemdeki ilk altı veriyi $(1970 - 1975)$ kullanarak,

i. regresyon denklemini bulunuz,

ii. 1980 yılı karını tahmin ediniz,

iii. %5 önem seviyesinde regresyon önem kontrolünü yapınız.

4. $1945 - 1980$ yılları arasında, bir ülkenin nüfus artışı 5 yıllık sayımlar itibariyle aşağıdaki tabloda gösterilmiştir.

Yıllar	1945	1950	1955	1960	1965	1970	1975	1980
Nüfus(milyon)	20	24	29	35	42	51	60	72

i. regresyon parabolunun denklemini bulunuz,

ii. 1982 yılı karını tahmin ediniz,

iii. Tahminin standart hatasını bulunuz.

iv. %5 önem seviyesinde regresyon önem kontrolünü yapınız.

5. Dördüncü problemin verileri için

i. lineer regresyon denklemini bulunuz,

ii. Tahminin standart hatasını bulunuz.

iii. %5 önem seviyesinde regresyon önem kontrolünü yapınız.

iv. Sonuçları, 4.problemin sonuçlarıyla karşılaştırarak yorumlayınız.

6. Yedi öğrencinin Matematik, Fizik ve Kimya derslerinden aldıkları notlar (10 üzerinden) aşağıdaki tabloda gösterilmiştir.

Matematik	5	8	3	7	4	6	8
Fizik	7	8	6	8	4	7	10
Kimya	6	8	5	7	4	6	9

Matematik ve Fizik derslerindeki başarının Kimya dersi başarısını etkilediğini düşünerek,

i. Regresyon denklemini bulunuz ($Y = a + b_1X_1 + b_2X_2$),

ii. Fizik dersinden 8, matematikten 5 alan bir öğrencinin kimya dersinden alacağı notu tahmin ediniz.

iii. Tahminin standart hatasını bulunuz.

iv. %1 önem seviyesinde regresyon önem kontrolünü yapınız.

7. Bir sınıftan alfabetik sıraya göre 10 öğrenci alınmış ve bu öğrencilerin bir derse ait bilim ve uygulama notları aşağıdaki tablo ile belirtilmiştir. Bilim ve uygulama notları arasındaki lineer korelasyon katsayısını bulunuz.

Uygulama	8	7	7	9	6	4	10	8	3	5
Bilim	9	6	5	8	7	4	9	9	1	6

8. Birinci problemin verileri için lineer determinasyon katsayısını bularak, %5 önem seviyesinde korelasyon katsayısının önem testini yapınız.

9. 8 ayrı firma tarafından üretilen belli bir eşyaya ait tüketici sıra numaraları ve fiatları aşağıdaki tabloda gösterilmiştir.

Markalar	Tüketici tercihi	Fiat (TL)
A	7	44950
B	4	52500
C	2	47995
D	6	49995
E	1	58000
F	3	54995
G	8	46995
H	5	53250

Bu verilere göre, tüketici tercihi ile malın fiatı arasında bir ilişkinin olduğu söylenebilir mi? Kararınızı, sıra korelasyon katsayısını bularak veriniz.

10. Aşağıdaki tabloda, şansa bağlı olarak seçilen 10 babanın ve bunların en yaşlı oğullarının boyları gösterilmektedir.

Baba boyu	165	160	170	163	173	157	178	168	173	170
Oğul boyu	173	168	173	165	175	168	173	165	180	170

Baba ve oğul boyları arasındaki sıra korelasyon katsayısını bulunuz.

Normal Eğri Altındaki Alan

Burada $\phi(0) = 0.0000$ olarak alınmıştır. Bazı kitaplarda z cetvelinde $\phi(0) = 0.5000$ olarak alınmtr. Bu cetvellerin biri dierine kolaylkla dntrlebilir. rnein, aadaki cetvele 0.5000 eklenerek, dier z cetveline geilir. Bu kitapta ki, baz rnekler, cetveller arasndaki iliki daha iyi anlalsn diye, farkl cetvellere gre zlmtr.

Z	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	.0398	.0438	.0478	.0517	.557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	0.832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.01179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2417	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

t-Dağılımının Yüzde Noktaları.

s.d	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.9995}$
1	0.3250	0.7270	1.376	3.078	6.3138	12.706	31.821	63.657	636.619
2	.2885	.6172	1.061	1.886	2.9200	4.3027	6.965	9.9248	31.598
3	.2766	.5840	.9780	1.638	2.3534	3.1825	4.5410	5.8409	12.9240
4	.2707	.5692	.941	1.533	2.1318	2.7764	3.747	4.6041	8.610
5	.2672	.5598	.920	1.476	2.0150	2.5706	8.365	4.0321	6.869
6	.2648	.5536	.906	1.440	1.9432	2.4469	3.143	3.7074	5.959
7	.2632	.5493	.896	1.415	1.8946	2.3646	2.998	3.4995	5.408
8	.2619	.5461	.889	1.397	1.8595	2.3060	2.896	3.3554	5.041
9	.2610	.5436	.883	1.383	1.8331	2.2622	2.821	3.2498	4.781
10	.2602	.5416	.879	1.372	1.8125	2.2281	2.764	3.1693	4.587
11	.2596	.5400	.876	1.363	1.7939	2.2010	2.718	3.1058	4.437
12	.2590	.5387	.873	1.356	1.7823	2.1788	2.681	3.0545	4.318
13	.2586	.5375	.870	1.350	1.7709	2.1604	2.650	3.0123	4.221
14	.2582	.5366	.868	1.345	1.7613	2.1448	2.624	2.9768	4.140
15	.2579	.5358	.866	1.341	1.7530	2.1315	2.602	2.9467	4.073
16	.2576	.5351	.865	1.337	1.7459	2.1199	2.583	2.9208	4.015
17	.2574	.5344	.863	1.333	1.7396	2.1098	2.567	2.8982	3.965
18	.2571	.5338	.862	1.330	1.7341	2.1009	2.552	2.8784	3.922
19	.2569	.5333	.861	1.328	1.7291	2.0930	2.539	2.8609	3.883
20	.2567	.5329	.860	1.325	1.7247	2.0860	2.528	2.8453	3.850
21	.2566	.5325	.859	1.323	1.7207	2.0796	2.518	2.8314	3.819
22	.2564	.5321	.858	1.321	1.7171	2.0739	2.508	2.8188	3.792
23	.2563	.5318	.858	1.319	1.7139	2.0687	2.500	2.9073	3.767
24	.2562	.5315	.857	1.318	1.7109	2.0639	2.492	2.7969	3.745
25	.2561	.5312	.856	1.316	1.7081	2.0595	2.485	2.7874	3.725
26	.2560	.5309	.856	1.315	1.7056	2.0555	2.479	2.7787	3.707
27	.2559	.5307	.855	1.314	1.7033	2.0518	2.473	2.7707	3.690
28	.2558	.5304	.855	1.313	1.7011	2.0484	2.467	2.7633	3.674
29	.2557	.5302	.854	1.311	1.6991	2.0452	2.462	2.7564	3.659
30	.2556	.5300	.854	1.310	1.6973	2.0423	2.457	2.7500	3.616
35	.2553	.5292	.8521	1.3062	1.6896	2.0301	2.438	2.7239	3.5919
40	.2550	.5286	.8507	1.3031	1.6839	2.0211	2.423	2.7045	3.5511
45	.2549	.5281	.8497	1.3007	1.6794	2.0141	2.412	2.6896	3.5207
50	.2547	.5278	.8489	1.2987	1.6759	2.0086	2.403	2.6778	3.4965
60	.2545	.5272	.8477	1.2959	1.6707	2.0003	2.390	2.6603	3.4606
70	.2543	.5268	.8468	1.2938	1.6669	1.9945	2.381	2.6480	3.4355
80	.2542	.5265	.8462	1.2922	1.6641	1.9901	2.374	2.6388	3.4169
90	.2541	.5263	.8457	1.2910	1.6620	1.9867	2.368	2.6316	3.4022
100	.2540	.5261	.8452	1.2901	1.6602	1.9840	2.364	2.6260	3.3909
120	.2539	.5258	.8446	1.2887	1.6577	1.9799	2.358	2.6175	3.3736
140	.2538	.5256	.8442	1.2876	1.6558	1.9771	2.353	2.6114	3.3615
160	.2538	.5255	.8439	1.2869	1.6545	1.9749	2.350	2.6070	3.3527
180	.2537	.5253	.8436	1.2863	1.6534	1.9733	2.347	2.6035	3.3456
200	.2537	.5252	.8434	1.2858	1.6525	1.9719	2.345	2.6006	3.3400
∞	.2533	.5244	.8416	1.2816	1.6449	1.9600	2.326	2.5758	3.2905

F-Dağılımının Yüzde Noktaları.

Pr							v_1						
$(F \leq f)$	v_2	1	2	3	4	5	6	7	8	9	10	12	15
0.95	1	161	200	216	225	230	234	237	239	241	242	244	246
0.975		648	800	864	900	922	937	948	957	963	969	977	985
0.99		4052	4999	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157
0.95	2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
0.975		38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4
0.99		98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
0.95	3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
0.975		17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3
0.99		34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9
0.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
0.975		12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66
0.99		21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2
0.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
0.975		10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43
0.99		16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72
0.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
0.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27
0.99		13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56
0.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
0.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57
0.99		12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	7.72	6.62	6.47	6.31
0.95	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
0.975		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10
0.99		11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
0.95	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
0.975		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77
0.99		10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
0.95	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
0.975		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52
0.99		10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
0.95	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
0.975		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18
0.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
0.95	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
0.975		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86
0.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52

χ^2 Dağılımının Kritik Değerleri

s.d	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	4.55	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

KAYNAKLAR

[**1.**] Akdeniz, F., *Olasılık ve İstatistik.* Ankara : Ankara Üni. Fen Fak., 1984.

[**2.**] Başar, A., *İstatistik.* Erzurum : Ata. Üni. İkt. İd. Bil. Fak., Ders Notu, 1995.

[**3.**] Cramer, H., *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1961.

[**4.**] Crawshaw, J. and J. Chambers, *A Concise Course in A-Level Statistics.* İkinci baskı. Bath: Stanley Thornes Ltd. 1990.

[**5.**] David, H.A., *Order Statistics.* İkinci baskı. New York: Willey, 1981.

[**6.**] Feller, W., *An Introduction to Probability Theory and Its Applications.* Cilt 1. Üçüncü Baskı. New York: Willey, 1968.

[**7.**] İdil, O., *İşletmeciler İçin Genel İstatistik.* Cilt 1. İstanbul: İ.Ü. İşletme Fak., 1989.

[**8.**] İnal, H.C. ve S. Günay, *Olasılık ve matematiksel İstatistik.* İkinci baskı. Ankara: Hacettepe Üni. Fen Fak., 1982.

[**9.**] Kreyszig, E., *Indroductory Mathematical Statistics.* New York: Willey, 1970.

[**10.**] Kreyszig, E., *Advanced Engineering Mathematics.* Yedinci baskı. New York: Willey, 1993.

[**11.**] Lipschutz, S., *Theory and Problems of Probability.* New York: McGraw-Hill Book Comp. 1968.

[**12.**] Meyer, P. L., *Indroductory Probability and Statistical Applications.* London: Addison-Wesley Pub. Comp., 1970.