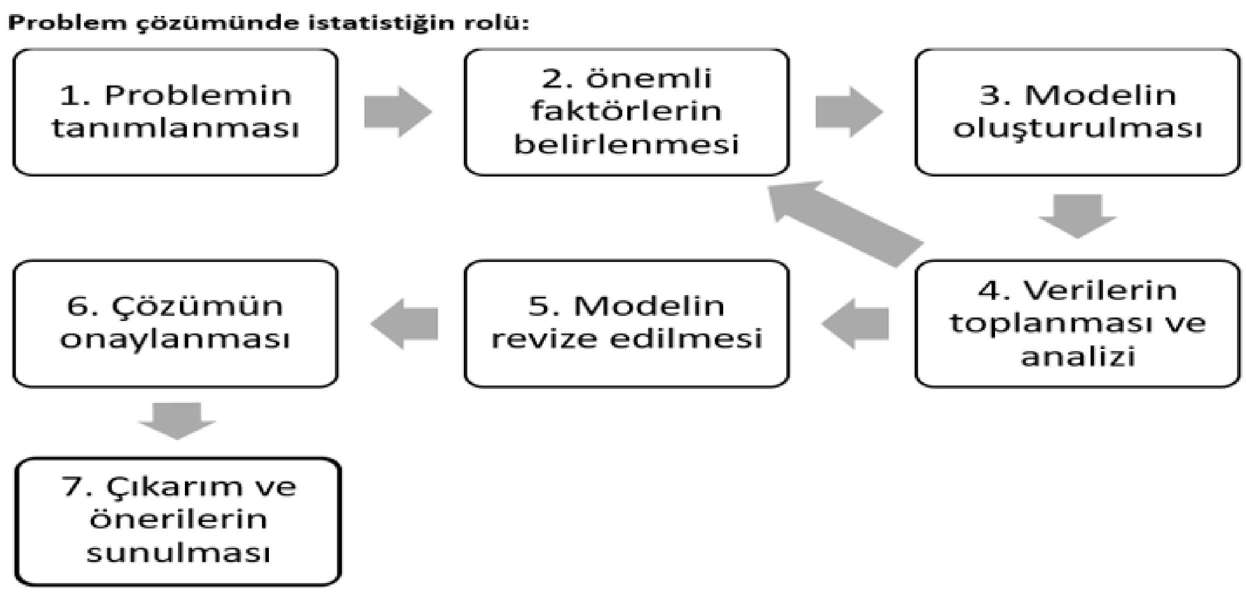
OLASILIK VE İSTATİSTİK DERS NOTU

1. TEMEL İSTATİSTİKSEL KAVRAMLAR



1.1. Tanımlamalar:

İstatistik bilimi; verilerin toplanması, düzenlenmesi, özetlenmesi, takdimi, analizi ve bu analizler vasıtasıyla elde edilen sonuçların yorumlanması ve bir karara bağlanması ile ilgilenir. Buna göre istatistiği ikiye ayırabiliriz: Betimsel istatistik ve çıkarımsal istatistik.

Frekans dağılımları, merkezi ölçüleri (aritmetik ortalama, geometrik ortalama, mod, medyan, …), dağılış ölçüleri (standart sapma, varyans, değişim aralığı, …), çarpıklık ve basıklık gibi konular verilerin özetlenmesi ve betimlenmesi ile ilgili olup betimsel istatistiğin konusudur. Örnekleme teorisi, tahmin, hipotez testleri, ilişki katsayıları ve regresyon analizi gibi konular ise çıkarımsal istatistiğin konusudur.

Kitle: Üzerinde ölçüm ve araştırmaların yapılacağı, belirli bir ya da birkaç özelliği incelenecek olan birimler topluluğudur. Örneğin; Denizli'deki üniversite öğrencilerinin aylık giderleri için kitle PAÜ öğrencileridir.

Tamsayım: Kitle hakkında tüm bilgilerin elde edilmesi kitleyi oluşturan tüm birimlerin incelenmesi demektir. Kitlenin tüm birimlerinin incelenmesi işlemine tam sayım denir.

Örneklem: Kitleye ulaşmanın mümkün olmadığı ya da yüksek maliyetler gerektirdiği durumlarda kitleden belirli yöntemlerle seçilen gözlemlerin oluşturduğu kümeye denir.

Serkan AKOĞUL

Örnekleme: Kitleden, kitle birim sayısından daha az sayıda birimin seçilmesi ile kütle hakkında tahmin yapma işlemlerine denir.

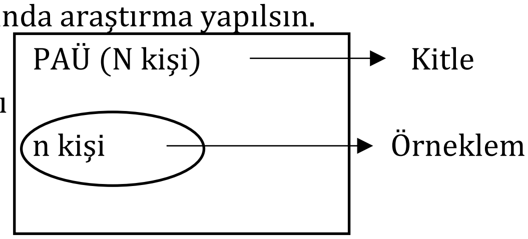
Gözlem: Belirli bir özellik hakkında bilgi ve veri toplamak için kayıt altına alınan her bir ölçüme denir.

Değişken: Gözlemden gözleme değişik değerler alabilen objelere, özelliklere ya da durumlara denir. Örneğin; cinsiyet, eğitim düzeyi, boy, kilo, aylık harcama.

Parametre ve Örneklem İstatistiği: Kitlenin tanımlayıcı sayısal değerlerine parametre örneklemin tanımlayıcı sayısal değerlerine de örneklem istatistiği denir. İstatistiksel çıkarımda amaç örneklem istatistiği kullanılarak kitle parametreleri hakkında çıkarımda bulunmaktır.

	Parametre	**Örneklem istatistiği**
Ortalama	μ	$\bar{x}$
Varyans	σ^2	s^2
Standart sapma	σ	s
Oran	Π	p
Gözlem sayısı	N	n

Örnek: PAÜ deki öğrencilerin aylık harcamaları hakkında araştırma yapılsın.
Kitle: PAÜ de kayıtlı öğrenci sayısı N=44 733
Parametre (μ): öğrencilerin ortalama aylık harcaması
Örneklem: Rasgele seçilen n=250 öğrenci
Örneklem istatistiği ($\bar{x}$): örneklemin ortalaması



1.2. Verilerin Toplanması

- **Geçmişe Dönük Araştırma:** Geçmişteki bir zaman diliminde toplamış verilerin bir kısmının ya da tamamının kullanılması durumudur. Maliyet ve zaman tasarrufu sağlamaktadır. Fakat verinin geçerliliği ve güvenirliliği şüphelidir. Zaman içerisinde gözlem kaybı yaşanmış olabilir. Veri toplanırken hatalar yapılmış olabilir. Tam amacımıza hitap etmiyor olabilir. Örnek: Eski laboratuvar çalışmasında elde edilen veriler.

- **Gözleme Dayalı Araştırma:** Bir süreç veya kitlenin belirli bir periyotta gerçekleştirilen rutin işlemler sırasın da verilerin gözlemlenip kayıt altına alınmasıdır. Örnek: Yeni kurulan elektrik santralinin bölgedeki sıcaklık değişimine etkisi.

- **Deney Tasarımı**: Sistemde yer alan kontrol edilebilir değişkenler (faktörler) üzerinde bilinçli değişiklikler yapılarak sistemin bu değişikliklere verdiği tepkiyi gözlemleyerek değişkenlerin çıktı performansının etkilerini belirleyebilmesidir. Örnek: Araç yakıt tüketimini etkileyen faktörler örneği.

1.3. Bazı Olasılığa Dayalı Örneklem Yöntemleri:

- ***Basit rasgele örneklem***: Bir kitleden her bir elemanın seçilme şansının eşit olduğu örnekleme yöntemidir. Örneğin, PAÜ aylık ortalama harcama örneği gibi rasgele 250 öğrenci seçilmesi.

- ***Sistematik örnekleme***: Örneklemi oluşturacak ilk eleman kitleden rasgele seçilir, diğer elemanların ise belirli ve aynı kural çerçevesinde seçimi ile belirlenen yöntemdir. Genel olarak (örneklem aralığı=Kitle büyüklüğü/örneklem büyüklüğü) formülü ile rasgele seçilen ilk elemana örneklem aralığı eklenerek oluşturulur. Örneğin sınıftaki 80 öğrenciden 10 kişilik örneklem seçelim. Örneklem aralığı=(80/10) dan 8 olup sınıf listesinden rasgele 2 öğrenci ilk eleman olarak seçilsin. O halde örneklemi 2, 10, 18, 26, 34, 42, 50, 58, 66 ve 74 sıra numarasına sahip öğrenciler oluşturacaktır.

- ***Tabakalı Örnekleme:*** Öncelikle kitle belirli bir çerçevede gruplandırılır ve her bir gruptan rasgele örneklem çekilerek birleştirilmesiyle elde edilen yöntemdir. Burada önemli olan gruplar kendi içinde homojen, gruplar arası ise heterojen olmalıdır. Örneğin, PAÜ deki öğrencilerin aylık ortalama harcamaları için kitleyi gruplandıralım. Tıp, Mühendislik, Fen Edebiyat, İktisadi ve İdari Bilimler, Mimarlık ve Tasarım gibi fakültelerden yoğunluklarına göre her birinden rasgele öğrenciler seçip örneklem oluşturmak.

- ***Küme örnekleme:*** Kitleyi ayrı ayrı temsil edebilen kendi içerisinde heterojen, aralarında homojen yapıya sahip kümelerin seçimine bağlı yöntemdir. Örneğin, sanayi bölgesinde tekstil fabrikalarında çalışan işçilerin çalışma koşulları üzerine araştırma yapılmak istensin. Aynı koşullara sahip her bir fabrika birer küme olmuş olup rasgele fabrikalar seçilerek örneklemi oluşturma yöntemidir.

1.4. Değişken ve değişkenlerin ölçülme düzeyleri

Değişken: Birimlerin farklı değerler alabildikleri nitelik veya niceliklerdir. Örneğin müşterilerin oluşturduğu bir kitle de; meslek, eğitim düzeyi, cinsiyet, yaş, aylık gelir gibi.

- **Nitel (Kategorik) değişkenler:** Gözlemlerin sınıflara ayrılarak açıklandığı ölçümlerdir. Cinsiyet, eğitim durumu, Kan grubu gibi.

- **Nicel (Sayısal) değişkenler:** Sayısal değerlerden oluşan ölçümlerdir. Boy, kilo, yaş, nüfus gibi. Sayısal değişkenler *kesikli* ve *sürekli* olmak üzere ikiye ayrılır. *Kesikli değişkenler*, belirli aralıkta tam değerler alabilen değişkenlerdir. Sınıftaki öğrenci sayısı, otoparktaki araba sayısı gibi. Sürekli değişkenler ise belirli aralıkta tüm değerleri alır. Ağırlık, uzunluk gibi.

  Değişkenlerin ölçümleri genel olarak dört farklı şekilde yapılabilir.

  - **Sınıflama Ölçeği**: Gözlemlerin sadece sınıflanması veya etiketlenmesiyle sınırlıdır. Sınıflar arasında büyüklük ya da sıra açısından bir ilişki bulunmaz. Kategorik değişkenlerdir. Örneğin, Cinsiyet, kan grubu, meslek grubu, Gözlük kullanımı (0: Gözlüksüz, 1: Gözlüklü) gibi.

  - **Sıralama Ölçeği**: Gözlemlerin belirli bir kritere göre sıralayarak etiketlendirilmesidir. Değişken değerleri bakımından önem sırası bulunur. Kategorik değişkenlerdir. Örneğin, Eğitim durumu (ilkokul, ilköğretim, lise, üniversite), Üniversite öğretim üyesi unvanları ve askeri rütbeler, ekonomik durum (1: çok kötü, 2: kötü, 3: orta, 4: İyi, 5: çok iyi).

  - **Eşit aralıklı Ölçek:** Bu ölçekte üzerinde durulan değişken belirli iki değer arasında sonsuz değer alabilir. Bu ölçekteki 0 değeri, ölçülen karakteristiğin olmadığını göstermez. Aynı şekilde ölçüm karakteristiklerinden biri diğerinin katlarıyla ifade edilemez. Sayısal değişkenlerdir. Örneğin, Sıcaklık ölçümleri. 0 derece, sıcaklığın olmadığını göstermez ve 4 derece, 2 derecenin iki katı değildir.

  - **Oran Ölçeği:** Zayıftan kuvvetliye doğru sıraladığımızda yukarıdaki ölçeklerin en hassas olanıdır. Ölçülen karakteristiğin 0 olması o karakteristiğin olmadığını gösterir. Aynı şekilde ölçülen bir karakteristik diğerinin katları ile ifade edilebilir. Sayısal değişkenlerdir. Örneğin; ağırlık, uzunluk, hız, gelir gibi. Sıfır ağırlığa sahip olmak yokluğu ifade ettiği gibi 50 kg mın iki katı 100 kg ifade eder. Tüm matematiksel işlemler yapılabilir.

2. FREKANS TABLOLARI

Her bir gözlemin veri setinde kaç kez tekrarlandığı ya da belirli bir aralıkta gözlemlerden kaçının bulunduğunu gösteren tablolara *frekans tabloları* denir. f ile gösterilir. n gözleme sahip veri için j inci sınıfın frekansı $f_j$ile, göreli frekansı da f_j/n ile gösterilir. Birikimli frekans ise genellikle "-den az" şeklinde yorumlanır ve her sınıfın frekansı kendinden önceki sınıfın frekansına eklenerek oluşturulur.

Bir veriyi özetlemek ve daha anlaşılır hale getirmek istediğimizde frekans tabloları kullanırız. Veride her bir X değerinin karşısına o değerin frekansı yazılır. Mesela 15 birimlik veride 5 tane farklı birim varsa bu 15 birim 5 sınıf halinde özetlenmiş olur. Farklı birimlerin çok fazla olması halinde her bir sınıf için belirli aralıklar oluşturularak bu aralığa düşen gözlem sayıları frekans olarak yazılır.

Örnek: Bir sitedeki 15 ailenin çocuk sayıları aşağıda verilmiştir. Frekans tablosu oluşturunuz.

0	2	1	3	1	2	1	2	2	1	3	0	4	3	1

Çocuk sayısı	Frekans	Birikimli frekans	Göreli frekans
0	2	2	$2/15 = 0.133$
1	5	7	$5/15 = 0.333$
2	4	11	$4/15 = 0.267$
3	3	14	$3/15 = 0.200$
4	1	15	$1/15 = 0.067$

- *<u>Sınıflandırılmış Frekans tablosu:</u>* İlgilenilen veri setinde sürekli değişken ya da fazla sayıda kesikli değişken içeren gözlemler varsa, değişkenin değer aralığı alt aralıklara bölünüp sınıflar oluşturulur ve her bir sınıfın frekansı hesaplanarak tablo oluşturulur. Sınıflandırılmış frekans tablosu için aşağıdaki adımlar sırası ile uygulanır.

  - Veri seti için n=gözlem sayısı, EB=en büyük değer, EK=en küçük değer belirlenir.
  - Sınıf sayısı k keyfi ya da $k \geq \sqrt{n}$ olacak şekilde seçilir.
  - Sınıf genişliği $h \geq \frac{EB-EK}{k}$ ile hesaplanır.
  - En küçük gözlem değeri ilk sınıfın alt limiti olarak seçilir ve bu değere sınıf genişliği ardışık olarak eklenerek diğer sınıfların alt limitleri oluşturulur.
  - Sınıflara düşen gözlem sayıları yani frekanslar belirlenir.

Örnek: 20 kişiye yapılan ankette kişilerin yaşları sorulmuş ve aşağıdaki yanıtlar elde edilmiştir. Sınıflandırılmış frekans tablosunu oluşturunuz.

35	35	32	40	48	28	35	32	35	52
48	41	47	35	22	33	34	52	20	22

- n=20 ve EB=52 ve EK=20
- Sınıf sayısı $k \geq \sqrt{20} = 4.5$ olup $k = 5$ alınır.
- Sınıf genişliği $h \geq \frac{52-20}{5} = 6.4$ olup $h = 7$ alınır.
- Sınıf alt limitler: 20, 27, 34, 41, 48 dir.
- Sınıflara düşün frekanslar tabloda verilir.

Yaş (X) için sınıflar	Frekans	Birikimli frekans	Göreli frekans
$20 \leq X < 27$ (20-27)	3	3	$^3/_{20} = 0.15$
$27 \leq X < 34$ (27-34)	4	7	$^4/_{20} = 0.20$
$34 \leq X < 41$ (34-41)	7	14	$^7/_{20} = 0.35$
$41 \leq X < 48$ (41-48)	2	16	$^2/_{20} = 0.10$
$48 \leq X < 55$ (48-55)	4	20	$^4/_{20} = 0.20$

Örnek: Bir sitedeki 27 konutun aylık doğalgaz tüketimi verileri aşağıdaki gibidir. Sınıflandırılmış frekans tablosunu oluşturunuz.

300	390	365	420	432	490	595	540	460	390	420	430	490	480
490	390	405	425	450	420	440	420	500	460	570	480	540	

- n=27 ve EB=595 ve EK=300
- Sınıf sayısı $k \geq \sqrt{27} = 5.2$ olup $k = 6$ alınır.
- Sınıf genişliği $h \geq \frac{595-300}{6} = 49.2$ olup $h = 50$ alınır.
- Sınıf alt limitler: 300, 350, 400, 450, 500, 550 dir.
- Sınıflara düşen frekanslar tabloda verilir.

Doğalgaz tüketimi	Frekans	Birikimli Frekans
300-350	1	1
350-400	4	5
400-450	9	14
450-500	7	21
500-550	3	24
550-600	3	27

3. MERKEZİ EĞİLİM ÖLÇÜLERİ

Gözlem değerlerinin etrafında toplandığı merkezi ifade eder. Genellikle duyarlı ve duyarlı olmayan ortalamalar olmak üzere iki ana başlık altınca incelenebilir.

3.1 Duyarlı (Analitik) Ortalamalar:

Veri setindeki tüm gözlem değerlerinden etkilenen ortalamalardır. Aritmetik ortalama, Geometrik ortalama ve Kareli ortalama gibi.

3.1.1 Aritmetik Ortalama: Veri setindeki gözlem değerlerinin toplamının, gözlem sayısına bölünmesiyle elde edilir. n gözleme sahip X değişkeninin aldığı değerler $x_1, x_2, \dots, x_n$ olmak üzere aritmetik ortalama ($\bar{x}$) aşağıdaki şekilde hesaplanır.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}$$

$\bar{x} = \frac{\sum x}{n}$ örneklem ortalaması olup, kitle ortalaması $\mu = \frac{\sum X}{N}$ ile gösterilir.

Örnek: 5 kişinin yaşları 20, 25, 28,21 ve 22 olmak üzere kişilerin yaş ortalaması

$$\bar{x} = \frac{20 + 25 + 28 + 21 + 22}{5} = 23.2$$

- Eğer veri seti frekans tablosuna sahip ise her bir gözlem frekansı ile çarpılarak toplam frekansa bölünerek ortalama hesaplanır.

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_j f_j}{n} = \frac{\sum(x * f)}{\sum f}$$

Örnek: Aşağıda yaşlara ilişkin verilen frekans tablosu için ortalama yaşı hesaplayınız.

Yaş	Kişi sayısı
20	4
22	3
24	2
25	5

$$\bar{x} = \frac{\sum(x * f)}{\sum f} = \frac{20 * 4 + 22 * 3 + 24 * 2 + 25 * 5}{4 + 3 + 2 + 5}$$

$$= \frac{319}{14} = 22.79$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise ilk önce her sınıf için *sınıf orta değeri=(sınıf alt limit + sınıf üst limit)/2* formülü ile bulunur. Benzer şekilde her bir sınıfın orta değeri ile sınıf frekansı çarpılıp toplam frekansa bölünerek ortalama hesaplanır. j inci sınıfın orta değeri m_j olmak üzere

$$\bar{x} = \frac{m_1 f_1 + m_2 f_2 + \dots + m_j f_j}{n} = \frac{\sum(m * f)}{\sum f}$$

Örnek: 100 öğrencinin istatistik dersinden aldıkları notlara ilişkin sınıflandırılmış frekans tablosuna göre sınıf için ortalama notu hesaplayınız.

Not	Öğrenci sayısı		Orta değer (m_j)	$m * f$
0-20	10		$\frac{0+20}{2} = 10$	10*10=100
20-40	15		$\frac{20+40}{2} = 30$	30*15=450
40-60	50		$\frac{40+60}{2} = 50$	50*50=2500
60-80	20		$\frac{60+80}{2} = 70$	70*20=1400
80-100	5		$\frac{80+100}{2} = 90$	90*5=450
Toplam	**100**			**4900**

$$\bar{x} = \frac{\sum(m * f)}{\sum f} = \frac{4900}{100} = 49$$

- Eğer veri setindeki her bir gözlem a_j gibi farklı ağırlıklara (öneme) sahip ise veri setinin ortalaması için ağırlıklı (tartılı) ortalama kullanılır. Her bir gözlem değeri ağırlığı ile çarpılarak toplam ağırlığa bölünerek ağırlıklı ortalama aşağıdaki şekilde hesaplanır.

$$\overline{x_a} = \frac{\sum(x_j * a_j)}{\sum a_j}$$

Örnek: PAÜ makine mühendisliği X isimli öğrencinin 2018 bahar yarıyılında almış olduğu dersler ve başarı notu aşağıdaki gibidir. Öğrencinin akademik ortalamasını hesaplayınız.

Dersler	AKTS	Başarı Notu	Ağırlıklı Not
Uygulamalı matematik	7	A1 (4.0)	7*4.0=28
Malzeme II	3.5	B1 (3.3)	3.5*3.3=11.55
Termodinamik II	3.5	A1 (4.0)	3.5*4.0=14
Mukavemet II	4	A1 (4.0)	4*4=16
Dinamik	3	B1 (3.3)	3*3.3=9.9
Lab II	1.5	A2 (3.7)	1.5*3.7=5.55
Sayısal Analiz	6	A1 (4.0)	6*4.0=24
Toplam	**28.5**	**26.3**	**109**

Ağırlıklı ortalama $\overline{x_a} = \frac{109}{28.5} = 3.82$ olur.

Eğer ağırlıkları hesaba katmazsak yanlış olan $\bar{x} = \frac{26.3}{7} = 3.76$ buluruz.

3.1.2 Geometrik Ortalama: Gözlem sonuçlarında mutlak farklar yerine oransal farklar ile ilgilenildiğinde geometrik ortalama (GO) kullanımı daha uygun olacaktır. G, gözlem değerlerinin birbiri ile çarpımının gözlem sayısı derecesinden köküne eşittir.

$$GO = \sqrt[n]{x_1 * x_2 * \ldots * x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

veya

$$log(GO) = \frac{1}{n}\sum_{i=1}^{n} log(x_i)$$

Örnek: Bir firma 2019, 2020, 2021 yıllarında sırasıyla 5, 10 ve 80 bin TL kar etmiş olsun. Bu firmanın karındaki ortalama büyüme oranı nedir?

Firma bir önceki yıla göre karını 2020 de 2 kat arttırmışken 2021 de 8 kat arttırmıştır. Aritmetik ortalaması $AO = (2 + 8)/2 = 5$ olup ortalama 5 kat attırdı dersek 2020 ve 2021 sırasıyla 25 ve 125 bin TL kar elle etmiş olur. Hâlbuki gerçek değerler böyle değildir.
Bu soruda geometrik ortalama kullanmak daha doğru olacaktır. $GO = \sqrt[2]{2 * 8} = \sqrt[2]{16} = 4$ olup ortalama 4 kat arttırdığında 2020 ve 2021 karları sırasıyla 20 ve 80 bin TL olacaktır. Önceki tahminlere nazaran daha iyi sonuçlar bulunmuştur.

- Eğer veri seti frekans tablosu şeklinde ise, $\sum f_j = n$ olup

$GO = \sqrt[n]{\prod x_j^{f_j}}$ veya $log(GO) = \frac{1}{n}\sum f_j log(x_j)$

- Eğer veri seti sınıflandırılmış frekans tablosu şeklinde ise, m_j sınıf orta değerleri

$GO = \sqrt[n]{\prod m_j^{f_j}}$ veya $log(GO) = \frac{1}{n}\sum f_j log(m_j)$

Örnek: Bir işyerinde çalışan 60 işçinin günlük ücret dağılımının geometrik ortalamasını hesaplayınız.

Ücret	İşçi sayısı		Orta değer (m_j)	$log\ m_j$	$f_j * log\ m_j$
15-20	20		17.5	1.2430	24.86
20-40	25		30	1.4771	36.93
40-80	15		60	1.7782	26.67
Toplam	**60**				**88.46**

$$logGO = \frac{1}{60}88.46$$
$$= 1.4743$$

$$GO = 10^{1.4743} = 29.81$$

3.1.3 Kareli ortalama: Veri setindeki değerlerin toplamı sıfır olduğunda aritmetik ortalama kullanılamaz. Yine veri setinde negatif değerlerin bulunduğu durumda da geometrik ortalama hesaplanamaz. Bu gibi durumlarda kareli ortalama kullanılır. Kareli ortalama (KO) veri setindeki değerlerin kareleri toplamının gözlem sayısına bölümünün karekökü ile hesaplanır.

$$KO = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Örnek: Aşağıdaki verinin kareli ortalamasını hesaplayınız.

x_i		x_i^2
-5		25
-3		9
0		0
2		4
6		36

$\sum x_i^2 = 74$ olup verinin kareli ortalaması

$$KO = \sqrt{\frac{74}{5}} = 3.85$$

- Eğer veri seti frekans tablosuna şeklinde ise, $\sum f_j = n$ olup

$$KO = \sqrt{\frac{\sum f_j * x_j^2}{n}}$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise $\sum f_j = n$ ve m_j sınıf orta değerleri

$$KO = \sqrt{\frac{\sum f_j * m_j^2}{n}}$$

Örnek: Aşağıdaki verinin kareli ortalamasını hesaplayınız.

Sınıflar	f_j		Orta değer (m_j)	m_j^2	$f_j * m_j^2$
5-15	3		10	100	300
15-25	8		20	400	3200
25-35	7		30	900	6300
35-45	2		40	1600	3200
Toplam	**20**				**13000**

$$KO = \sqrt{\frac{13000}{20}} = 25.5$$

3.2 Duyarsız (Analitik Olmayan) Ortalamalar:

Veri setindeki gözlem değerlerinden etkilenmeyip sıklık ve sıralamaya önem veren ortalamalardır. Mod, Medyan ve Kartiller gibi.

3.2.1 Mod (Tepe Değer): Bir veri setinde en çok tekrarlanan ya da frekansı en yüksek olan değere veri setinin Mod u denir. Her değer yalnız bir kez ya da tüm değerler eşit miktarda tekrar ediyorsa veri setinin modu yoktur denir.

Örnek: Bir mahallede yer alan 16 binaya ait binadaki daire sayıları tabloda verilmiştir. Mahalledeki daire sayısı için AO ve Mod bulunuz.

Daire sayısı	Frekans
2	4
4	2
6	7
8	1
10	2
Toplam	**16**

$$\bar{x} = \frac{\sum(x * f)}{\sum f} = \frac{2*4+4*2+6*7+8*1+10*2}{16} = 5.375$$

$$Mod = 6$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise Mod aşağıdaki formül kullanılarak hesaplanır. Frekansı en yüksek olan sınıf Mod sınıfı olmak üzere

  L_{Mod}: Mod sınıfının alt değeri

  h_{Mod}: Mod sınıfının genişliği

  Δ_1: Mod sınıfının frekansı ile bir önceki sınıfın frekansı farkı

  Δ_2: Mod sınıfının frekansı ile sonraki sınıfın frekansı farkı

$$Mod = L_{Mod} + h_{Mod}\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

Örnek: 100 bilyenin çaplarına ilişkin verilen aşağıdaki verinin mod unu bulunuz.

Bilye çapı	Bilye sayısı
10-15	10
15-20	20
20-25	30
25-30	22
30-35	18
Toplam	**100**

Frekansı en yüksek olan sınıf mod sınıfıdır yani 30 frekansa sahip 20-25 mod sınıfıdır.

$L_{Mod} = 20$ $\quad h_{Mod} = 5$

$\Delta_1 = 30 - 20 = 10$ $\quad \Delta_2 = 30 - 22 = 8$

$$Mod = 20 + 5\left(\frac{10}{10+8}\right) = 22.78$$

3.2.2 Medyan (Ortanca): Verilerin küçükten büyüğe doğru sıralandığında verilerin ortasında kalan değerdir. n gözleme sahip veri seti için eğer n tek ise $\left(\frac{n+1}{2}\right)$ inci terim medyan olur. Eğer n çift ise $\left(\frac{n}{2}\right)$ ve $\left(\frac{n}{2}+1\right)$ inci terimlerin ortalaması medyan olarak alınır.

Örnek: X ve Y veri setleri için medyanı hesaplayınız.

X	4	8	12	15	17

n=5 tek olup (5+1)/2=3 üncü terim olan 12 değeri X verisinin Medyanıdır.

Y	2	4	6	8	10	18

n=6 çift olup 6/2=3 üncü ve (6/2)+1=4 üncü terimler 6 ve 8 in ortalaması (6+8)/2=7 değeri de Y verisinin Medyanıdır.

Örnek: Aşağıda yaşlara ilişkin verilen frekans tablosu için Medyanı hesaplayınız.

Yaş	**Kişi sayısı**	**Birikimli frekans**
20	4	4
22	3	7
24	2	9
25	5	14

Toplam frekans ya da gözlem sayısı n=14 çift olup 7. ve 8. terimin ortalaması medyan olacaktır. Birikimli frekans değerlerine bakılarak 7. terim 22 olup 8. terimde 24 tür.
O halde medyan (22+24)/2=23 olur.

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise medyan aşağıdaki formül kullanılarak hesaplanır. Verilerin ortancasını içeren sınıf Medyan sınıfı olmak üzere

  n: gözlem sayısı
  L_M: Medyan sınıfının alt değeri
  h_M: Medyan sınıfının genişliği
  f_M: Medyan sınıfının frekansı
  f_B: Medyan sınıfından önceki sınıfların frekansları toplamı

$$Medyan(M) = L_M + \frac{h_M}{f_M}\left(\frac{n}{2} - f_B\right)$$

Örnek: Veri setinin Medyanını bulunuz.

Hammadde tüketimi	**Firma sayısı**	**Birikimli frekans**
5-15	80	80
15-25	100	180
25-35	200	380
35-45	520	900
45-55	100	1000
Toplam	**1000**	

n=1000 çift olup 500 üncü ve 501 inci sıradaki terimin ortalaması medyan olacaktır. Medyan sınıfı demek medyanın bulunduğu sınıf olup 500 ve 501 terim 35-45 arasında yer alacak olup bu sınıf medyan sınıfıdır. Bu terimler ayrı sınıflara düşecek olursa medyan sınıfı 500 üncü terimin bulunduğu sınıf alınır.
$L_M = 35$; $h_M = 45 - 35 = 10$; $f_M = 520$; $f_B = 380$

$$M = L_M + \frac{h_M}{f_M}\left(\frac{n}{2} - f_B\right) = 35 + \frac{10}{520}\left(\frac{1000}{2} - 380\right) = 37.3$$

Örnek: Üniversitedeki 100 öğrencinin kantinde geçirdikleri sürelerine ilişkin verilere göre veri setinin AO, M ve Mod bulunuz.

Süre(x)	Öğrenci sayısı(f)		B. Frekans	Orta değer(m)	$m * f$
20-30	10		10 (1-10)	25	250
30-40	35		45 (11-45)	35	1225
40-50	45		90 (46-90)	45	2025
50-60	5		95 (91-95)	55	275
60-70	3		98 (96-98)	65	195
70-80	2		100 (99-100)	75	150
Toplam	**100**				**4120**

Aritmetik ortalama $\bar{x} = \frac{\sum(m*f)}{\sum f} = \frac{4120}{100} = 41.20$

n=100 çift olup 50. ve 51. terimin ortalaması Medyan olacağı için medyan sınıfı 40-50 dir.

$L_M = 40$; $h_M = 50 - 40 = 10$; $f_M = 45$; $f_B = 45$;

$$M = 40 + \frac{10}{45}\left(\frac{100}{2} - 45\right) = 40.90$$

Frekansı en yüksek olan sınıf mod sınıfıdır yani 45 frekansa sahip 40-50 mod sınıfıdır.

$L_{Mod} = 40 \quad h_{Mod} = 10 \quad \Delta_1 = 45 - 35 = 10 \qquad \Delta_2 = 45 - 5 = 40$

$$Mod = 40 + 10\left(\frac{10}{10+40}\right) = 42$$

3.2.3 Çeyreklikler (Kartiller): Veri setinin değişim aralığını dört eşit parçaya bölen değerledir. Medyan verilerin küçükten büyüğe doğru sıralandığında ortada kalan değer olup ikinci çeyreklik (Q_2) olarak adlandırılır. Yani verilerin %50 si Medyanın altında %50 si de üzerindedir. Birinci çeyreklik (Q_1), verilerin ilk %25 nin altında bulunduran değerdir. Benzer şekilde üçüncü çeyreklik de (Q_3) verilerin ilk %75 altında bulunduran değerdir.

Küçükten büyüğe doğru sıralanmış veriler için hesaplanan değerler tam sayı ise Q_1 değeri $\frac{n+1}{4}$ üncü terim, Q_2 değeri $\frac{2(n+1)}{4}$ üncü terim, Q_3 değeri de $\frac{3(n+1)}{4}$ üncü terim olarak alınır. Eğer bu hesaplamalar tam sayı değilse enterpolasyon yöntemi ile çeyreklikler hesaplanır.

Örnek: 11 gözleme sahip veri setinin çeyrekliklerini bulunuz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
96	97	101	104	105	108	110	115	120	124	125

$$\frac{11+1}{4} = 3.\,gözlem \qquad Q_1 = 101$$

$$\frac{2(11+1)}{4} = 6.\,gözlem \qquad Q_2 = M = 108$$

$$\frac{3(11+1)}{4} = 9.\,gözlem \qquad Q_3 = 120$$

Örnek: 12 gözleme sahip veri setinin çeyrekliklerini bulunuz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
96	97	101	104	105	108	110	115	120	124	125	130

$$\frac{12+1}{4} = 3.25\ gözlem \qquad Q_1 = 101 + (104 - 101) * 0.25 = 101.75$$

$$\frac{2(12+1)}{4} = 6.5\ gözlem \qquad Q_2 = 108 + (110 - 108) * 0.50 = 109$$

$$\frac{3(12+1)}{4} = 9.75\ gözlem \qquad Q_3 = 120 + (124 - 120) * 0.75 = 123$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip çeyreklikler aşağıdaki şekilde hesaplanır.

  n: gözlem sayısı
  L_1: Birinci çeyreklik sınıfının alt değeri
  L_3: Üçüncü çeyreklik sınıfının alt değeri
  h_1: Birinci çeyreklik sınıfının genişliği
  h_3: Üçüncü çeyreklik sınıfının genişliği
  f_1: Birinci çeyreklik sınıfının frekansı
  f_3: Üçüncü çeyreklik sınıfının frekansı
  f_{B1}: Birinci çeyreklik sınıfından önceki sınıfların frekansları toplamı
  f_{B3}: Üçüncü çeyreklik sınıfından önceki sınıfların frekansları toplamı

$$Q_1 = L_1 + \frac{h_1}{f_1}\left(\frac{n}{4} - f_{B1}\right)$$

$$Q_3 = L_3 + \frac{h_3}{f_3}\left(\frac{3n}{4} - f_{B3}\right)$$

şeklinde hesaplanır.

Örnek: Tabloda verilen 70 öğrencinin ağırlıklarına ilişkin veri setinin birinci, ikinci ve üçüncü çeyrekliklerini bulunuz.

Ağırlıklar	Frekans	Birikimli Frekans
44-51	8	8
51-58	11	19
58-65	17	36
65-72	15	51
72-79	10	61
79-86	6	67
86-93	3	70

n=70 olup Medyan yani Q_2 35 inci ve 36 ıncı gözlemlerin ortasında olacağı için 58-65 arasında yer alacaktır.

$$M = Q_2 = L_M + \frac{h_M}{f_M}\left(\frac{n}{2} - f_B\right) = 58 + \frac{7}{17}\left(\frac{70}{2} - 19\right) = 64.59$$

Q_1 için $(70+1)/4 = 17.75$ inci gözlem 17 inci ve 18 inci gözlemlerin arasında bir değer olacağı için Q_1 birikimli frekanslara bakıldığında 51-58 arasında olacaktır.

$$Q_1 = L_1 + \frac{h_1}{f_1}\left(\frac{n}{4} - f_{B1}\right) = 51 + \frac{7}{11}\left(\frac{70}{4} - 8\right) = 57.04$$

Q_3 için $3*(70+1)/4 = 53.25$ inci gözlem 53 üncü ve 54 üncü gözlemlerin arasında bir değer olacağı için Q_3 birikimli frekanslara bakıldığında 72-79 arasında olacaktır.

$$Q_3 = L_3 + \frac{h_3}{f_3}\left(\frac{3n}{4} - f_{B3}\right) = 72 + \frac{7}{10}\left(\frac{3*70}{4} - 51\right) = 73.05 \text{ olur.}$$

4. DAĞILIŞ ÖLÇÜLERİ

Gözlem değerlerinin birbirlerine göre konumlarını, birbirine yakınlık ve uzaklıklarını yansıdan değerlerdir. Bu bölümde Değişim aralığı, Varyans ve Standart sapma, Değişim katsayısı, çeyreklik aralığı anlatılacaktır.

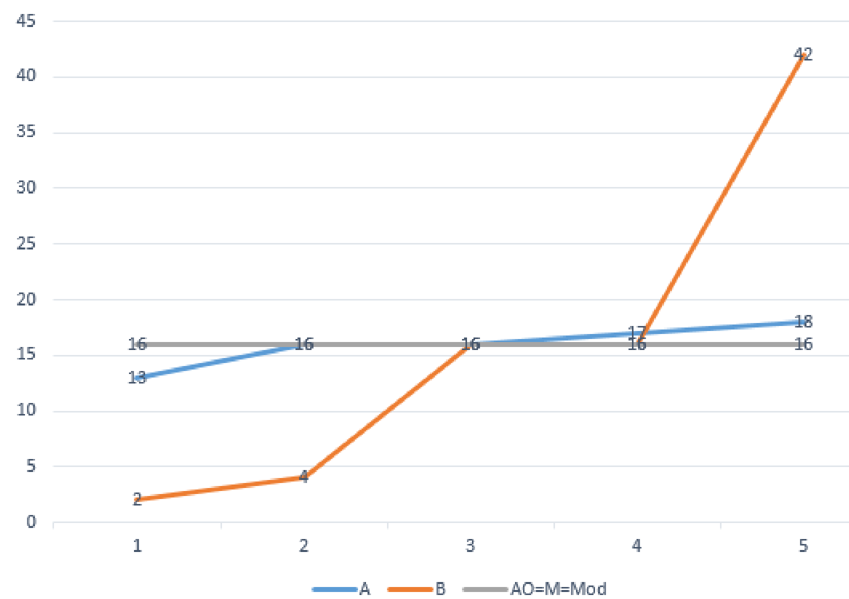
Dağılış ölçülerinin önemini daha iyi anlamak için aşağıdaki örneği inceleyelim.

Örnek: A ve B örneklemleri için AO, M ve Mod değerlerini hesaplayınız ve bu veri setlerini karşılaştırınız.

A	13	16	16	17	18
B	2	4	16	16	42

$\bar{x}_A = M_A = Mod_A = 16$

$\bar{x}_B = M_B = Mod_B = 16$



4.1 Değişim aralığı:

Değişim aralığı: Bir veri setinde en büyük (EB) değer ile en küçük değer (EK) arasındaki farktır. Verinin yayılımını gösterir. $Değişim\ Aralığı\ (DA) = EB - EK$

Örnek: $DA_A = 18 - 13 = 5$ ve $DA_B = 42 - 2 = 40$ olup B veri seti A veri setine göre yayılıma daha fazladır diyebiliriz. Yani B de değişkenlik daha fazladır.

4.2 Varyans ve Standart Sapma: Veri setindeki gözlem değerleri ile aritmetik ortalamanın farklarının karelerinin ortalamasına varyans denir ve varyansın karekökü de standart sapma olarak adlandırılır.

Kitle için varyans: $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$ olup Standart sapma: $\sigma = \sqrt{\sigma^2}$

Örneklem için varyans: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ olup Standart sapma: $s = \sqrt{s^2}$

Örnek:

$$s_A^2 = \frac{(13-16)^2 + (16-16)^2 + (16-16)^2 + (17-16)^2 + (18-16)^2}{5-1} = 3.5$$

$$s_B^2 = \frac{(2-16)^2 + (4-16)^2 + (16-16)^2 + (16-16)^2 + (42-16)^2}{5-1} = 254$$

Varyansı büyük olan veri setinde değişkenlik daha fazladır. Bu pek istenilmeyen bir durumdur. Varyansın büyük olması verinin heterojen yapıya sahip olması ve tahminlerde daha fazla sapmalar olması anlamına gelir.

- Eğer veri seti frekans tablosuna sahip ise veri setinin varyansı şeklinde hesaplanır.

$$s^2 = \frac{\sum(x-\bar{x})^2 * f}{n-1}$$

Örnek: Aşağıda yaşlara ilişkin verilen örneklemin standart sapmasını bulunuz.

Yaş	**Kişi sayısı**
20	4
22	3
24	2
25	5

$$\bar{x} = \frac{\sum(x*f)}{\sum f} = \frac{319}{14} = 22.79$$

$$s^2 = \frac{\sum(x-\bar{x})^2 * f}{n-1}$$

$$= \frac{(20-22.79)^2 * 4 + \cdots + (25-22.79)^2 * 5}{14-1} = 4.64$$

$$s = \sqrt{4.64} = 2.15$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise m_j sınıf orta değeri olmak üzere

$$s^2 = \frac{\sum(m-\bar{x})^2 * f}{n-1}$$

şeklinde hesaplanır.

Örnek: Rasgele seçilen 100 öğrencinin istatistik dersinden aldıkları notlara ilişkin sınıflandırılmış frekans tablosuna göre sınıf için ortalama notu hesaplayınız.

Not	$\boldsymbol{f}$		$\boldsymbol{m}$	$\boldsymbol{m * f}$	$(\boldsymbol{m - \bar{x}})$	$(\boldsymbol{m - \bar{x}})^{\boldsymbol{2}}$	$(\boldsymbol{m - \bar{x}})^{\boldsymbol{2}} * \boldsymbol{f}$
0-20	10		10	10*10=100	-39	1521	15210
20-40	15		30	30*15=450	-19	361	5415
40-60	50		50	50*50=2500	1	1	50
60-80	20		70	70*20=1400	21	441	8820
80-100	5		90	90*5=450	41	1681	8405
Toplam	**100**			**4900**			**37900**

$\bar{x} = \frac{\sum(m*f)}{\sum f} = \frac{4900}{100} = 49$ $\quad s^2 = \frac{\sum(m-\bar{x})^2 * f}{n-1} = \frac{37900}{100-1} = 382.83$ $\quad s = \sqrt{382.83} = 19.57$

4.3 Değişim Katsayısı (DK): Gözlemlerin birimlerinden arındırılmış bir dağılış ölçüsüdür. Birimleri farklı verilerin karşılaştırılmasında sıklıkla kullanılır. DK küçük olan veri seti değişkenlik bakımından daha homojen olduğu söylenebilir ve bu tercih edilen bir durumdur.

Kitle için: $DK = \frac{\sigma}{\mu}$ olup örneklem için: $DK = \frac{s}{\bar{x}}$

Örnek: $DK_A = \frac{1.87}{16} = 0.12$ olup $DK_B = \frac{15.94}{16} = 0.99$ olarak bulunur. $DK_A < DK_B$ olduğundan A örneklemi B ye göre daha homojen olduğu söylenebilir.

4.4 Kartil (Çeyreklik) Aralığı (KA): Veri setinin bir yayılma ölçüsü olup aykırı değerlerin tespit edilmesinde kullanılır.

$KA = Q_3 - Q_1$ ile hesaplanır. $(Q_1 - 1.5 * KA)$ ve $(Q_3 + 1.5 * KA)$ sınırlarının dışına düşen gözlem aykırı değer olarak değerlendirilir.

Örnek: 11 gözleme sahip veri setinin çeyreklikleri $Q_1 = 101$,$Q_2 = M = 108$ ve $Q_3 = 120$ ise $KA = Q_3 - Q_1 = 120 - 101 = 19$ olup Aykırı değer sınırları $(101 - 1.5 * 19) = 72.5$ ile $(120 + 1.5 * 19) = 148.5$ dir. Yani 72.5 dan az ya da 148.5 dan fazla değerler aykırı değerler olarak değerlendirilir.

5. GRAFİKLER

Veriyi özetlemenin görsel biçimidir. Genellikle kategorik değişkenler için daire ve çubuk grafiği, sayısal değişkenler için histogram, poligon ve kutu grafiği kullanılır.

5.1. Daire (Pasta) Grafiği:

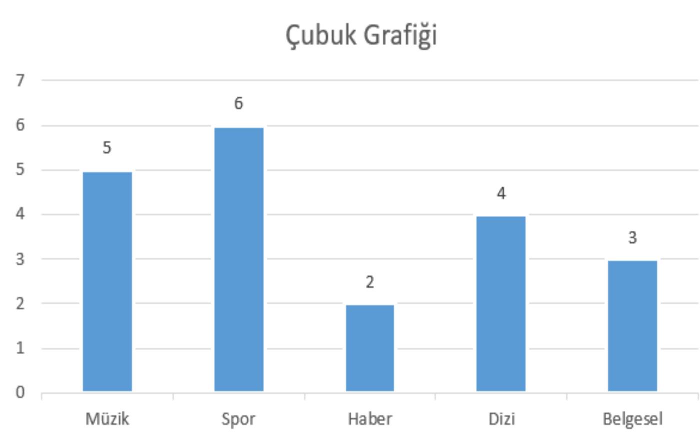
Kategorik değişkenlerin frekans dağılımlarının grafiksel gösterimi için sıklıkla kullanılır. n gözleme sahip veriler için f_j: j inci sınıf frekansı, $\frac{f_j}{n} * 100$: j inci sınıfın dairedeki yüzdelik alan, $\frac{f_j}{n} * 360$: j inci sınıfın dairedeki açı değeri olarak alınır.

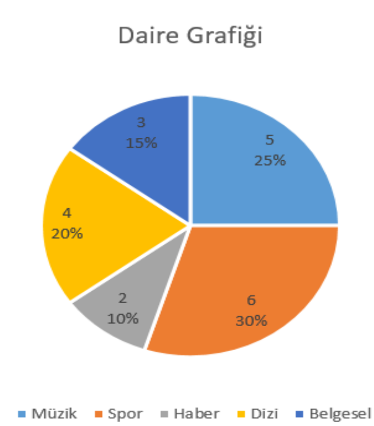
5.2. Çubuk (Sütun) grafiği:

Bu grafikte yatay eksende değişkenin kategorileri, dikey eksende ise bunların frekanslarının gösterildiği eşit genişliklere sahip çubuklardan oluşan grafiklerdir.

Örnek: Bir sınıftaki 20 öğrencinin en çok izlediği TV programına ait frekans tablosuna göre daire ve çubuk grafiğini oluşturunuz.

TV programı	Frekans	Yüzdesi	Açısı
Müzik	5	$\frac{5}{20} * 100 = 25$	$\frac{5}{20} * 360 = 90$
Spor	6	$\frac{6}{20} * 100 = 30$	$\frac{6}{20} * 360 = 108$
Haber	2	$\frac{2}{20} * 100 = 10$	$\frac{2}{20} * 360 = 36$
Dizi	4	$\frac{4}{20} * 100 = 20$	$\frac{4}{20} * 360 = 72$
Belgesel	3	$\frac{3}{20} * 100 = 15$	$\frac{3}{20} * 360 = 54$
Toplam	**20**	**100**	**360**

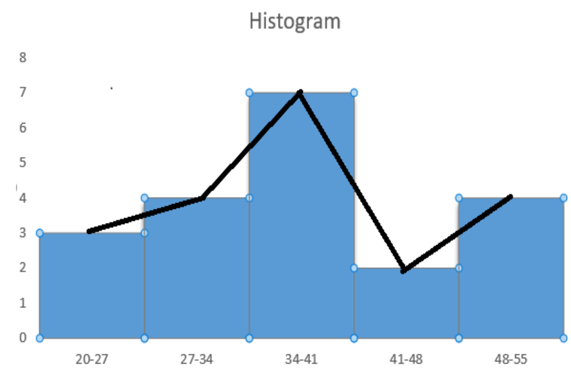


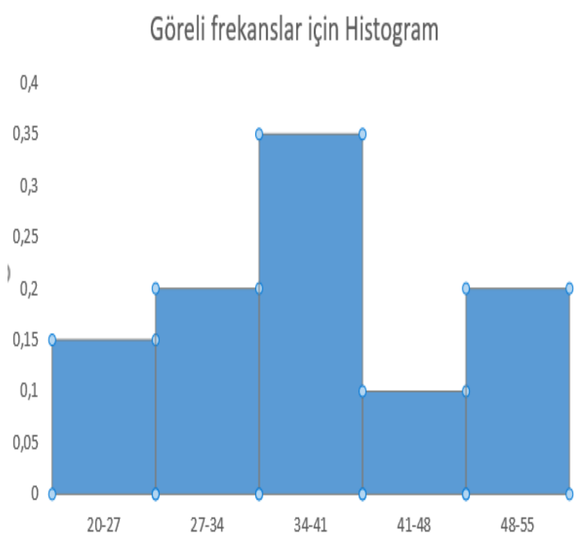


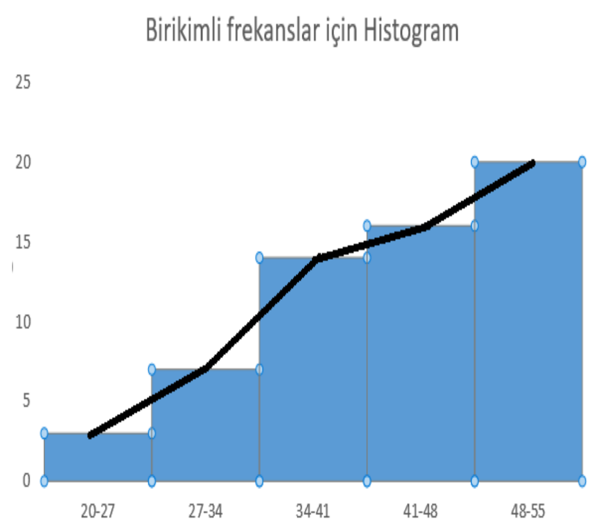
5.3. Histogram ve Poligon: Histogram bir dikdörtgenler dizisidir. Bu dikdörtgenlerin tabanları sınıflandırılmış frekans tablosundaki her bir sınıfın genişliğini, yükseklikleri ise sınıfın frekansını gösterir. Bu dikdörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe *frekans poligonu (diyagram)* denir. Göreli ya da birikimli frekans dağılımlarına ait histogramları elde etmek için bu histogramdaki dikey eksene göreli ya da veya birikimli frekanslar yerleştirilmelidir. Birikimli frekans poligonlarına *ojiv* eğrileri de denir.

Örnek: 20 kişinin yaşlarına ilişkin sınıflandırılmış frekans tablosuna göre histogram ve poligon grafiklerini çiziniz.

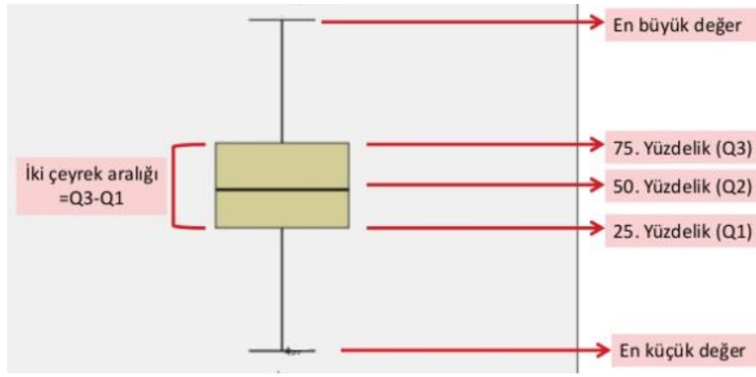
Yaş	Frekans	Birikimli frekans	Göreli frekans
20-27	3	3	$3/_{20} = 0.15$
27-34	4	7	$4/_{20} = 0.20$
34-41	7	14	$7/_{20} = 0.35$
41-48	2	16	$2/_{20} = 0.10$
48-55	4	20	$4/_{20} = 0.20$







5.4. Kutu Grafiği: Bir veri setinin merkez, yayılım, simetri ve aykırı değer gibi birçok özelliği hakkında bilgi veren kutu şeklindeki gösterimlerdir. Kutunun içerisinde yer alan medyan çizgisi verinin konumu hakkında bilgi verirken Kartil (Çeyreklik) aralığı verinin yayılımı hakkında bilgi verir.



Örnek: 12 gözleme sahip veri setinin kutu grafiğini çiziniz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
96	97	101	104	105	108	110	115	120	124	125	130

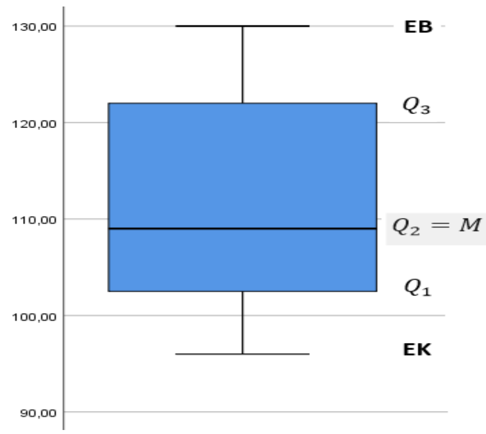
$Q_1 = 101.75$ ve $Q_2 = 109$ ve $Q_3 = 123$ olup
$KA = Q_3 - Q_1 = 123 - 101.75 = 21.25$ dir.
Aykırı değerlerin tespiti için

$$(Q_1 - 1.5 * KA) = 69.86$$

ile

$$(Q_3 + 1.5 * KA) = 154.86$$

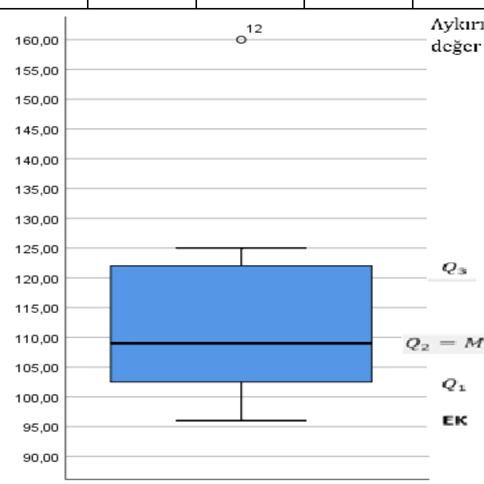
aralığının dışında kalan değerler aykırı değerlerdir.
Veri setimizde aykırı değerler bulunmamaktadır.



Örnek: 12 gözleme sahip veri setinin kutu grafiğini çiziniz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
96	97	101	104	105	108	110	115	120	124	125	160

$Q_1 = 101.75$ ve $Q_2 = 109$ ve $Q_3 = 123$ olup
$KA = 123 - 101.75 = 21.25$ dir.
69.86 *ile* 154.86 aralığının dışında kalan değerler aykırı değerlerdir.
Veri setimizde aykırı değer 12 inci gözlem olup 160 değeridir. Kutu grafiğinde aykırı değerler kutu grafiği dışında noktasal olarak gösterilir.



6. OLASILIĞA GİRİŞ

6.1 Küme Kuramı

Küme: İyi tanımlı nesneler topluluğuna küme denir. Büyük harflerle gösterilir. Kümedeki her bir nesneye de kümenin elemanı denir. Genel olarak uygulamalarda bir kümeyi belirlemek için iki yol vardır:

- Küme sonlu elemanlı ise onun elemanları parantez içinde yazarak gösterilebilir.
- Kümenim öğesi olabilecek herhangi bir nesnenin sağlamak zorunda olduğu özellik parantez içinde tanımlanarak küme belirtilebilir.

Örnek: Elemanları 2, 4, 6 ve 8 tamsayıları olan bir A kümesi:
$A = \{2, 4, 6,8\}$ şeklinde yazılır.

Örnek: B kümesi A nın elemanlarının küplerinden meydana gelmektedir. O halde B kümesi;
$B = \{8, 64, 216,512\}$ ya da;
$B = \{x^3 | x = 2, 4, 6, 8\}$ şeklinde yazılır.

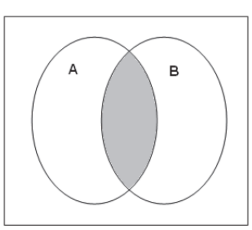
Boş Küme: Elemanı olamayan kümeye denir ve $\emptyset$ ya da $\{\quad\}$ ile gösterilir.

Alt Küme: A kümesinin her elemanı B kümesinin de elemanı ise A ya B kümesinin alt kümesi denir ve $A \subset B$ ile gösterilir. Boş küme her kümenin alt kümesidir.

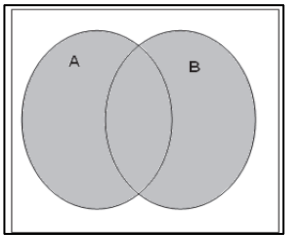
Evrensel Küme: Üzerinde çalışılan kümelerin hepsini kapsayan kümeye denir ve E ile gösterilir. Üzerinde çalışılan kümelerin her birini alt küme kabul eden evrensel kümenin;
E, n elemanlı sonlu bir küme olmak üzere E ve $\emptyset$' yi de kapsayan 2^n farklı alt kümesi vardır.

Örnek: $E = \{1, 3, 5\}$ ile evrensel kümeyi gösterelim.
Üç elemanlı bir evrensel kümenin alt kümelerinin sayısı $2^3 = 8$ olmak üzere;
$\emptyset$, $\{1\}$, $\{3\}$, $\{5\}$, $\{1, 3\}$, $\{1, 5\}$, $\{3, 5\}$, $\{1, 3, 5\}$ kümelerinin tümü E'nin alt kümeleridir.

Kesişim Kümesi: Hem A'nın hem de B'nin elemanlarının oluşturduğu kümeye denir ve
$A \cap B = \{x \mid x \in A \text{ ve } x \in B\}$ ile gösterilir.

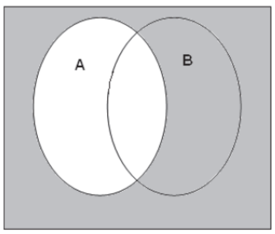


Birleşim Kümesi: A ya da B den an az birine ait elemanlarının oluşturduğu kümeye denir ve $A \cup B = \{x \mid x \in A \; veya \;\; x \in B\}$ ile gösterilir.

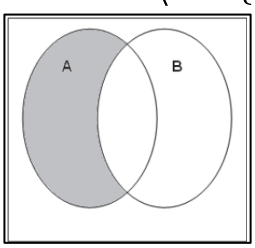


Ayrık Küme: Ortak elemanı olmayan yani $A \cap B = \emptyset$ ise A ve B kümeleri ayrıktır denir.

Küme Tümleyeni: A kümesinin tümleyeni A^t ya da A' ile gösterilir. A kümesinde bulunmayıp diğer tüm kümelerde bulunun elemanlardan oluşur. $A' = \{x \mid x \in E \; ve \; x \notin A\}$



Fark Kümesi: A kümesinde olup B kümesinde olmayan elemanlar dan oluşan kümeye denir. $A - B = A \setminus B = \{x \mid x \in A \; veya \;\; x \in B\}$ ile gösterilir.



Küme Özellikleri: A, B ve C kümeleri E evrensel kümenin alt kümeleri olmak üzere

- *Özdeşlik:* $A \cup \emptyset = A$, $\quad A \cup E = E$, $\quad A \cap \emptyset = \emptyset$, $\quad A \cap E = A$
- *Tümleme:* $A \cup A' = E$, $\quad A \cap A' = \emptyset$, $\quad (A')' = A$
- *Değişme:* $A \cup B = B \cup A$, $\quad A \cap B = B \cap A$
- *Birleşme:* $(A \cup B) \cup C = A \cup (B \cup C)$, $\quad (A \cap B) \cap C = A \cap (B \cap C)$
- *Dağılma:* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- *De Morgan:* $(A \cup B)' = A' \cap B'$, $\quad (A \cap B)' = A' \cup B'$

Örnek: Bir zar atılsın. $A = \{1, 2, 3\}$ ve $B = \{4, 5\}$ olayları tanımlansın. De Morgan Kuralının sağlandığını gösteriniz.

$$\left.\begin{array}{l} (A \cup B)' = \{6\} \\ A' = \{4, 5, 6\} \\ B' = \{1, 2, 3, 6\} \\ A' \cap B' = \{6\} \end{array}\right\} \qquad (A \cup B)' = A' \cap B'$$

$$\left.\begin{array}{l} (A \cap B)' = \{1, 2, 3,4,5, 6\} \\ A' = \{4, 5, 6\} \\ B' = \{1, 2, 3, 6\} \\ A' \cup B' = \{1, 2, 3,4,5, 6\} \end{array}\right\} \qquad (A \cap B)' = A' \cup B'$$

6.2 Sayma Teknikleri

Klasik olasılığın diğer bir ifade ile eşit olasılıklı olayların geçerli olduğu durumlarda; örnek uzayının eleman sayısı ve ilgilenilen olayın eleman sayısının belirlenmesi gerekmektedir.

Örnek uzayı ve olay sayısını belirleyen sayma tekniklerinde ise Toplama ve Çarpma kuralı olmak üzere iki temel prensip bulunmaktadır.

Toplama Kuralı: k adımlı bir işlem, birinci adım n_1 farklı şekilde, ikinci adım n_2 farklı şekilde, benzer şekilde de k ıncı adım n_k farkı şekilde yapılıyor ve bu adımlar ayrık ise bu işlem $n_1 + n_2 + \cdots + n_k$ farklı şeklilde yapılır.

Örnek: İstanbul'dan İzmir'e 2 farklı tren seferi, 4 farklı havayolu firması, 40 farklı otobüs firması ve 1 adet denizyolu firması ile gidilebildiğine göre İstanbul'dan İzmir'e kaç farklı şekilde gidilir? $2 + 4 + 40 + 1 = 47$

Çarpma Kuralı: k adımlı bir işlem, birinci adım n_1 farklı şekilden herhangi birinde yapıldıktan sonra ikinci adım n_2 farklı şekilde, benzer şekilde de k ıncı adım n_k farklı şekilde yapılıyor ve bu adımlar bağımsız ise bu işlem $n_1 * n_2 * \dots * n_k$ farklı şeklilde yapılır.

Örnek: Bir üniversitede 3 farklı tarih, 4 farklı edebiyat ve 2 farklı matematik dersinin verildiğini varsayalım.

Bir öğrenci bu derslerden sadece birini 3+4+2=9 farklı şekilde seçebilir.

Bir öğrenci her birinden birer tane seçmesi gerekiyorsa bu seçimi 3*4*2=24 farklı şekilde yapabilir.

Not: Örnek uzayı ve olay sayısının büyük olduğu durumlarda sayma yöntemleri olarak Permütasyon ve Kombinasyon kullanılmaktadır.

6.3 Faktöriyel

1 den n ye kadar olan pozitif tam sayıların çarpımına n faktöriyel denir ve $n!$ İle gösterilir.

$n! = 1 * 2 * 3 * \ldots * (n-2) * (n-1) * n$ şeklinde hesaplanır. $n! = n * (n-1)!$ şeklinde tanımlanabilir.

Burada özel olarak $0! = 1$ kabul edilir.

Örnek: $1! = 1$, $2! = 1 * 2 = 2$, $3! = 1 * 2 * 3 = 6$, $4! = 1 * 2 * 3 * 4 = 24$
$5! = 1 * 2 * 3 * 4 * 5 = 120$, $6! = 1 * 2 * 3 * 4 * 5 * 6 = 720$

6.4 Permütasyon

Bir kümenin elemanlarının bir kısmının ya da tamamının belirli bir sıralanmasına veya düzenlenmesine ***Permütasyon*** denir.

- Tümü birlikte kullanılan n farklı nesnenin permütasyonları sayısı $P(n,n) = n!$ dir.

- Bir defada r tanesi alınarak yinelemeden n farklı nesnenin permütasyonu $P(n,r) = n!/(n-r)!$ dir.

Örnek: Bir tiyatro gişesinde bilet almak isteyen 3 kişi kaç farklı şekilde sıraya girebilir.

Kişiler A, B ve C olsun. $Küme = \{A, B, C\}$ nin tüm permütasyonları ABC, ACB, BAC, BCA, CAB ve CBA olup toplam 6 farklı şekilde. $P(3,3) = 3! = 6$ dır.

Örnek: Bir haber sunucusu 5 farklı haberden 3 ünü kaç farklı şekilde sunabilir?

$$P(n,r) = \frac{n!}{(n-r)!} = \frac{5!}{(5-3)!} = 60$$

Örnek: 1,2,3,4 ve 5 rakamlarını kullanarak elde edilecek dört basamaklı sayılar için;

a. Kaç farklı sayı yazılabilir? $5 * 5 * 5 * 5 = 5^4 = 625$
b. Rakamları farklı kaç sayı yazılabilir? $P(5,4) = \frac{5!}{(5-4)!} = 120$
c. Rakamları farklı kaç tek sayı yazılabilir? $P(3,1) * P(4,3) = 3 * 24 = 72$
d. 4000 den büyük rakamları farklı kaç tane sayı yazılabilir?

$$P(2,1) * P(4,3) = 2 * 24 = 48$$

- Bir çember ya da yuvarlak masa etrafında düzenlenecek n farklı nesnenin permütasyonları sayısı $(n-1)!$ dir.

Örnek: Evli olmayan 2 Bayan ve 3 evli çift bir yuvarlak masa etrafında

a. Hiçbir koşul olmadan;

   Toplam 8 kişi olup $(8-1)! = 7! = 5040$

b. İki bekâr bayan yan yana gelmeyecek şekilde;

   Tümleyenden gidelim. İki bayan yan yana $2!$ farklı şekilde oturur. Bunları bir kişi olarak görelim toplam 7 kişi yuvarlak masa etrafına $(7-1)!$ farklı şekilde olup yan yana $2! * (7-1)!$ olur. Tüm durumdan çıkaralım. O halde yan yana gelmeyecek şekilde $(8-1)! - 2! * (7-1)! = 3600$

   Ya da 2 bekar bayan hariç geriye kalan 6 kişi yuvarlak masaya $(6-1)!$ şekilde. 6 kişinin arasındaki 6 boşluğa 2 kişi de $P(6,2)$ farklı şekilde olup istenilen durum $(6-1)! * P(6,2) = 3600$

c. Evli çiftler yan yana oturacak şekilde;

   3 evli çift yan yana $2! * 2! * 2!$ farklı şekilde, bunları bir görüp toplam 5 kişide yuvarlak masa etrafında $(5-1)!$ farklı şekilde olup istenilen durum

$$2! * 2! * 2! * (5-1)! = 192$$

- Tekrarlı Permütasyon; kümedeki nesnelerin r_1 tanesi birinci çeşit, r_2 tanesi ikinci çeşit, …, r_k tanesi k ıncı çeşit olan $r_1 + r_2 + \cdots + r_k = n$ elemanlı nesnelerin farklı permütasyonları sayısı

$$\frac{n!}{r_1! * r_2! * \ldots * \ r_k!}$$

Örnek: Duvar yapımında kullanılan özdeş 7 tuğlanın 3 ü kırmızı ve 4 ü beyaz ise bu tuğlalar; $\frac{7!}{3! * 4!} = 35$ farklı şekilde sıralanır.

Örnek: "mississippi" kelimesinin harfleri kullanılarak anlamlı ya da anlamsız yazılacak kelimelerin sayısı $\frac{11!}{1! * 4! * 2! * 4!} = 34650$ dır.

6.5 Kombinasyon

Bir defada r tanesi alınan n farklı nesnenin seçimine nesnelerin kombinasyonu denir

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

şeklinde hesaplanır. $P(n,r) = r! * C(n,r)$ olduğu açıktır.

Örnek: 4 evli çift arasından 3 kişilik bir kurul seçilecektir.

a. Her hangi bir koşul olmaksızın; $\binom{n}{r} = \binom{8}{3} = \frac{8!}{3!*(8-3)!} = 56$

b. Kurulda 2 kadın ve 1 erkek olmak zorunda; $\binom{4}{2} * \binom{4}{1} = 6 * 4 = 24$

c. Kurulda en az 2 erkek olsun; $\binom{4}{2} * \binom{4}{1} + \binom{4}{3} * \binom{4}{0} = 28$

d. Evli çiftler aynı kurulda bulunmasın; $\binom{8}{3} - \binom{4}{1}\binom{6}{1} = 32$

Örnek: $\{1,2,3,4,5,6,7,8,9\}$ kümesinin;

a. Tüm alt kümelerinin sayısı $\binom{9}{0} + \binom{9}{1} + \cdots + \binom{9}{9} = 2^9 = 512$

b. Alt kümelerinin kaçında 1 ve 2 birlikte bulunmaz?
Tüm durumdan birlikte bulunması durumlarını çıkaralım. 1 ve 2 yi kümeden çıkaralım geriye 7 eleman kalır ve bunların toplam alt küme sayısı $2^7 = 128$ olup bu kadar alt kümeye 1 ve 2 yi eklersek beraber bulunmuş olurlar. O halde istenilen durum $2^9 - 2^7 = 512 - 128 = 384$

c. 5 elemanlı alt kümelerinin kaçında 6 bulunmaz? $\binom{9}{5} - \binom{1}{1} * \binom{8}{4} = 126 - 70 = 56$

d. 6 elemanlı alt kümelerinin kaçında 8 ve 9 birlikte bulunur? $\binom{2}{2}\binom{7}{4} = 35$

6.6 Binom Teoremi

$r \leq n$ olmak üzere r ve n pozitif tam sayıları için binom açılımı

$$(a+x)^n = \binom{n}{0} a^n x^0 + \binom{n}{1} a^{n-1} x^1 + \cdots + \binom{n}{r} a^{n-r} x^r + \cdots + \binom{n}{n} a^0 x^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} x^k$$

şeklindedir. Burada $a = 1$ ve $x = 1$ alırsak $(1+1)^n = 2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}$ olur.

Örnek: $(2x-3)^5$ açılımında x^3 ün katsayısı nedir?

$$(2x-3)^5 = \sum_{k=0}^{5} \binom{5}{k} (-3)^{5-k} (2x)^k$$

x^3 ün katsayısı için $k = 3$ alalım. O halde $\binom{5}{3} (-3)^{5-3} (2x)^3 = 720x^3$

7. OLASILIK VE KOSULLU OLASILIK

17 yy.'da şans oyunlarıyla birlikte kullanılmaya başlanan olasılık, uygulamalı matematiğin bir dalı olarak gelişim göstermiş ve istatistiksel yorumlamada önemli uygulama alanı bulmuştur.

Olasılık, bir olayın meydana gelme şansının sayısal ifadesidir. Bir olayın sonucundan kesin olarak emin olunmadığında, belirli sonuçların olasılıklarından yani bu sonuçların ne kadar olası olduklarından bahsedilebilmektedir. Dolayısıyla, kesin olmayışlık ve belirsizlik gibi durumlarda karar vermede olasılık teorisi önemli rol oynamaktadır.

Olasılık, bilinmezliğin ya da belirsizliğin bir ölçüsü olup, bir durumun olması ya da olmamasının matematiksel değeridir. Şans oyunları, zar atma, kart çekme gibi.

7.1 Örnek Uzay ve Olaylar

Örnek Uzay: Bir deneyin tüm olanaklı sonuçlarının kümesine denir ve S ile gösterilir.

Olay: Örnek uzayın alt kümesine denir.

Örnek: Hilesiz bir madeni para atılması deneyinde örnek uzay $S = \{Y, T\}$ olup paranın yazı gelmesi $A = \{Y\}$ alt kümesi de bir olaydır.

Kesin ve İmkânsız olay: Bir olay, S örnek uzayının bir alt kümesi olduğundan S nin kendisi ve boş kümede birer olaydır. S olayına kesin olay, ∅ ye imkânsız olay denir.

Örnek: Hilesiz bir madeni para atılması deneyinde paranın yazı veya tura gelmesi olayı $S = \{Y, T\}$ olup kesin olaydır, paranın dik gelmesi ∅ olup imkânsız olaydır.

Ayrık ve Ayrık Olmayan olaylar: İki olay aynı anda gerçekleşemiyorsa yanı iki olayın kesişim kümesi boş küme ise bu iki olaya <u>*ayrık olay*</u>, aksi halde <u>*ayrık olmayan olay*</u> denir.

Örnek: Hilesiz bir madeni para 2 kez atılsın. A olayı bir kez yazı gelmesi, B olayı en az bir yazı gelmesi ve C olayı da yazı gelmemesi olarak tanımlansın.

$S = \{YY, YT, TY, TT\}$, $A = \{YT, TY\}$, $B = \{YY, YT, TY\}$, $C = \{TT\}$

$A \cap B = \{YT, TY\} \neq \emptyset$ olup A ile B ayrık olmayan olaylardır.

$A \cap C = \emptyset$ olup A ile C ayrık olaylardır.

$B \cap C = \emptyset$ olup B ile C ayrık olaylardır.

Bağımsız ve Bağımlı Olaylar: Bir olayın sonucu diğer olayın sonucunu etkilemiyorsa böyle olaylara <u>*bağımsız olaylar,*</u> bir olayın sonucu diğer olayın sonucunu etkiliyorsa da böyle olaylara <u>*bağımlı olaylar*</u> denir.

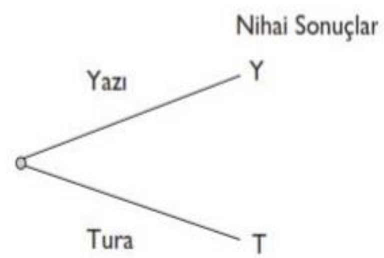
Örnek: Bir madeni para ile bir zar birlikte havaya atılıyor. Zarın üst yüzüne çift sayı gelmesi ve paranın da tura gelmesi olayı iki bağımsız olaylardır.

Örnek: Bir sınıftaki öğrencilerin isimleri kartlara yazılıp torbaya atılıyor. Çekilen kartı torbaya geri atmamak şartıyla art arda çekilen iki karttan ilki sınıf başkanı, ikincisi sınıf yardımcısı olacaktır. Başkan seçilen kişi yardımcı olamayacağı için, yani seçilen kartı tekrar seçemeyeceğimiz için (iadesiz) bu iki olay bağımlı olaylardır.

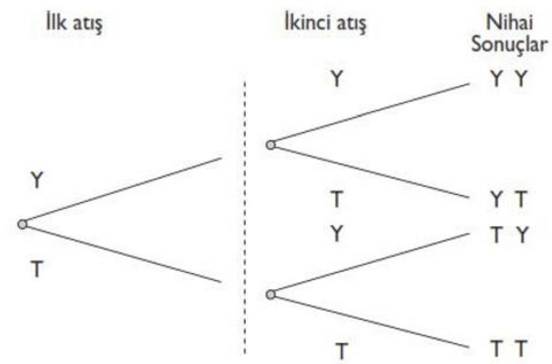
7.2 Ağaç Diyagramı

Ağaç diyagramları, bir dizi ardışık denemelerden oluşan deneyler veya olayların olası tüm sonuçlarını bulabilmek için kullanılan bir araçtır. Diyagramdaki her bir dal, ait olduğu denemeye ilişkin mümkün bir sonucu temsil eder.

Örnek: Madeni paranın bir kez ve iki kez atılması deneyleri için ağaç diyagramını çiziniz.

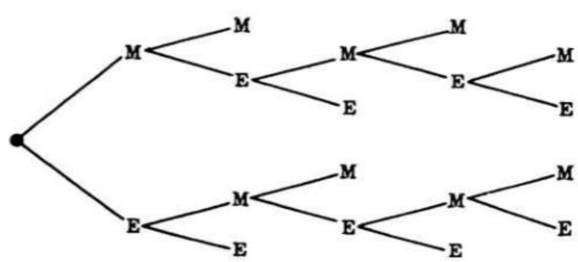


Bir kez atılması deneyi $S = \{Y, T\}$



İki kez atılması deneyi $S = \{YY, YT, TY, TT\}$

Örnek: Mert ve Emre, kendi aralarında bir satranç turnuvası düzenleyecektir. Arka arkaya iki ya da toplamda üç oyun kazanan kişi bu turnuvanın galibi olacaktır. Turnuva kaç farklı şekilde gerçekleşeceğini yazınız.



$S = \{MM, MEMM, MEMEM, MEMEE, MEE, EMM, EMEMM, EMEME, EMEE, EE\}$ olup turnuva 10 farklı şekilde gerçekleşebilir.

7.3 Olasılık ve Aksiyomları

Olasılık, bir olayın meydana gelmesi şansı olup P ile gösterilir. S örnek uzayında bir A olayının olasılığını tanımlamak için iki yol vardır.

a) **Klasik Yaklaşım:** Bir deney, eşit olasılıklı n farklı sonuç verirse ve bu sonuçların m tanesi bir A olayına uygun ise, A olayının gerçekleşmesi olasılığı

$$P(A) = \frac{m}{n} = \frac{\{uygun\ sonuçların\ sayısı\}}{\{Tüm\ sonuçların\ sayısı\}}$$

Örnek: Zarın bir kez atılması deneyinde, çift gelmesi, 6 gelmesi ve en fazla 4 gelmesi olasılıklarını bulunuz.

$S = \{1,2,3,4,5,6\}$ olup $s(S) = 6$

$A_1 = \{2,4,6\}$ ve $s(A_1) = 3$ olup $P(A_1) = \frac{3}{6} = \frac{1}{2}$

$A_2 = \{6\}$ ve $s(A_2) = 1$ olup $P(A_1) = \frac{1}{6}$

$A_3 = \{1,2,3,4\}$ ve $s(A_3) = 4$ olup $P(A_1) = \frac{4}{6} = \frac{2}{3}$

b) **Göreli frekans yaklaşım**: Eşit olasılıklı olmayan sonuçların olasılık hesabında, n kez tekrarlanan deneyde f kez A olayı gözlemleniyorsa, A olayının gerçekleşmesi olasılığı $P(A) = f/n$ dir.

Örnek: A marka üründen rasgele seçilen 200 tanesinin 10 tanesi bozuk olduğuna göre, rasgele seçilen A marka ürünün sağlam olması olasılığı nedir.

Bozuk ürün 10 olup Sağlam 190 ürün vardır. O halde $P(\text{Sağlam}) = \frac{190}{200} = 0.95$ dir. Bozuk olması olasılığı $P(\text{Bozuk}) = \frac{10}{200} = 0.05 = 1 - P(\text{Sağlam})$ dır.

Olasılık Aksiyomları: S bir deneyin örnek uzayı ve A_i ler de örnek uzayda olaylar olsun.

- $P(A_i) \geq 0$
- $P(S) = 1$
- A_i ler ikişerli ayrık olaylar olmak üzere $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$

aksiyomlarını sağlayan P olasılık fonksiyonu, $P(A_i)$ de A_i olayının olasılığı olarak adlandırılır.

Bazı Olasılık Kuralları: S örnek uzayında A ve B birer olay olmak üzere;

- Herhangi bir A olayı için $0 \leq P(A) \leq 1$, $P(S) = 1$ ve $P(\emptyset) = 0$ dır.
- Eğer $A \subset B$ ise $P(A) \leq P(B)$ dir.
- Herhangi bir $A \subset S$ olayı için $P(A) \leq 1$.
- A olayının tümleyeni A' olmak üzere, A nın gerçekleşmemesi olasılığı $P(A')$ olup $P(A') = 1 - P(A)$.

Örnek: Bir zar atılsın. $S = \{1,2,3,4,5,6\}$ dir. A olayı zarın 6 gelmesi olayı ise $A = \{6\}$ ve $A' = \{1,2,3,4,5\}$ olup $P(A) = 1/6$ ve $P(A') = 1 - P(A) = 1 - 1/6 = 5/6$ dır.

- A ile B olaylarının birleşimi $A \cup B$ ve kesişimi $A \cap B$ olmak üzere $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ dir.

- A ve B ayrık olaylar ise $A \cap B = \emptyset$ olup $P(A \cap B) = 0$ ve $P(A \cup B) = P(A) + P(B)$.

Örnek: Bir zar atılsın $S = \{1,2,3,4,5,6\}$ dir. A olayı zarın tek gelmesi, B olayı 4 den büyük gelmesi ve C olayı 2 gelmesi olsun.

$A = \{1,3,5\}$, $B = \{5,6\}$ ve $C = \{2\}$ olup $A \cap B = \{5\}$, $A \cap C = \emptyset$, $B \cap C = \emptyset$,
$A \cup B = \{1,3,5,6\}$, $A \cup C = \{1,2,3,5\}$, $B \cup C = \{2,5,6\}$
$P(A) = 3/6$, $P(B) = 2/6$, $P(C) = 1/6$
$P(A \cap B) = 1/6$, $P(A \cap C) = 0$, $P(B \cap C) = 0$
$P(A \cup B) = 4/6$, $P(A \cup C) = 4/6$, $P(B \cup C) = 3/6$
$\frac{4}{6} = P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6}$
$\frac{4}{6} = P(A \cup C) = P(A) + P(C) = \frac{3}{6} + \frac{1}{6}$
$\frac{3}{6} = P(B \cup C) = P(B) + P(C) = \frac{2}{6} + \frac{1}{6}$

- Eğer B olayının gerçekleştiği biliniyorken A olayının olma olasılığına ***koşullu olasılık*** denir ve $P(A|B) = \frac{P(A \cap B)}{P(B)}$ dir. Benzer şekilde A olayının gerçekleştiği biliniyorken B olayının olma olasılığı da $P(B|A) = \frac{P(A \cap B)}{P(A)}$ dir.

- $P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$ dir.

- Bir olayın ortaya çıkması diğer olayın ortaya çıkma olasılığını etkilemiyorsa bu iki olaya ***bağımsız olaylar*** denir ve A ve B olayları bağımsız ise $P(A|B) = P(A)$ ve $P(B|A) = P(B)$ olur. Böylece $P(A \cap B) = P(A) * P(B)$ dir. Eşitliliğin sağlanmadığı durumda olaylar ***bağımlıdır***.

Örnek: Bir depodaki 100 ürünün üretildiği makine ve kalite durumuna göre dağılımı tablodaki gibidir.

		Ürün Kalitesi		
		Sağlam	**Bozuk**	**Toplam**
Makine	**X**	48	12	60
	Y	32	8	40
	Toplam	80	20	100

a) $P(X) = \frac{60}{100} = 0.6$, $P(Y) = \frac{40}{100} = 0.4$, $P(S) = \frac{80}{100} = 0.8$, $P(B) = \frac{20}{100} = 0.2$

b) $P(X|B) = \frac{P(X \cap B)}{P(B)} = \frac{12/100}{20/100} = \frac{12}{20} = 0.6 = P(X)$ olduğundan X ile B olayları bağımsızdırlar. $P(X \cap B) = 0.12 = P(X) * P(B) = 0.6 * 0.2$ elde edilir.

c) $P(Y|S) = \frac{P(Y \cap S)}{P(S)} = \frac{32/100}{80/100} = \frac{32}{80} = 0.4 = P(Y)$ olduğundan Y ile S olayları bağımsızdırlar. $P(Y \cap S) = 0.32 = P(Y) * P(S) = 0.4 * 0.8$ elde edilir.

Örnek: 400 öğrencinin cinsiyet ve istatistik dersi başarı durumuna göre dağılımı tablodaki gibidir.

		Cinsiyet		
		Erkek(E)	Kadın(K)	Toplam
Başarı Durumu	Başarılı(B)	80	20	100
	Başarısız(BS)	100	200	300
	Toplam	180	220	400

a) $P(E) = \frac{180}{400}$, $P(K) = \frac{220}{400}$, $P(B) = \frac{100}{400}$, $P(BS) = \frac{300}{400}$

b) $P(E \cap B) = \frac{80}{400}$, $P(E \cap BS) = \frac{100}{400}$, $P(K \cap B) = \frac{20}{400}$, $P(K \cap BS) = \frac{200}{400}$,

c) $P(E \cup B) = \frac{180}{400} + \frac{100}{400} - \frac{80}{400} = \frac{200}{400}$, $P(K \cup BS) = \frac{220}{400} + \frac{300}{400} - \frac{200}{400} = \frac{320}{400}$

d) Öğrencinin dersten başarılı olduğu biliniyorken erkek olması olasılığı

$$P(E|B) = \frac{P(E \cap B)}{P(B)} = \frac{80/400}{100/400} = \frac{80}{100}$$

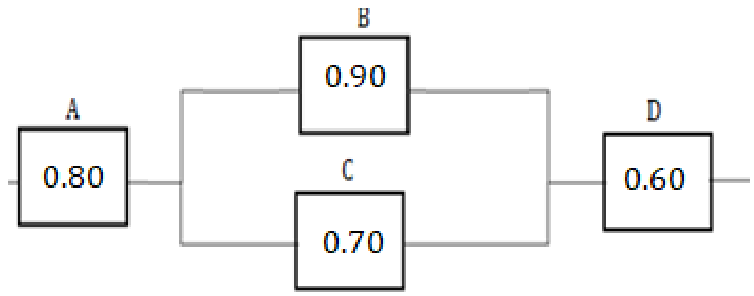
e) Öğrencinin kadın olduğu biliniyorken başarısız olması olasılığı

$$P(BS|K) = \frac{P(BS \cap K)}{P(K)} = \frac{200/400}{220/400} = \frac{200}{220}$$

f) $P(E) = \frac{180}{400} = 0.45$ ve $P(E|B) = \frac{80}{100} = 0.80$ olduğundan Erkek ile Başarı olayları bağımlıdır. $P(E \cap B) = 0.20 \neq P(E) * P(B) = 0.1125$

g) $P(BS) = \frac{300}{400} = 0.75$ ve $P(BS|K) = \frac{200}{220} = 0.91$ olup Kadın ile Başarısız olayları bağımlıdır. $P(BS \cap K) = 0.50 \neq P(BS) * P(K) = 0.4125$

Örnek: Bir elektrik sistemine ilişkin diyagrama göre bileşenlerin başarı davranışları bağımsız ise sistemin çalışma olasılığı nedir.



Sistemin Çalışması olasılığı $Ç = A \cap (B \cup C) \cap D$ olayının olasılığıdır.

$$P(Ç) = P(A \cap (B \cup C) \cap D) = P(A) * P(B \cup C) * P(D)$$

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0.90 + 0.70 - (0.90 * 0.70) = 0.97$$

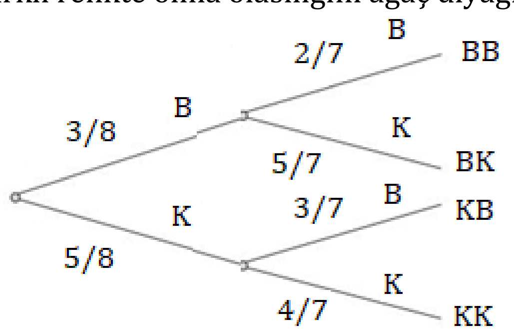
ya da

$$P(B \cup C) = 1 - P(B \cup C)' = 1 - P(B' \cap C') = 1 - P(B') * P(C')$$
$$= 1 - (0.10 * 0.30) = 0.97$$
$$P(Ç) = P(A) * P(B \cup C) * P(D) = 0.80 * 0.97 * 0.60 = 0.4656$$

II. yol:

$$Ç = A \cap (B \cup C) \cap D = [(A \cap B) \cup (A \cap C)] \cap D = (A \cap B \cap D) \cup (A \cap C \cap D)$$
$$P(Ç) = P(A \cap B \cap D) + P(A \cap C \cap D) - P(A \cap B \cap D \cap A \cap C \cap D)$$
$$= P(A) * P(B) * P(D) + P(A) * P(C) * P(D) - P(A) * P(B) * P(C) * P(D)$$
$$= 0.80 * 0.90 * 0.60 + 0.80 * 0.70 * 0.60 - 0.80 * 0.90 * 0.70 * 0.60$$
$$= 0.4656$$

Örnek: Bir torbada 3 beyaz ve 5 kırmızı top vardır. Bu torbadan rasgele art arda iki top çekilsin. Her iki topunda farklı renkte olma olasılığını ağaç diyagramını çizerek bulunuz.

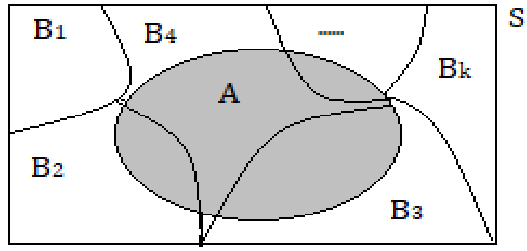


$S = \{BB, BK, KB, KK\}$ olup istenilen durum BK ya da KB dallarının gerçekleşmesidir. Dallar ayrık olayları simgelediği için istenilen olasılık

$$P(BK \cup KB) = P(BK) + P(KB) = \frac{3}{8} * \frac{5}{7} + \frac{5}{8} * \frac{3}{7} = \frac{15}{28} = 0.5357$$

7.4 Örnek Uzayın Parçalanışı, Toplam Olasılık ve Bayes Teoremi

7.4.1 Örnek uzayın parçalanışı:



- $B_i \cap B_j = \emptyset \ (i = 1, \dots, j, \dots k \ ve \ i \neq j)$
- $\bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \dots \cup B_k = S$
- $P(B_i) > 0$ koşulunu sağlayan B_i ler S örnek uzayının bir parçalanışı olup $P(B_1) + P(B_2) + \cdots + P(B_k) = 1$ dir

7.4.2 Toplam Olasılık Formülü: B_i ler S örnek uzayının bir parçalanışı ise S deki herhangi bir A olayının gerçekleşmesi olasılığı

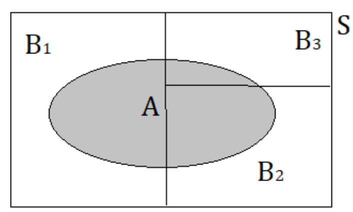
$$P(A) = \sum_{i=1}^{k} P(B_i) * P(A|B_i)$$

7.4.3 Bayes Teoremi: A olayının gerçekleştiği biliniyorken B_r olayının koşullu olasılığı

$$P(B_r|A) = \frac{P(B_r \cap A)}{P(A)} = \frac{P(B_r) * P(A|B_r)}{\sum_{i=1}^{k} P(B_i) * P(A|B_i)}$$

Örnek: Conta üreten bir fabrikada toplam üretimin %50 si B_1 makinesinde, %30 u B_2 makinesinde ve kalanı da B_3 makinesinde gerçekleşmektedir. Bu makinelerin üretimde sırasıyla %5 i, %4 ü ve %3 ü kusurlu gerçekleşmektedir.

a) Bir günlük üretimin ardından rasgele seçilen bir contanın kusurlu olması
b) Rasgele seçilen contanın kusurlu olduğu biliniyorken, bu contanın B_2 makinesinde üretilmiş olması olasılıklarını bulunuz.



$B_1 = \{B_1 \ de \ üretilmiş \ olması\}$
$B_2 = \{B_2 \ de \ üretilmiş \ olması\}$
$B_3 = \{B_3 \ de \ üretilmiş \ olması\}$
$S = B_1 \cup B_2 \cup B_3$ dir.
$A = \{Kusurlu \ olması\}$ olayı olsun.

$$P(A) = \sum_{i=1}^{3} P(B_i) * P(A|B_i) = P(B_1) * P(A|B_1) + P(B_2) * P(A|B_2) + P(B_3) * P(A|B_3)$$
$$= \frac{50}{100} * \frac{5}{100} + \frac{30}{100} * \frac{4}{100} + \frac{20}{100} * \frac{3}{100} = \frac{430}{10000} = 0.0430$$

$$P(B_2|A) = \frac{P(B_2)*P(A|B_2)}{P(A)} = \frac{\frac{30}{100}*\frac{4}{100}}{\frac{430}{10000}} = \frac{12}{43} = 0.2791$$

8. RASGELE DEĞİŞKENLER

Örnek uzaydaki her bir eleman, reel sayılarda tanımlı bir değişkenin değeri olarak alınırsa böyle değişkenlere *rasgele değişken (rd)* denir.

Bir rasgele değişkenin alabileceği değerler tam sayı, sonlu ya da sayılabilir sonsuz ise Kesikli rd, bir aralıktaki tüm değerleri ya da sayılamaz sonsuzlukta değerler alıyorsa Sürekli rd olarak adlandırılır. Rd ler X,Y,Z gibi büyük harflerle alabileceği değerlerde x,y,z gibi küçük harflerle gösterilir.

Örnek:

i. X rd bir paranın iki kez atılmasın deneyinde elde edilen turaların sayısı olmak üzere $S = \{YY, TY, YT, TT\}$ olup X Kesikli rd alabileceği değerler $X = \{x_1 = 0, x_2 = 1, x_3 = 2\}$ dir.
ii. Y rd iki zarın yuvarlanması deneyinde üste gelen sayıların toplamı olmak üzere $S = \{(1,1), (1,2), \dots (6,6)\}$ olup Y Kesikli rd alabileceği değerler $Y = \{y_1 = 2, y_2 = 3, \dots, y_{11} = 12\}$ dir.
iii. Z Kesikli rd bir dersi alan öğrenci sayısı ise $Z = \{0,1,2, \dots\}$
iv. T yeni doğan bir sağlıklı bebeğin ağırlığı (gr) ise T sürekli rd $3000 \leq t \leq 3500$
v. Yükseklik, ağırlık, sıcaklık, uzunluk, hacim gibi ölçüm ya da tartı ile elde edilen değişkenler sürekli rd dir.

8.1 Kesikli Rasgele Değişkenler ve Dağılımları

8.1.1 Kesikli RD Olasılık Fonksiyonu:
X kesikli rd nin her bir x olası değerinin gerçekleşme olasılığı $P(X = x) = f(x)$ şeklinde ifade edilir.

$$0 \leq f(x_i) \leq 1$$
$$\sum f(x_i) = 1$$

özelliklerini sağlayan $f(x)$, X kesikli rd nin *olasılık fonksiyonu (of)* dur.

8.1.2 Kesikli RD Birikimli Dağılım Fonksiyonu:
X kesikli rd nin bir x değerine eşit ve küçük olması olasılığı $P(X \leq x) = F(x)$ şeklinde ifade edilir ve $F(x)$ e X kesikli rd nin *birikimli dağılım fonksiyonu (bdf)* denir. X in BDF

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

$$0 \leq F(x) \leq 1$$

$$x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$$

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

özelliklerini sağlar.

8.1.3 Kesikli RD Beklenen Değeri (Ortalaması): X rd nin beklenen değeri $E(X)$ ile gösterilir ve ağırlıklı ortalama ya da kitle ortalaması (μ) olarak da adlandırılır. X kesikli RD nin beklenen değeri

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

şeklindedir.

Y, X kesikli rd nin bir fonksiyonu $Y = g(x)$ ise $E(Y) = E(g(x)) = \sum_{i=1}^{N} g(x_i) f(x_i)$ dir.
$a, b \in \mathbb{R}$ için $g(x) = aX + b$ ise $E(g(x)) = E(aX + b) = E(aX) + E(b) = aE(X) + b$

8.1.4 Kesikli RD Varyansı: Varyans, bir rasgele değişkenin aldığı değerlerin ortalamadan ne kadar saptığının bir yayılım ölçüsüdür. X rd nin varyansı $V(X)$ ile gösterilir. Kitle varyansı (σ^2) olarak da ifade edilir. X kesikli rd nin varyansı

$$\sigma^2 = V(X) = E[(x-\mu)^2] = \sum_{i=1}^{N} (x_i - \mu)^2 f(x_i)$$

$$\sigma^2 = V(X) = \sum_{i=1}^{N} x_i^2 f(x_i) - \mu^2 = E(X^2) - \left(E(X)\right)^2$$

şeklindedir. Standart sapma varyansın karekökü olup $\sigma = \sqrt{V(X)}$ dir.

Y, X kesikli rd nin bir fonksiyonu $Y = g(x)$ ise $V(Y) = E\left[\left(g(x) - E\left(g(x)\right)^2\right)\right]$ dir.

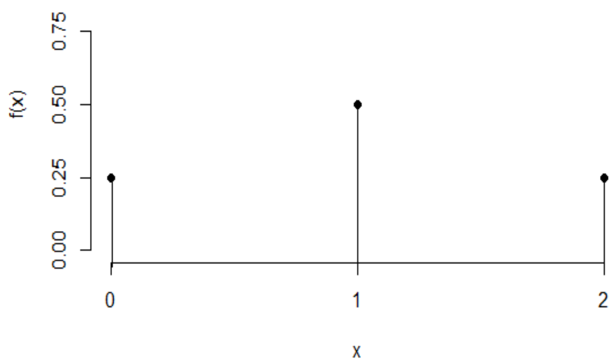
$a, b \in \mathbb{R}$ için $g(x) = aX + b$ ise $V(g(x)) = V(aX + b) = V(aX) + V(b) = a^2 V(X) + 0$

Örnek: X rasgele değişkeni bir paranın iki kez atılmasın deneyinde elde edilen turaların sayısı olsun. X kesikli bir rd dir.
$S = \{YY, TY, YT, TT\}$ olup $X = \{x_1 = 0, x_2 = 1, x_3 = 2\}$ dir.

X kesikli rd nin of $f(x)$ ve grafiği

$x_1 = 0$ (hiç tura gelmemesi), $\{YY\}$ olayı olup $\quad P(X = 0) = f(0) = 1/4 = 0.25$
$x_2 = 1$ (bir tura gelmesi), $\{TY, YT\}$ olayı olup $\quad P(X = 1) = f(1) = 2/4 = 0.50$
$x_3 = 2$ (iki tura gelmesi), $\{TT\}$ olayı olup $\quad P(X = 2) = f(2) = 1/4 = 0.25$

$0 \le f(x_i) \le 1$ ve $\sum_{i=1}^{3} f(x_i) = f(0) + f(1) + f(2) = 1$ olup $f(x) = \begin{cases} 0.25 & ; \ x = 0 \\ 0.50 & ; \ x = 1 \\ 0.25 & ; \ x = 2 \\ 0 & ; \ d.d. \end{cases}$



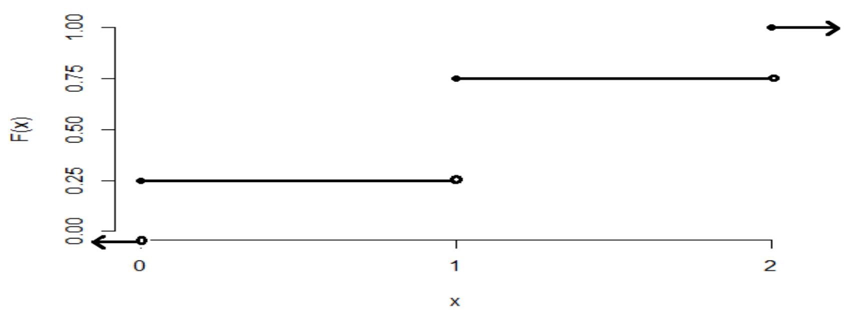
X kesikli rd bdf $F(x) = \sum_{x_i \le x} f(x_i)$ ve grafiği

$x_1 = 0$ için $\quad P(X \le x_1 = 0) = F(0) = \sum_{x_i \le 0} f(x_i) = f(0) = \frac{1}{4} = 0.25$

$x_2 = 1$ için $P(X \le x_2 = 1) = F(1) = \sum_{x_i \le 1} f(x_i) = f(0) + f(1) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4} = 0.75$

$x_3 = 2$ için $P(X \le x_3 = 2) = F(2) = \sum_{x_i \le 2} f(x_i) = f(0) + f(1) + f(2) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$

$$F(x) = \begin{cases} 0 & ; \quad x < 0 \\ 0.25 & ; \ 0 \le x < 1 \\ 0.75 & ; \ 1 \le x < 2 \\ 1 & ; \quad 2 \le x \end{cases}$$



X kesikli rd beklenen değeri ve varyansı;

$$\mu = E(X) = \sum_{i=1}^{3} x_i f(x_i) = 0 * 0.25 + 1 * 0.50 + 2 * 0.25 = 1$$

$$\begin{aligned} \sigma^2 = V(X) &= \sum_{i=1}^{3} (x_i - \mu)^2 f(x_i) = (0-1)^2 * 0.25 + (1-1)^2 * 0.50 + (2-1)^2 * 0.25 \\ &= \sum_{i=1}^{N} {x_i}^2 f(x_i) - \mu^2 = (0^2 * 0.25 + 1^2 * 0.50 + 2^2 * 0.25) - 1^2 \\ &= 0.5 \end{aligned}$$

Örnek: X rd nin olasılık fonksiyonu $x = 1,2,3,4$ için $f(x) = \frac{1}{k} * x$ olduğuna göre;

a) $k = ?$
b) $F(x)$
c) $P(X = 2)$
d) $P(X = 1.5)$
e) $P(X \leq 3)$
f) $P(X > 2.5)$
g) $E(X)$
h) $V(X)$

değerlerini bulunuz.

Çözüm: $f(x)$ olasılık fonksiyonu ise $0 \leq f(x_i) \leq 1$ olup $\sum_{i=1}^{4} f(x_i) = 1$ olmalı.

$$\sum_{i=1}^{4} f(x_i) = \frac{1}{k} * 1 + \frac{1}{k} * 2 + \frac{1}{k} * 3 + \frac{1}{k} * 4 = 1 \Rightarrow \frac{10}{k} = 1 \Rightarrow k = 10$$

olup $f(x) = \begin{cases} \frac{1}{10} * x & ;\ x = 1,2,3,4 \\ 0 & ;\ d.d. \end{cases}$

a) $k = 10$

b) $F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{i=1}^{x} \frac{1}{10} x = \frac{1}{10}(1 + 2 + \cdots + x) = \frac{1}{10}\left(\frac{x(x+1)}{2}\right)$ olup

$$F(x) = \begin{cases} 0 & ;\quad x < 1 \\ \dfrac{x(x+1)}{20} & ;\ 1 \leq x < 4 \\ 1 & ;\quad 4 \leq x \end{cases}$$

c) $P(X = 2) = f(2) = \frac{1}{10} * 2 = 0.2$

d) $P(X = 1.5) = f(1.5) = 0$

e) $P(X \leq 3) = f(1) + f(2) + f(3) = 0.1 + 0.2 + 0.3 = 0.6$

$P(X \leq 3) = F(3) = \frac{3(3+1)}{20} = 0.6$

f) $P(X > 2.5) = f(3) + f(4) = 0.3 + 0.4 = 0.7$

$P(X > 2.5) = 1 - P(X \leq 2.5) = 1 - F(2.5) = 1 - F(2) = 1 - \frac{2(2+1)}{20} = 0.7$

g) $E(X) = \mu = \sum_{i=1}^{4} x_i f(x_i) = 1 * 0.1 + 2 * 0.2 + 3 * 0.3 + 4 * 0.4 = 3$

h) $V(X) = \sum_{i=1}^{4} {x_i}^2 f(x_i) - \mu^2 = (1^2 * 0.1 + 2^2 * 0.2 + 3^2 * 0.3 + 4^2 * 0.4) - 3^2 = 1$

8.2 Bazı Önemli Kesikli Dağılımlar

Rastlantısal olgulara veya süreçlere ilişkin bir araştırma yaparken elimizdeki rasgele değişkenin olasılık dağılımını bilmek varsayımlarda ve hesaplamalarda büyük kolaylık sağlamaktadır. Bu bölümde kesikli dağılımlardan Binom, Geometrik ve Poisson dağılımları incelenecektir.

8.2.1 Binom Dağılımı: Her bağımsız denemede sadece iki olası sonuca ulaşılabilen deneylerin modellenmesinde Binom dağılımı kullanılabilir. Örneğin madeni paranın atılması deneyi ya yazı ya da tura gelmesiyle sonuçlanması, üretilen bir ürünün kusurlu ya da kusursuz olması, bir soruya sadece evet ve hayır cevabının verilmesi gibi.

X rd birbirinden bağımsız n denemede başarılı olanların sayısı ve p de her bir bağımsız denemedeki başarı olasılığı olmak üzere binom dağılımı

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \; x = 0,1,\dots,n$$
$$E(X) = \mu = n * p$$
$$\sigma^2 = V(X) = n * p * (1-p)$$

dir. $X \sim Bin(n,p)$ ile gösterilir.

Örnek: Madeni bir para 4 kez atılsın.

a) Üç tura gelmesi olasılığı nedir?
b) En az bir tura gelmesi olasılığı nedir?
c) En fazla iki tura gelmesi olasılığı nedir?
d) Bu deney çok fazla tekrar edilirse ortalama kaç kez tura gelmesi beklenir?

Çözüm: Deney iki olası durumla sonuçlanabilir Tura-Yazı. X rd elde edilen turaların sayısı olup istenilen durum bizim için başarı olacağından tura gelmesi olasılığı $p = 1/2$ ve $n = 4$ olup $X \sim bin(4, 1/2)$ dir. X rd olasılık fonksiyonu

$$P(X = x) = f(x) = \binom{4}{x} (1/2)^x (1-1/2)^{4-x} \quad ; \; x = 0,1,2,3,4$$

a) $P(X = 3) = f(3) = \binom{4}{3} (1/2)^3 (1/2)^{4-3} = \frac{1}{4} = 0.25$

b) $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - f(0)$

$$= 1 - \binom{4}{0} (\tfrac{1}{2})^0 \left(\tfrac{1}{2}\right)^{4-0} = \frac{15}{16} = 0.9375$$

c) $P(X \leq 2) = f(2) + f(1) + f(0) = \frac{11}{16} = 0.6875$

d) $E(X) = \mu = 4 * 1/2 = 2$

Örnek: Bir bilgisayar donanım imalatçısı firmanın imal ettiği çiplerin %10 unun arızalı gerisinin sağlam olduğu bilinmektedir. Rasgele seçilen 10 ürünün;

a) Sadece 8 inin sağlam olması olasılığı nedir?
b) En fazla 2 ürünün sağlam olması olasılığı nedir?
c) Çipin sağlam sayısının beklenen değeri nedir?

Çözüm: Deney iki olası durumla sonuçlanabilir Sağlam-Arızalı. X rd sağlam ürünlerin sayısı olup istenilen durum bizim için başarı olacağından sağlam olması olasılığı $p = 0.90$ ve $n = 10$ olup $X \sim bin(10, 0.90)$ dir. X rd olasılık fonksiyonu

$$P(X = x) = f(x) = \binom{10}{x} (0.90)^x (1 - 0.90)^{10-x} \quad ; \; x = 0,1,\dots,10$$

a) $P(X = 8) = f(8) = \binom{10}{8} (0.90)^8 (0.10)^{10-8} = 0.19$
b) $P(X \leq 2) = f(2) + f(1) + f(0) = 3.6 * 10^{-7}$
c) $E(X) = \mu = 10 * 0.90 = 9$

8.2.2 Geometrik Dağılım: Tekrarlanan bağımsız bernoulli denemelerinde p başarı olasılığı olmak üzere *ilk başarılı sonuca ulaşılan* deneme sayısı olan X rd geometrik dağılıma sahip olup olasılık fonksiyonu

$$P(X = x) = f(x) = p * (1 - p)^{x-1} \quad , x = 1,2,3,\dots$$

şeklinde olup $E(X) = 1/p$ ve $V(X) = \frac{(1-p)}{p^2}$ dır. $X \sim Geo(p)$ ile gösterilir.

Örnek: Bir avcının hedefi vurması olasılığı 0.20 dir. Bu avcının hedefi beşinci atışta ilk defa vurması olasılığı nedir.

Çözüm: $p = 0.20$ olup $X \sim Geo(0.20)$ dir. $f(x) = 0.20 * (1 - 0.20)^{x-1} \; , x = 1,2,3,\dots$

$$f(5) = 0.20 * (1 - 0.20)^{5-1} = 0.0819$$

Örnek: Bir firmanın ürünlerinin %5 inin kusurlu olduğu bilinmektedir. Montaj hizmeti sağlayan bir bayii, bu firmanın ürünlerinde ilk kez 8 inci montajda kusurlu ürünle karşılaşması olasılığı nedir.

Çözüm: $p = 0.05$ olup $X \sim Geo(0.05)$ dir. $f(x) = 0.05 * (1 - 0.05)^{x-1} \; , x = 1,2,3,\dots$

$$f(8) = 0.05 * (1 - 0.05)^{8-1} = 0.0349$$

8.2.3 Poisson Dağılımı: Belirli bir zaman diliminde veya belirli bir mekânda bağımsız ve rasgele gerçekleşen olayların modellenmesinde poisson dağılımı kullanılabilir. Örneğin; bir kavşakta mesai saatleri içerisinde meydana gelen kazaların sayısı, bir havaalanına her saat inen uçakların sayısı, üretilen bir bakır telin uzunluğunun belirli bir aralıktaki kusur sayısı gibi.

İlgili olayın belirli bir aralıktaki ortalama tekrar sayısını λ biliniyor olsun. O halde poisson dağılımına sahip X rd olasılık fonksiyonu

$$P(X=x)=f(x)=\frac{\lambda^x e^{-\lambda}}{x!} \quad , \qquad x=0,1,2,\ldots$$

şeklinde olup $E(X)=V(X)=\lambda$ dır. $X \sim Poi(\lambda)$ ile gösterilir.

Örnek: Üretilen bakır bir kabloda milimetre başına ortalama 2 kusur bulunduğu bilinmektedir.

a) Bir milimetre kabloda tam 3 kusur bulunması olasılığı nedir?
b) Dört milimetre kabloda tam 10 kusur bulunması olasılığı nedir?
c) İki milimetre kabloda en az 2 kusur bulunması olasılığı nedir?

Çözüm: Belirli bir uzunluktaki kusur sayısını poisson dağılımı ile modelleyebiliriz. Bir milimetredeki ortalama kusur sayısı 2 olup

a) 1mm kabloda ortalama 2 kusur $\lambda=2$ olup $f(3)=\frac{2^3 e^{-2}}{3!}=0.1804$
b) 4mm kabloda ortalama 8 kusur $\lambda=8$ olup $f(10)=\frac{8^{10} e^{-8}}{10!}=0.0993$
c) 2mm kabloda ortalama 4 kusur $\lambda=4$ olup
$P(X\geq 2)=1-P(X<2)=1-\big(f(0)+f(1)\big)=0.9084$

Örnek: Belirli bir limana her gün ortalama olarak 10 petrol tankeri gelmektedir. Limandaki tesisler günde en fazla 15 tankerin işlemini yapabilmektedir. Belirli bir günde tankerlerin geri çevrilmesi olasılığı nedir?

Çözüm: Her gün gelen tankerlerin sayısı X rd ve ortalama tanker sayısı $\lambda=10$ olup $X \sim Poi(10)$ olur. Tankerlerin geri çevrilmesi durumu gelen tankerlerin 15 in üzerinde olması durumunda olacağı için istenilen olasılık

$$P(X>15)=1-P(X\leq 15)=1-\sum_{x=0}^{15}\frac{10^x e^{-10}}{x!}=0.0487$$

8.3 Sürekli Rasgele Değişkenler ve Dağılımları

X sürekli rd ise x in alabileceği değerler aralık ya da aralıklar kümesidir.
$-\infty < x < \infty,\qquad a < x \leq b,\qquad x < c,\qquad d \leq x$ gibi.

8.3.1 Sürekli RD Olasılık Yoğunluk Fonksiyonu: D_x , X sürekli rd için tanım kümesi olmak üzere

$$0 \leq f(x) \leq 1$$
$$\int_{D_x} f(x)dx = 1$$

özelliklerini sağlayan $f(x)$ e *olasılık yoğunluk fonksiyonu (oyf)* denir. X rd nin a ile b arasında olması olasılığı $P(a \leq x \leq b) = \int_a^b f(x)dx$ dir.

$$P(X = a) = P(a \leq x \leq a) = \int_a^a f(x)dx = 0$$

$$P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b) = \int_a^b f(x)dx$$

$$P(x \leq c) = P(x < c) = \int_{D_x}^c f(x)dx$$

$$P(d \leq x) = P(d < x) = \int_d^{D_x} f(x)dx$$

8.3.2 Sürekli RD Birikimli Dağılım Fonksiyonu: X sürekli rd nin bir x değerine eşit ve küçük olması olasılığı $P(X \leq x) = F(x)$ şeklinde ifade edilir ve $F(x)$ e X sürekli rd nin *birikimli dağılım fonksiyonu (bdf)* denir. X in BDF

$$P(X \leq x) = F(x) = \int_{D_x}^x f(x)dx$$
$$f(x) = \frac{dF(x)}{dx}$$
$$0 \leq F(x) \leq 1$$
$$x_1 \leq x_2 \ \Rightarrow F(x_1) \leq F(x_2)$$
$$P(a \leq x \leq b) = F(b) - F(a)$$
$$P(X > x) = 1 - P(X \leq x) = 1 - F(x)$$

özelliklerini sağlar.

8.3.3 Sürekli RD Beklenen Değeri (Ortalaması): X sürekli RD nin beklenen değeri

$$\mu = E(X) = \int_{D_x} xf(x)dx$$

şeklindedir.

Y, X sürekli rd nin bir fonksiyonu $Y = g(x)$ ise $E(Y) = E(g(x)) = \int_{D_x} g(x)f(x)dx$ dir.
$a, b \in \mathbb{R}$ için $g(x) = aX + b$ ise $E(g(x)) = E(aX + b) = E(aX) + E(b) = aE(X) + b$

8.3.4 Sürekli RD Varyansı: X sürekli rd nin varyansı

$$\sigma^2 = V(X) = \int_{D_x} (x - \mu)^2 f(x)dx = \int_{D_x} x^2 f(x)dx - \mu^2 = E(X^2) - \big(E(X)\big)^2$$

şeklindedir. Standart sapma varyansı karekökü olup $\sigma = \sqrt{V(X)}$ dir.

Y, X sürekli rd nin bir fonksiyonu $Y = g(x)$ ise $V(Y) = E\left[\left(g(x) - E\big(g(x)\big)^2\right)\right]$ dir.

$a, b \in \mathbb{R}$ için $g(x) = aX + b$ ise $V(g(x)) = V(aX + b) = V(aX) + V(b) = a^2V(X) + 0$

Örnek: X rd ni ince bir bakır telde ölçülen akımı göstersin. $f(x) = 0.1$ ve $0 \leq x \leq 10$ olup;

a) Ölçülen akımın 5 ma (mili amper) den az olması olasılığı nedir?
b) Ölçülen akımın 4 ile 8 ma arasında olması olasılığı nedir?
c) Ölçülen akımın 7 ma den fazla olması olasılığı nedir?
d) $F(x)$ bulunuz.
e) $E(X)$ ve $V(X)$ bulunuz.

Çözüm: $D_x = [0,10]$ aralığı olup $0 \leq f(x) = 0.1 \leq 1$ ve $\int_0^{10} 0.1dx = 1$

a) $P(x < 5) = \int_0^5 0.1dx = 0.5$
b) $P(4 < x < 8) = \int_4^8 0.1dx = 0.4$
c) $P(x > 7) = \int_7^{10} 0.1dx = 0.3$
d) $P(X \leq x) = F(x) = \int_0^x 0.1dx = 0.1x$
e) $E(X) = \int_0^{10} x * 0.1\, dx = 5$ ve $V(X) = \int_0^{10} x^2 * 0.1\, dx - 5^2 = 8.3333$

Örnek: X rd olasılık yoğunluk fonksiyonu $f(x) = \begin{cases} kx & 0 < x < 2 \\ 0 & d.d \end{cases}$ olsun.

a) k=?
b) $F(x)$ =?
c) $P(1 \leq X < 3/2)$ =?
d) $P\left(X > 1/2 \big/ X < 1\right)$ =?
e) $Y = 3X + 2$ rasgele değişkeni için $E(Y)$ yi hesaplayınız.

Çözüm: $D_x = (0,2)$ aralığı olup $0 \leq f(x) \leq 1$ ve $\int_0^2 kx\,dx = 1$ olmalı. $k = \frac{1}{2}$ olur.

a) $k = 1/2$
b) $F(x) = \int_0^x \frac{x}{2} dx = \frac{x^2}{4}$
c) $P(1 \leq X < 3/2) = \int_1^{3/2} \frac{x}{2}\,dx = 0.3125$
d) $P\left(X > 1/2 \big/ X < 1\right) = \frac{P\left(\frac{1}{2} \leq X < 1\right)}{P(X<1)} = \frac{0.1875}{0.25} = 0.75$
e) $E(X) = \int_0^2 x * \frac{x}{2} dx = 1.3333$ dir. $Y = 3X + 2 \Rightarrow E(Y) = 3E(X) + 2 = 3 * 1.3333 + 2 = 6$

Örnek: X rd olasılık yoğunluk fonksiyonu $f(x) = \begin{cases} kx^2 & 1 \leq x \leq 4 \\ 0 & d.d \end{cases}$ olsun.

a) k=?
b) $F(x)$ =?
c) $P(2 \leq X < 3)$ =?
d) $E(X)$
e) $V(X)$.

Çözüm: $\int_1^4 kx^2\,dx = 1$ olmalı. $k = \frac{1}{21}$ olur. $f(x) = \frac{1}{21}x^2$

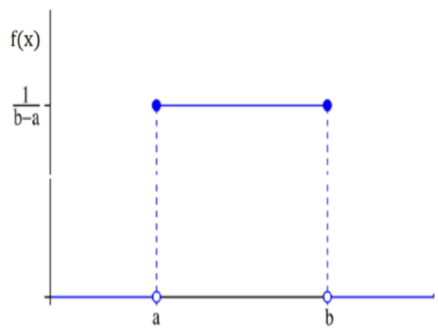
a) $k = 1/21$
b) $F(x) = \int_1^x \frac{1}{21}x^2 dx = \frac{x^3-1}{63}$
c) $P(2 \leq X < 3) = \int_2^3 \frac{1}{21}x^2\,dx = \frac{19}{63} = 0.3016$
d) $E(X) = \int_1^4 x * \frac{1}{21}x^2 dx = \frac{85}{28} = 3.0357$
e) $V(X) = E(X^2) - E(X)^2 = \int_1^4 x^2 * \frac{1}{21}x^2 dx - E(X)^2 = \frac{341}{35} - \left(\frac{85}{28}\right)^2 = 0.5274$

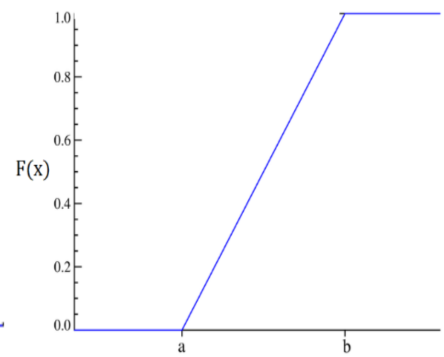
8.4 Bazı Önemli Sürekli Dağılımlar

8.4.1 Düzgün (Uniform) dağılım: Tanımlı olduğu aralıktaki her sürekli değer için aynı sabit olasılık dağılımların modellenmesinde kullanılır. X sürekli rd nin olasılık yoğunluk fonksiyonu

$$f(x) = \frac{1}{b-a} \ , \qquad a \le x \le b$$

ile verilen X rd, [a,b] aralığında düzgün dağılmıştır denir. $E(X) = \frac{a+b}{2}$ ve $V(X) = \frac{(b-a)^2}{12}$ dır. $X \sim Uni(a,b)$ ile gösterilir. X rd oyf ve bdf aşağıdaki şekildedir.





Örnek: X rd [-1,1] aralığında düzgün dağılıma sahip olduğuna göre, X rd nin 0.25 ile 0.75 arasında olması olasılığı ve varyansını bulunuz.

Çözüm: $[\mathrm{a,b}] = [-1,1]$ olup $f(x) = \frac{1}{1-(-1)} = \frac{1}{2} = 0.5$ dir.

$$P(0.25 \le X \le 0.75) = \int_{0.25}^{0.75} 0.5dx = 0.25$$

$$E(X) = \int_{-1}^{1} x * 0.5\, dx = 0$$

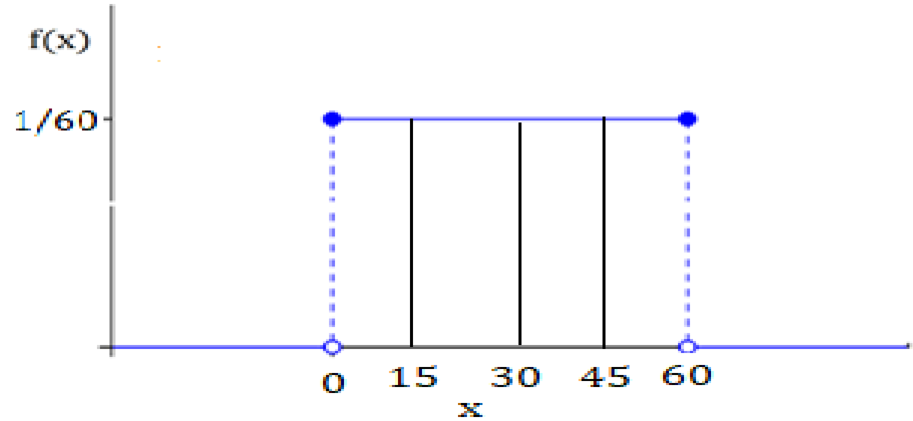
$$V(X) = E(X^2) - E(X)^2 = \int_{-1}^{1} x^2 * 0.5dx - 0^2 = \frac{1}{3} = 0.3333$$

$E(X) = \frac{a+b}{2} = \frac{-1+1}{2} = 0$ ve $V(X) = \frac{(b-a)^2}{12} = \frac{(1-(-1))^2}{12} = \frac{1}{3}$

Örnek: Denizli belediyesi 320 Üniversite-Karahasanlı hat otobüsü hafta içi 08.00-18.00 arası belirli bir durağa 15 dk bir gelmektedir. Bir yolcu durağa 10.00 ile 11.00 arasında düzgün dağılmış bir zamanda varırsa,

a) Otobüs için 5 dk dan az
b) Otobüs için en az 12 dk beklemesi olasılıklarını bulunuz.

Çözüm: Saat 10.00 ile 11.00 arasındaki 1 saatlik diliminde X sürekli rd ni yolcunun durağa geliş zamanı (dk) olsun. O halde $[a,b] = [0,60]$ aralığında $f(x) = \frac{1}{60-0} = \frac{1}{60}$ olur. O tobüsün durağa gelme zamanı (10.00-10.15-10.30-10.45-11.00) dır.



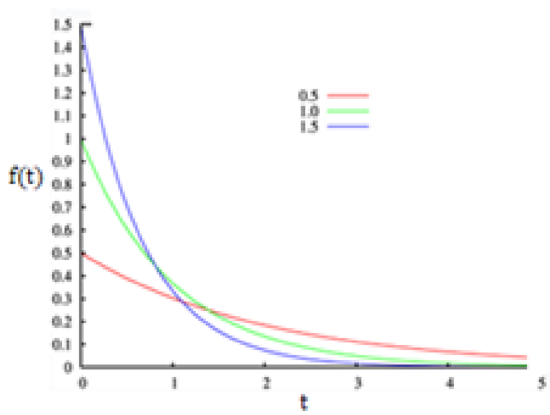
a) 5 dk az beklemesi için
$$P(10 < X < 15) + P(25 < X < 30) + P(40 < X < 45) + P(55 < X < 60)$$
$$= \frac{5}{60} + \frac{5}{60} + \frac{5}{60} + \frac{5}{60} = \frac{20}{60} = 0.3333$$

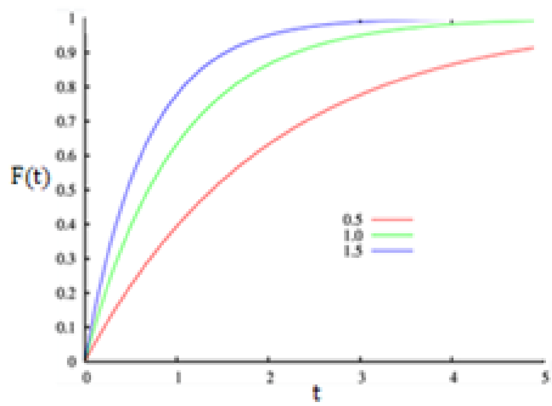
b) En az 12 dk beklemesi
$$P(0 < X < 3) + P(15 < X < 18) + P(30 < X < 33) + P(45 < X < 48)$$
$$= \frac{3}{60} + \frac{3}{60} + \frac{3}{60} + \frac{3}{60} = \frac{12}{60} = 0.20$$

8.4.2 Üstel dağılım: Belirli bir olayın gerçekleşmesine kadar geçen sürenin ya da iki olay arasında geçen sürenin modellenmesinde kullanılır. Örneğin, Bir bankaya gelen müşterilerin arasındaki bekleme süresi, Bir acil servise gelen hastaların arasındaki geçen süre, bir bakır teldeki iki kusur arasındaki uzunluk gibi. T sürekli rd nin olasılık yoğunluk fonksiyonu

$$f(t) = \frac{1}{\alpha} e^{-\frac{t}{\alpha}} , \qquad t > 0$$

ile verilen T rd üstel dağılıma sahiptir. $E(T) = \alpha$ ve $V(T) = \alpha^2$ dır. $T \sim üstel(\alpha)$ ile gösterilir. Üstel dağılımın bdf $F(t) = \int_0^t \frac{1}{\alpha} e^{-\frac{t}{\alpha}} \, dt = 1 - e^{-\frac{t}{\alpha}}$ olup, T rd farklı α parametresine oyf ve bdf grafikleri aşağıdaki şekildedir.





Örnek: Bir taksi durağına gelen müşteriler arasında ortalama bekleme süresi 5 dk dır. Durağa gelen bir müşterinin 10 dk fazla beklemesi olasılığı nedir?
Çözüm: T rd M1 ile M2 arasında bekleme süresi, ortalaması $E(T) = \alpha = 5$ ve $V(T) = 25$ olup $T \sim üstel(5)$ olur.

$$f(t) = \frac{1}{5} e^{-\frac{t}{5}} \text{ olup } P(T > 10) = \int_{10}^{\infty} \frac{1}{5} e^{-\frac{t}{5}} dt = \frac{1}{5} \frac{e^{-\frac{t}{5}}}{-\frac{1}{5}} \Bigg|_{10}^{\infty} = (-0) - (-e^{-2}) = 0.135$$

- **Poisson ve Üstel dağılım arasındaki ilişki:** X rd belirli bir zaman aralığında ilgili olayın ortaya çıkma sayısı ve λ ortalaması olsun. O halde $X \sim Poi(\lambda)$ olup $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ dir.

T poisson dağılmış bir rasgele değişkenin biri diğerini izleyen iki olay arasındaki süre olarak tanımlanırsa $T \sim üstel(\alpha = \frac{1}{\lambda})$ olur. $f(t) = \lambda e^{-\lambda t}$

Örnek: Bir acil servise her 15 dakikada ortalama 3 hasta gelmektedir. Bu servise gelen hastalar arası geçen sürenin 8 dk. veya daha fazla çıkması olasılığı nedir?

Çözüm: X rd 1 dk içinde gelen hasta sayısı olsun. Birim zamanda gelen ortalama hasta sayısı $\lambda = 3\, hasta\ /15\, dk = 0.2\, hasta/\, dk$. O halde T rd iki hasta arasında geçen süre ve ortalaması $\alpha = 1/\lambda = 1/0.2 = 5\, dk$ olup $f(t) = \frac{1}{5} e^{-\frac{t}{5}}$ dir. Ya da $F(t) = 1 - e^{-\frac{t}{5}}$

$$P(T \geq 8) = \int_8^\infty \frac{1}{5} e^{-\frac{t}{5}} dt$$

$$= 1 - P(T < 8) = 1 - F(8) = 1 - \left(1 - e^{-\frac{8}{5}}\right) = 0.2019$$

Örnek: Bir kafeteryada müşterilere hizmet verme süresi ortalaması 4 dk dır. Bu kişinin 7 günün 5 gününde 3 dk dan az bir sürede hizmet verme olasılığını bulunuz?

Çözüm: T rd hizmet için geçen süre ve ortalaması $\alpha = 4$ dk olup $f(t) = \frac{1}{4} e^{-\frac{t}{4}}$ ya da $F(t) = 1 - e^{-\frac{t}{4}}$ dir. Önce bu kişinin 3dk dan daha az sürede hizmet vermesi olasılığını bulalım.

$$P(T < 3) = F(3) = 1 - e^{-\frac{3}{4}} = 0.5276$$

X rd 7 gün içinde 3 dk dan az bir sürede hizmet görülen gün sayısı olsun. Bağımsız 7 gün içinde 3 dk dan az bir sürede hizmet görülmesi başarı durumu olup olasılığı $p = 0.5276$ olup $X \sim Bin(7, 0.5276)$ dir.

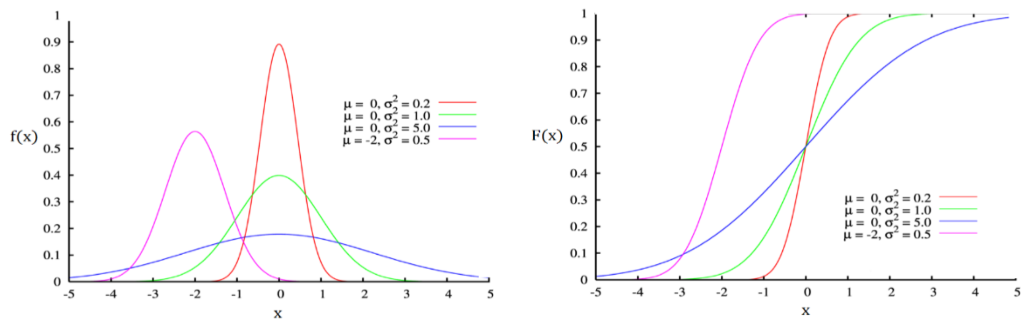
$$f(x) = \binom{7}{x} (0.5276)^x (1 - 0.5276)^{7-x}, x = 0,1,2,3,4,5,6,7$$

$$P(X = 5) = f(5) = \binom{7}{5} (0.5276)^5 (1 - 0.5276)^{7-5} = 0.1916$$

8.4.3 Normal dağılım: Sürekli bir rd nin dağılımı için en yaygın kullanılan model normal dağılımdır. Merkezi limit teoremine göre tekrarlanan her rasgele deneyde, tekrar sayısı arttıkça tekrarların ortalaması olan rd normal dağılıma yakınsar. X rd normal dağılıma sahipse oyf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad , -\infty < x, \mu < \infty \ , \qquad \sigma > 0$$

şeklindedir. $E(X) = \mu$ ve $V(X) = \sigma^2$ olup $X \sim N(\mu, \sigma^2)$ dir. Farklı ortalama ve varyansa göre normal dağılımın oyf ve bdf grafiği aşağıdaki şekildedir.



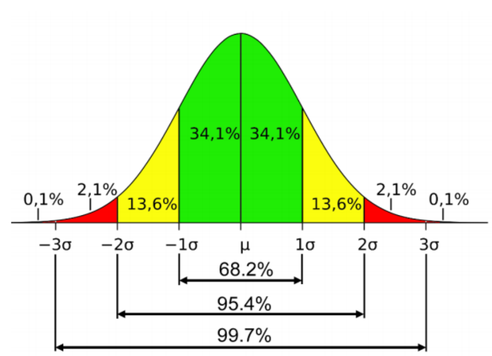
- **Normal dağılımın bazı temel özellikleri**

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

  ii. $P(a \le X \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

  iii. μ ye göre simetrik olup $\mu = M = Mod$ dur. $\int_{-\infty}^{\mu} f(x)dx = \int_{\mu}^{\infty} f(x)dx = 0.5$

  iv. $P(\mu - \sigma \le X \le \mu + \sigma) = 0.6826$
  $P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9544$
  $P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9974$



- **Standart Normal Dağılım**

Ortalaması sıfır ($\mu = 0$) ve varyansı bir ($\sigma^2 = 1$) olan normal dağılıma standart normal dağılım denir. Z standart normal dağılıma sahip rd olsun. $Z \sim N(0,1)$ ile gösterilir ve

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \ , \qquad -\infty < z < \infty$$

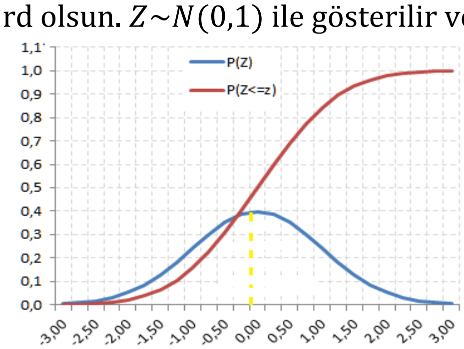
oyf sahiptir. Z rd nin bdf

$P(Z \le z) = P(Z < z) = F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

$P(Z > z) = 1 - P(Z \le z) = 1 - F(z)$

$P(a \le Z \le b) = P(Z \le b) - P(Z \le a) = F(b) - F(a)$

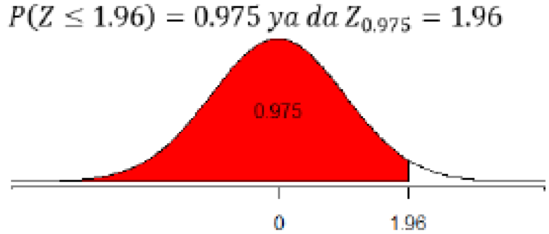
$P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z)$



Birikimli (Kümülatif) Standart Normal Dağılım Tablosu

Z_t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

$P(Z \leq 1.96) = 0.975$ *ya da* $Z_{0.975} = 1.96$



Bazı birikimli olasılıklar için tablo değerleri

Olasılık	0.90	0.95	0.975	0.99	0.995	0.999
Z_t	1.28	1.64	1.96	2.33	2.58	3.09

Örnek: $Z \sim N(0,1)$ olmak üzere Z rd nin 1.96 dan küçük değerler alması olasılığı nedir?

$1.96 = 1.90 + 0.06$ olarak düşünelim. Birinci sutundan 1.90, birinci satırdan da 0.06 değerlerinin kesiştikleri kutucuktur.

$$P(Z \leq 1.96) = 0.9750$$

Örnek: $Z \sim N(0,1)$ olmak üzere,

a) $P(Z \leq 1.5) = 0.9332$

b) $P(Z \geq 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$

c) $P(0.55 < Z \leq 1.77) = P(Z \leq 1.77) - P(Z < 0.55) = 0.9616 - 0.7088 = 0.2528$

d) $P(Z < -2.5) = P(Z > 2.5) = 1 - P(Z \leq 2.5) = 1 - 0.9938 = 0.0062$

e) $P(-2.62 < Z < 0.5) = P(Z < 0.5) - P(Z < -2.62)$
$= P(Z < 0.5) - \big(1 - P(Z < 2.62)\big) = 0.6915 - (1 - 0.9956) = 0.6871$

f) $P(Z < a) = 0.9975 \Rightarrow a = 2.81$

g) $P(Z < a) = 0.95 \Rightarrow a = 1.64$

Teorem: $X \sim N(\mu, \sigma^2)$ olmak üzere $Z = \frac{X-\mu}{\sigma}$ olarak tanımlanırsa $Z \sim N(0,1)$ dır.

Örnek: $X \sim N(\mu = 100, \sigma^2 = 25)$ olmak üzere $P(90 \leq X < 105) = ?$

$$P(90 \leq X < 105) = \int_{90}^{105} \frac{1}{5\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-100}{5}\right)^2} dx = ?$$

$$\begin{aligned} P(90 \leq X < 105) &= P\left(\frac{90-100}{5} \leq \frac{X-100}{5} < \frac{105-100}{5}\right) = P(-2 \leq Z < 1) \\ &= P(Z \leq 1) - P(Z < -2) = P(Z \leq 1) - \big(1 - P(Z < 2)\big) \\ &= 0.8413 - (1 - 0.9772) = 0.8185 \end{aligned}$$

Örnek: İstatistik dersi vize not ortalaması 68 ve standart sapması 2 olan normal dağılıma sahip olsun.

a) Dersi alan her hangi bir öğrencinin notunun 72 den fazla olması olasılığı nedir?

$$P(X > 72) = P\left(Z > \frac{72-68}{2}\right) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

b) Dersi alan öğrencilerin yüzde kaçının notu 66 ile 70 arasındadır?

$$\begin{aligned} P(66 < X < 70) &= P\left(\frac{66-68}{2} < Z < \frac{70-68}{2}\right) = P(-1 < Z < 1) \\ &= (Z < 1) - P(Z < -1) = P(Z < 1) - \big(1 - P(Z < 1)\big) \\ &= 0.8413 - (1 - 0.8413) = 0.6825 = \%68.25 \end{aligned}$$

c) 65 in altında not alan 14 öğrenci olduğuna göre yaklaşık kaç öğrenci dersi almıştır.

$$P(X < 65) = P\left(Z < \frac{65-68}{2}\right) = P(Z < -1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$$

O halde sınıfın %6.68 i 65 in altında not almıştır. Sınıfta toplam N tane öğrenci varsa $N * 0.0668 = 14 \Rightarrow N \cong 210$ kişi vardır.

Teorem: $X \sim N(\mu, \sigma^2)$ olsun. $Y = aX + b$ şeklinde tanımlı rd

$$E(Y) = aE(X) + b = a\mu + b \text{ ve } V(Y) = a^2V(X) = a^2\sigma^2 \text{ olup } Y \sim N(a\mu + b, a^2\sigma^2) \text{ dır.}$$

Örnek: $X \sim N(3,2)$ olsun. $Y = 5X + 4$ rd $Y \sim N(5 * 3 + 4, 5^2 * 2) = N(19, 50)$

Teorem: $X_1, X_2, \dots, X_n \sim N(\mu_i, \sigma_i^2)$ $(i = 1,2, \dots, n)$ bağımsız rd olmak üzere

$$X_1 + X_2 + \cdots + X_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$$

Örnek: $X_1 \sim N(2,1)$ ve $X_2 \sim N(3,2)$ bağımsız rd olmak üzere;

$$S = X_1 + X_2 \Rightarrow S \sim N(2 + 3, 1 + 2) = N(5,3)$$

$$D = X_1 - X_2 \Rightarrow D = X_1 + (-1)X_2$$
$$\Rightarrow E(D) = E(X_1) + E\big((-1)X_2\big) = 2 + (-1) * 3 = -1$$
$$\Rightarrow V(D) = V(X_1) + V\big((-1)X_2\big) = 1 + (-1)^2 * 2 = 3$$
$$\Rightarrow D \sim N(2 - 3, 1 + 2) = N(-1,3)$$

Örnek: Bir şehre ait yıllık yağış miktarının ortalaması 300mm ve standart sapması 10mm olan normal dağılıma sahip olsun. Yıllara göre yağış miktarları bağımsız olduğuna göre

a) Gelecek 2 yıl süresince toplam yağışın 625 mm aşması olasılığı nedir?
b) Takip eden iki yılın yağış miktarları farkının 15 mm den az olması olasılığı nedir.

Çözüm: $\mu_i = 300$ ve $\sigma_i^2 = 100$ olup $X_i \sim N(300,100)$ dır.

a) $S = X_1 + X_2$ ise $S \sim N(300 + 300, 100 + 100) = N(600,200)$

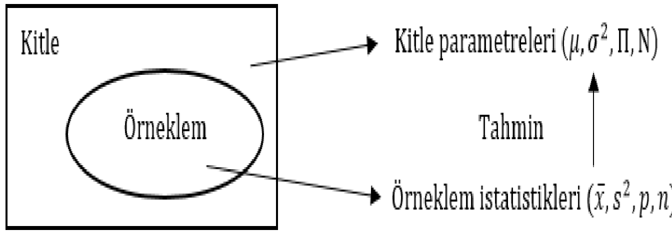
$$P(S > 625) = P\left(Z > \frac{625 - 600}{\sqrt{200}}\right) = P(Z > 1.77) = 1 - P(Z < 1.77) = 1 - 0.9616$$
$$= 0.0384 \cong \%4$$

b) $D = X_1 - X_2$ ise $D \sim N(300 - 300, 100 + 100) = N(0,200)$

$$P(D < 15) = P\left(Z < \frac{15 - 0}{\sqrt{200}}\right) = P(Z < 1.06) = 0.8554 \cong \%86$$

9. ÖRNEKLEME DAĞILIMI

Örneklem istatistiklerini kullanarak bilinmeyen kitle parametrelerini tahmin etme yöntemlerinin genel adına *çıkarımsal istatistik* denir.



Kitleden seçilen örneklemin istatistikleri ($\bar{x}$, s^2,p gibi) elde edilen her bir örneklem için değişkenlik göstereceğinden bu istatistikler birer rasgele değişkenlerdir ($\bar{X}$, S^2,P gibi) ve birer dağılıma sahiptirler.

Tanım: N birimlik bir kitleden n büyüklüğünde çekilebilecek bütün örneklemler için hesaplanan herhangi bir istatistiğin dağılımına *örnekleme dağılımı* denir.

9.1 Örneklem Ortalamasının Örnekleme Dağılımı

Tanım: μ ortalamalı σ^2 varyanslı N birimlik bir kitleden çekilebilecek (iadeli N^n ya da iadesiz $C(N,n)$ tane) bütün örneklemlerin ortalamasından oluşan $\bar{X}$ rasgele değişkeninin dağılımına *ortalamanın örnekleme dağılımı* denir.

$E(\bar{X}) = \mu_{\bar{X}} = \mu$ olup iadeli örneklem için $V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ iadesiz örneklem için $V(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} * \frac{N-n}{N-1}$ dir.

Not: Genel olarak biz kitledeki gözlem sayısını bilmeyiz ve sonsuz büyüklükte kabul ederiz. Bu nedenle $\left({}^{N-n}/_{N-1}\right)$ değeri büyük N ler (ya da $n / N \leq 0.05$) için 1'e yakınsayacağından iadeli ve iadesiz örneklemler için varyans $\sigma_{\bar{X}}^2 = \sigma^2/n$ olarak alınır ve standart sapması (hatası) da $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ dir.

Örnek: N=3 birimlik bir kitlenin elemanları $X_1 = 2, X_2 = 4\ ve\ X_3 = 6$ olsun. Bu kitleden iadeli ve iadesiz olarak n=2 birimlik tüm mümkün örneklemler çekilsin. Örnekleme dağılımını oluşturup, dağılımın beklenen değeri (ortalaması) ve varyansını hesaplayınız.

Çözüm: Öncelikle kitle ortalaması ve varyansını bulalım.

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{2+4+6}{3} = 4$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N} = \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} = 2.67$$

a) *İadeli örneklemler*: 3 birimlik bir kitleden iadeli olarak alınan 2 birimlik örneklemlerin sayısı $N^n = 3^2 = 9$ dur. Bunlar

No	Tüm mümkün örneklemler	Örneklem ortalamaları ($\bar{X}_i$)
1	2 ve 2	2
2	2 ve 4	3
3	2 ve 6	4
4	4 ve 2	3
5	4 ve 4	4
6	4 ve 6	5
7	6 ve 2	4
8	6 ve 4	5
9	6 ve 6	6

Ortalamanın Örnekleme Dağılımı		
Örneklem ortalamaları ($\bar{X}$)	**Frekanslar (f)**	**Göreli frekanslar (olasılıklar $f(\bar{X})$)**
2	1	1/9
3	2	2/9
4	3	3/9
5	2	2/9
6	1	1/9
Toplam	**9**	**1**

Dağılımın beklenen değeri

$$E(\bar{X}) = \sum \bar{X} * f(\bar{X}) = 2 * \frac{1}{9} + 3 * \frac{2}{9} + 4 * \frac{3}{9} + 5 * \frac{2}{9} + 6 * \frac{1}{9} = \frac{36}{9} = 4$$
$$= \frac{\sum \bar{X} * f}{\sum f} = \frac{2 * 1 + 3 * 2 + 4 * 3 + 5 * 2 + 6 * 1}{9} = 4$$

olup $E(\bar{X}) = \mu = 4$

Dağılımın varyansı

$$V(\bar{X}) = \sum \left(\bar{X} - E(\bar{X})\right)^2 * f(\bar{X}) = (2-4)^2 * \frac{1}{9} + (3-4)^2 * \frac{2}{9} + \cdots + (6-4)^2 * \frac{1}{9} = 1.33$$
$$= \frac{\sum \left(\bar{X} - E(\bar{X})\right)^2 * f}{\sum f} = \frac{(2-4)^2 * 1 + (3-4)^2 * 2 + \cdots + (6-4)^2 * 1}{9} = 1.33$$

olup $V(\bar{X}) = \frac{\sigma^2}{n} = \frac{2.67}{2} = 1.33$

b) *İadesiz örneklemler*: 3 birimlik bir kitleden iadesiz olarak alınan 2 birimlik örneklemlerin sayısı $C(3,2) = 3$ dur. Bunlar

No	Tüm mümkün örneklemler	Örneklem ortalamaları ($\bar{X}_i$)
1	2 ve 4	3
2	2 ve 6	4
3	4 ve 6	5

Ortalamanın Örnekleme Dağılımı		
Örneklem ortalamaları ($\bar{X}$)	**Frekanslar (f)**	**Göreli frekanslar (olasılıklar $f(\bar{X})$)**
3	1	1/3
4	1	1/3
5	1	1/3
Toplam	**3**	**1**

Dağılımın beklenen değeri

$$E(\bar{X}) = \sum \bar{X} * f(\bar{X}) = 3 * \frac{1}{3} + 4 * \frac{1}{3} + 5 * \frac{1}{3} = \frac{12}{3} = 4$$
$$= \frac{\sum \bar{X} * f}{\sum f} = \frac{3 * 1 + 4 * 1 + 5 * 1 +}{3} = 4$$

olup $E(\bar{X}) = \mu = 4$

Dağılımın varyansı

$$V(\bar{X}) = \sum \left(\bar{X} - E(\bar{X})\right)^2 * f(\bar{X}) = (3-4)^2 * \frac{1}{3} + (4-4)^2 * \frac{1}{3} + (5-4)^2 * \frac{1}{3} = 0.67$$
$$= \frac{\sum \left(\bar{X} - E(\bar{X})\right)^2 * f}{\sum f} = \frac{(3-4)^2 * 1 + (4-4)^2 * 1 + (5-4)^2 * 1}{3} = 0.67$$

olup $V(\bar{X}) = \frac{\sigma^2}{n} * \frac{N-n}{N-1} = \frac{2.67}{2} * \frac{3-2}{3-1} = 0.67$

Teorem: μ ortalamalı σ^2 varyanslı ***normal dağılıma*** sahip bir kitleden çekilen örneklemin ortalamasının örnekleme dağılımı $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ olup $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ olur.

Teorem (Merkezi Limit): μ ortalamalı σ^2 varyanslı bir kitleden $n \geq 30$ olacak şekildeki örneklemin ortalamasının örnekleme dağılımı $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ sahiptir. O halde $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ olur.

Örnek: Elektronik parça üretimi yapan bir firma ortalaması 100 Ω ve standart sapması 10 Ω olan rezistanslar imal etmektedir. Rezistansın dağılımı normaldir. Örneklem hacmi 25 olan rezistansların oluşturduğu örneklemin ortalamasının 95 Ω değerinin altında olması ihtimali nedir?

Çözüm: Normal dağılıma sahip ortalaması ve varyansı bilinen kitleden alınan örneklem ortalamasının dağılımı da normal olup $n = 25$, $\mu_{\bar{X}} = 100$ ve $\sigma_{\bar{X}} = 10/\sqrt{25}$ dir.

$$P(\bar{X} < 95) = P\left(Z < \frac{95 - 100}{10/\sqrt{25}}\right) = P(Z < -2.5) = 1 - P(Z < 2.5) = 1 - 0.9938 = 0.0062$$

Örnek: Demir testere bilendikten sonraki kullanım ömrü ortalaması 60 saat ve varyansı 36 olan dağılıma sahip olsun. İmal edilen bu testerelerden rasgele seçilen 64 birimlik örneklemin ortalamasının 58 ile 61 saat arasında olması olasılığı nedir?

Çözüm: Kitle normal olmasa da çekilen örneklem büyüklüğü $n = 64 \geq 30$ olup $\bar{X} \sim N(\mu = 60, \frac{\sigma^2}{n} = \frac{36}{64})$ olur. O halde bizden istenilen olasılık

$$P(58 < \bar{X} < 61) = P\left(\frac{58 - 60}{6/\sqrt{64}} < Z < \frac{61 - 60}{6/\sqrt{64}}\right) = P(-2.67 < Z < 1.33)$$
$$= P(Z < 1.33) - \big(1 - P(Z < 2.67)\big) = 0.9082 - (1 - 0.9962) = 0.9044$$

Örnek: X sürekli rasgele değişkeninin olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} 1/4 & 0 \leq x \leq 4 \\ 0 & d.d \end{cases}$$

şeklindedir. Bu kitleden rasgele seçilen 50 birimlik örneklemin ortalamasının 2.5 dan az olması olasılığı nedir?

Çözüm: Örneklemin gözlem sayısı $n = 50 \geq 30$ olup örneklem ortalamasının örnekleme dağılımı normal dağılıma sahiptir. Öncelikle kitlenin ortalaması ve varyansını bulalım. X r.d. düzgün dağılıma sahip olup $\mu = \frac{0+4}{2} = 2$ ve $\sigma^2 = \frac{(4-0)^2}{12} = \frac{4}{3} = 1.33$ dür. Ya da

$$\mu = E(X) = \int_0^4 x * \frac{1}{4} dx = \frac{1}{4} * \frac{x^2}{2}\Big|_0^4 = 2$$

$$\sigma^2 = V(X) = E(X^2) - E(X)^2 = \int_0^4 x^2 * \frac{1}{4} dx - 2^2 = \frac{1}{4} * \frac{x^3}{3}\Big|_0^4 = \frac{16}{3} - 4 = \frac{4}{3} = 1.33$$

O halde $\mu_{\bar{X}} = 2$ ve $\sigma_{\bar{X}}^2 = \frac{1.33}{50}$ olup $\bar{X} \sim N(2, \frac{1.33}{50})$ olur.

$$P(\bar{X} < 2.5) = P\left(Z < \frac{2.5 - 2}{\sqrt{1.33}/\sqrt{50}}\right) = P(Z < 3.07) = 0.9989$$

10. PARAMETRE TAHMİNİ

Parametre tahminini nokta tahmini ve aralık tahmini olarak iki kısımda incelenir.

10.1 Nokta Tahmini

10.1 Nokta Tahmini: θ kitlenin bilinmeyen parametresi olmak üzere, örneklemden elde edilen $\hat{\theta}$ değerine θ nın tahmincisi denir.

Varsayalım ki bir mühendis otomobil şasisi üzerinde kullanılan bir parçanın kopma mukavemetini analiz ediyor olsun. Mühendis parçaların oluşturduğu kitleye ait ortalama kopma mukavemetini tahmin etmek için örneklem verisinin ortalamasını kullanacaktır. Bu sayı nokta tahmini olarak bilinir. Örneğin μ nün nokta tahmini $\bar{x}$ olup $\hat{\mu} = \bar{x}$ şeklinde gösterilir.

Örnek: Bir bilgisayar şirketinin ürettiği ekran kartlarından rasgele seçilen 5 tanesinin ömürleri 12, 15, 18, 20 ve 24 ay olarak bulunuyor.

Bu bilgiye dayanarak şirketin ürettiği tüm ekran kartlarının ortalama ömrünün nokta tahmini $\hat{\mu} = \bar{x} = \frac{12+15+18+20+24}{5} = 17.80$ ay olarak hesaplanır.

Benzer şekilde kitlenin varyansı σ^2 nin nokta tahmini de örneklemin varyansı olup $\widehat{\sigma^2} = s^2 = 21.2$ dir.

10.2 Aralık Tahmini (Güven Aralığı): Kitle parametresi θ yı nokta tahminleme yerine belirli bir hata payı (e) içeren bir aralıkla da tahmin edebiliriz. O halde $\left[\hat{\theta} - e,\ \hat{\theta} + e\right]$ aralığına θ nın aralık tahmini denir.

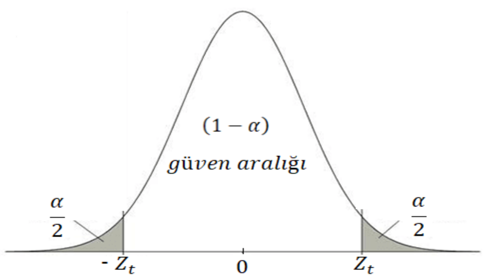
$$P\left[\hat{\theta} - e \leq \theta \leq \hat{\theta} + e\right] = 1 - \alpha$$

eşitliğinde $(1 - \alpha)$ değeri, θ nın $\left[\hat{\theta} - e,\ \hat{\theta} + e\right]$ aralığında bulunması olasılığını gösterir. Burada α anlam seviyesi genellikle 0.01, 0.05 ve 0.10 alınır ve sırasıyla kitle parametresi için %99, %95 ve %90 güven aralıkları belirlenmiş olur.

Örneğin $\alpha = 0.05$ alınırsa bunun anlamı kitle parametresinin %95 güven aralığı demek olur ve 100 tekrarda elde edilen tahminin en az 95'i belirlenen aralıkta olduğunun göstergesidir.

10.2.1 Kitle ortalaması (μ) için aralık tahmini: Kitle ortalamasının aralık tahmini için 4 farklı durum söz konusudur. Bu durumlar kitlenin varyansının bilinip bilinmediği ve kitlenin normal dağılıma sahip olup olmadığıdır.

❖ Kitlenin varyansı (σ^2) <u>biliniyorsa</u> ve kitle <u>normal</u> dağılıma sahipse örneklem büyüklüğü (n) ne olursa olsun ortalamanın örnekleme dağılımı da normal olup



$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$
$$P(-z_t \leq Z \leq z_t) = 1-\alpha$$
$$P\left(-z_t \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_t\right) = 1-\alpha$$
$$P\left(\bar{X}-z_t\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+z_t\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Burada $z_t = z_{1-\frac{\alpha}{2}}$ dir. Çünkü z_t değerinden eşit ve küçük olması olasılıkları toplam $(1-\alpha)+\frac{\alpha}{2} = 1-\frac{\alpha}{2}$ dir.

Örneğin; $\alpha = 0.05$ için $z_t = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$

O halde μ nün $1-\alpha$ güven arağı $\bar{x}-\left[z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \leq \mu \leq \bar{x}+\left[z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ dir.

Örnek: Varyansı 225 olan normal dağılmış bir kitleden 25 birim büyüklüğünde örneklem çekilmiş ve ortalaması 110 olarak bulunmuştur. Kitle ortalamasının %95 güven aralığını belirleyiniz.

Çözüm: $\sigma^2 = 225$ biliniyor ve kitle normal olup örneklemin gözlem sayısına bakılmaksızın kitle ortalamasının aralık tahmini $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ dur. $\sigma^2 = 225$ olup $\sigma = 15,\ \bar{x} = 110$ ve $n = 25$ dir. $1-\alpha = 0.95$ ise $\alpha = 0.05$ olup $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ dır.

O halde μ nün %95 güven aralığı

$$110-\left[1.96*\frac{15}{\sqrt{25}}\right] \leq \mu \leq 110+\left[1.96*\frac{15}{\sqrt{25}}\right] \Rightarrow 104.12 \leq \mu \leq 115.88$$

Yani kitle ortalaması %95 ihtimalle 104.12 ile 115.88 arasındadır.

❖ Kitlenin varyansı (σ^2) <u>biliniyor</u> ve kitlenin dağılımı <u>normal değilse</u> örneklem büyüklüğü $n \geq 30$ olduğunda merkezi limit teoreminden dolayı ortalamanın örnekleme dağılımı da normal olup $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ dir. Yine kitle ortalaması μ nün aralık tahmini $\bar{x}-\left[z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \leq \mu \leq \bar{x}+\left[z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ dir.

Örnek: Önceki bilgilere dayanarak ehliyet sınavına giren adayların aldıkları puanların standart sapması 25 puan olduğu bilinmektedir. Bu sınava giren adaylardan rasgele seçilen 80 adayın puan ortalaması 75 olduğuna göre kitlenin ortalama puanının %90 güven aralığını belirleyiniz.

Çözüm: Standart sapma $\sigma = 25$ olup varyans biliniyor demektir. Lakin kitlenin dağılımı bilinmediğinden güven aralığı için sadece örneklemin gözlem sayısı $n \geq 30$ olması durumunda Z tablo değerleri kullanılabilir. Burada $n = 80 \geq 30$ olup güven aralığı $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ ile belirlenebilir. $\bar{x} = 75$, $1 - \alpha = 0.90$ olup $\alpha = 0.10$ dir.

Buradan $z_{1-\frac{0.10}{2}} = z_{0.95} = 1.64$ dır. O halde μ nün %90 güven aralığı

$$75 - \left[1.64 * \frac{25}{\sqrt{80}}\right] \leq \mu \ \leq 75 + \left[1.64 * \frac{25}{\sqrt{80}}\right] \Rightarrow 70.42 \leq \mu \ \leq 79.58$$

Yani kitle puan ortalaması %90 ihtimalle 70.42 ile 79.58 arasındadır.

❖ Kitlenin varyansı (σ^2) bilinmiyorsa ve kitle normal dağılıma sahipse kitlenin varyansı (σ^2) yerine onun nokta tahmini olan örneklem varyansı (s^2) kullanılır. μ nün $1 - \alpha$ güven aralığı örneklemdeki gözlem sayısına göre sayısına (n) göre;

$n \geq 30$ ise $Z = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(0,1)$ olup $\bar{x} - \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right] \leq \mu \ \leq \bar{x} + \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$

$n < 30$ ise $T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ olup $\bar{x} - \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right] \leq \mu \ \leq \bar{x} + \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right]$

ile belirlenir. Burada $t_{1-\frac{\alpha}{2};\,(n-1)}$ değeri $n - 1$ serbestlik dereceli t dağılımının $1 - \frac{\alpha}{2}$ olasılığına sahip tablo değeridir.

Örnek: Normal dağılan bir kitleden alınan 36 birimlik örneklemin ortalaması 25 ve varyansı 16 olduğuna göre kitle ortalamasının %95 güven aralığını belirleyiniz.

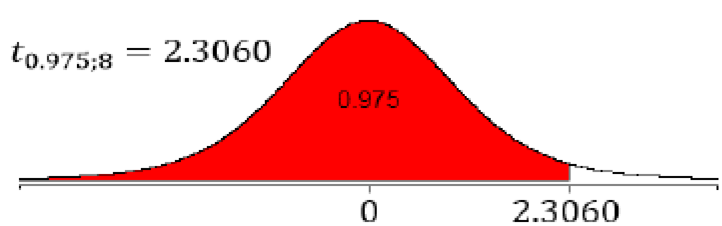
Çözüm: σ^2 bilinmiyor lakin ve kitle normal olup $n = 36 \geq 30$ olduğundan σ yerine onun tahmini s yi alarak kitle ortalamasının aralık tahmini $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$ ile hesaplanır. Burada $\bar{x} = 25$, $s^2 = 16$ ve $\alpha = 0.05$ olup $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ dir.

$$25 - \left[1.96 * \frac{4}{\sqrt{36}}\right] \leq \mu \ \leq 25 + \left[1.96 * \frac{4}{\sqrt{36}}\right] \Rightarrow 23.69 \leq \mu \ \leq 26.31$$

Yani kitle ortalaması %95 ihtimalle 23.69 ile 26.31 arasındadır.

Birikimli (Kümülatif) Student's t - Dağılım Tablosu

s.d.	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.999}$	$t_{0.9995}$
1	0.7265	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
35	0.5292	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400	3.5911
40	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
45	0.5281	0.8497	1.3006	1.6794	2.0141	2.4121	2.6896	3.2815	3.5203
50	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
60	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.5268	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.5263	0.8456	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
200	0.5258	0.8446	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
∞	0.5256	0.8442	1.2876	1.6558	1.9771	2.3533	2.6114	3.1495	3.3614



Örnek: Normal dağıldığı bilinen bir kitleden alınan 16 birimlik bir örneklemin ortalaması 35 ve standart sapması 4.6 olarak bulunmuştur. Kitle ortalamasının $\alpha = 0.05$ anlam seviyesinde güven aralığını belirleyiniz.

Çözüm: σ^2 bilinmiyor lakin ve kitle normal olup $n = 16 < 30$ olduğundan σ yerine onun tahmini s yi alarak kitle ortalamasının aralık tahmini $\bar{x} \mp \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right]$ ile hesaplanır. Burada $\bar{x} = 35, s = 4.6$ ve $\alpha = 0.05$ dir. Bizden %95 güven aralığı istenmektedir.

$t_{1-\frac{\alpha}{2};\,(n-1)} = t_{1-\frac{0.05}{2};\,(16-1)} = t_{0.975;\,15} = 2.1314$ olup

$$35 - \left[2.1314 * \frac{4.6}{\sqrt{16}}\right] \leq \mu \ \leq 35 + \left[2.1314 * \frac{4.6}{\sqrt{16}}\right] \Rightarrow 32.55 \leq \mu \ \leq 37.45$$

Yani kitle ortalaması %95 ihtimalle 32.55 ile 37.45 arasındadır.

❖ Kitlenin varyansı (σ^2) bilinmiyorsa ve kitle normal değilse uygulama açısından sadece $n \geq 30$ için $\bar{x} - \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right] \leq \mu \ \leq \bar{x} + \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$ dir.

Örnek: Bir otelin lobisine gelen müşterilerin ortalama kaç dakika burada vakit geçirdikleri araştırmak istenmektedir. Rasgele seçilen 36 kişinin lobide kaç dakika kaldıkları gözlemlenmiş ve kayıt altına alınmıştır. Bu verilerin ortalaması 25 ve standart sapması 4 olarak hesaplanmıştır. Lobiye gelen tüm müşterilerin ortalama geçirdikleri sürenin %95 güven aralığını bulunuz.

Çözüm: σ^2 ve kitlenin dağılımı bilinmiyor lakin $n = 36 \geq 30$ olduğundan kitle ortalamasının aralık tahmini $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$ ile hesaplanır. Burada $\bar{x} = 25, s = 4$ ve $1 - \alpha = 0.95$ olup $\alpha = 0.05$ dir. Buradan $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ olup

$$25 - \left[1.96 * \frac{4}{\sqrt{36}}\right] \leq \mu \ \leq 25 + \left[1.96 * \frac{4}{\sqrt{36}}\right] \Rightarrow 23.69 \leq \mu \ \leq 26.31$$

Yani %95 ihtimalle müşteriler lobide ortalama 23.69 ile 26.31 dakika arasında vakit geçirmektedir.

Not: μ nün $1 - \alpha$ güven aralığını şöyle özetleyebiliriz:

- ❖ σ^2 bilinivor ve kitle normal ise n ne olursa olsun $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$
- ❖ σ^2 biliniyor ve kitle normal değilse ancak $n \geq 30$ ise $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$
- ❖ σ^2 bilinmiyor ve kitle normal olsun olmasın $n \geq 30$ ise $\bar{x} \mp \left[z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$
- ❖ σ^2 bilinmiyor ve kitle normal olsun $n < 30$ ise $\bar{x} \mp \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right]$

Örnek: Normal dağıldığı bilinen bir kitleden alınan örneklemin {1,2,3,4,5,6,7,8,9} ölçüm değerlerini kullanarak kitle ortalamasının %95 güven aralığını bulunuz.

Çözüm: σ^2 bilinmiyor lakin ve kitle normal olup $n = 9 < 30$ olduğu için güven aralığı $\bar{x} \mp \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right]$ ile hesaplanır. Burada örneklemin ortalaması ve standart sapmasının hesaplanması gerekmektedir.

$$\bar{x} = \frac{1+2+\cdots+9}{9} = 5 \;\Rightarrow s^2 = \frac{(1-5)^2+(2-5)^2+\cdots+(9-5)^2}{9-1} = 7.5 \;\Rightarrow s = 2.74$$

$$t_{1-\frac{\alpha}{2};\,(n-1)} = t_{1-\frac{0.05}{2};\,(9-1)} = t_{0.975;\,8} = 2.3060$$

$$5 - \left[2.3060 * \frac{2.74}{\sqrt{9}}\right] \le \mu \;\le 5 + \left[2.3060 * \frac{2.74}{\sqrt{9}}\right] \Rightarrow 2.89 \le \mu \;\le 7.11$$

Kitle ortalaması %95 ihtimalle 2.89 ile 7.11 arasındadır.

Örnek: Bir araba üreticisi rasgele seçtiği 6 arabanın CO_2 emisyon değerlerini {100.1, 99.9, 100.7, 99.3, 99,5, 100.5} g/km olarak ölçmüştür. Verilerin normal dağıldığı varsayımı altında üretilen arabaların ortalama CO_2 emisyon değerlerinin %90 güven aralığını bulunuz.

Çözüm: σ^2 bilinmiyor lakin ve kitle normal olup $n = 6 < 30$ olduğu için güven aralığı $\bar{x} \mp \left[t_{1-\frac{\alpha}{2};\,(n-1)} \frac{s}{\sqrt{n}}\right]$ ile hesaplanır.

$$\bar{x} = \frac{100.1+\cdots+100.5}{6} = 100$$

$$s^2 = \frac{(100.1-100)^2+\cdots+(100.5-100)^2}{6-1} = 0.3 \Rightarrow s = 0.55$$

$$t_{1-\frac{\alpha}{2};\,(n-1)} = t_{1-\frac{0.10}{2};\,(6-1)} = t_{0.95;\,5} = 2.0150$$

$$100 - \left[2.0150 * \frac{0.55}{\sqrt{6}}\right] \le \mu \;\le 100 + \left[2.0150 * \frac{0.55}{\sqrt{6}}\right] \Rightarrow 99.54 \le \mu \;\le 100.46$$

Yani üretilen arabaların ortalama CO_2 emisyon değerlerinin %90 güvenle 99.54 ile 100.46 arasındadır.

Örnek: Bir hastalığın tedavisinde yeni bir yöntem geliştirilmiştir. Rasgele seçilen 12 hasta sözü edilen bu yöntem ile tedavi edilerek iyileşinceye kadar geçen süreler {10, 12, 8, 9, 9, 17, 9, 14, 3, 7, 12, 10} gün şeklindedir. Verilerin normal dağıldığı varsayımı altında $\alpha = 0.05$ anlam seviyesinde hastaların ortalama iyileşme süresi için güven aralığını bulunuz.

Çözüm: Kitle normal olup $n = 12 < 30$ olduğu t tablo değeri kullanılır.

$$\bar{x} = \frac{10+12+\cdots+10}{12} = 10 \Rightarrow s^2 = 12.55 \;\Rightarrow s = 3.54$$

$$t_{1-\frac{\alpha}{2};\,(n-1)} = t_{1-\frac{0.05}{2};\,(12-1)} = t_{0.975;\,11} = 2.2010$$

$$10-\left[2.2010*\frac{3.54}{\sqrt{12}}\right]\leq\mu\ \leq 10+\left[2.2010*\frac{3.54}{\sqrt{12}}\right]\Rightarrow 7.75\leq\mu\ \leq 12.24$$

O halde hastaların bu yöntem ile ortalama iyileşme süresi %95 güvenle 7.75 ile 12.24 gün arasındadır.

Örnek: Belirli bir mahalleden rasgele seçilen 10 binanın yaşlarına ilişkin verilen $\{3,3,3,4,4,5,6,7,7,8\}$ yıl olarak belirlenmiştir. Mahalledeki binaların yaş değerlerinin normal dağıldığı varsayımı altında mahalledeki binaların ortalama yaşlarına ilişkin $\alpha=0.01$ anlam seviyesinde güven aralığını bulunuz.

Çözüm: Kitle normal olup $n=10<30$ olduğu t tablo değeri kullanılır.

$$\bar{x}=\frac{3+3+\cdots+8}{10}=5\Rightarrow s^2=3.56\ \Rightarrow s=1.89$$

$$t_{1-\frac{\alpha}{2};\,(n-1)}=t_{1-\frac{0.01}{2};\,(10-1)}=t_{0.995;\,9}=3.2498$$

$$5-\left[3.2498*\frac{1.89}{\sqrt{10}}\right]\leq\mu\ \leq 5+\left[3.2498*\frac{1.89}{\sqrt{10}}\right]\Rightarrow 3.06\leq\mu\ \leq 6.94$$

Örnek: Bir çay fabrikası ürettiği 1 kilogramlık çay paketlerinin ortalama ağırlığı için %95 güven aralığını belirlemek istemektedir. Üretilen çay paketlerinden rasgele seçilen 49 tanesinin ortalama ağırlığı 995 gr ve standart sapması 70 gr olduğuna göre güven aralığını oluşturunuz.

Çözüm: σ^2 ve kitlenin dağılımı bilinmiyor lakin $n=49\geq 30$ olduğundan kitle ortalamasının aralık tahmini $\bar{x}\mp\left[z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$ ile hesaplanır.

Burada $\bar{x}=995$, $s=70$ ve $\alpha=0.05$ olup $z_{1-\frac{\alpha}{2}}=z_{0.975}=1.96$ olur.

$$995-\left[1.96*\frac{70}{\sqrt{49}}\right]\leq\mu\ \leq 995+\left[1.96*\frac{70}{\sqrt{49}}\right]\Rightarrow 975.40\leq\mu\ \leq 1014.60$$

10.2.2 Kitle oranı (Π) için aralık tahmini: Bir fabrikada üretilen ürünlerin kusurlu ya da kusursuz olması, bir fakültedeki öğrencilerin başarılı ya da başarısız olması gibi iki sonuçlu kitlelerin oran parametresi $\Pi = R/N$ ile hesaplanır ($0 \leq \Pi \leq 1$). Burada N kitledeki eleman sayısı olup R de ilgilenilen özelliğe sahip gözlem sayısıdır.

Kitleden alınan n boyutlu örneklemde ilgilenilen özelliğe sahip r tane gözlem varsa örneklemin oranı $p = r/n$ olacaktır. Eğer Π, 0 veya 1 e çok yakın değil ve n de yeterince büyükse ($np > 5$ veya $n(1-p) > 5$) örnek oranının dağılımı ortalaması $\mu = \Pi$ ve varyansı $\sigma_p^2 = \Pi(1-\Pi)/n$ olan normal dağılıma sahiptir. Kitle oranı Π bilinmediği durumlarda tahmini olan p alınarak örnek varyansı $s_p^2 = \frac{p(1-p)}{n} \Rightarrow s_p = \sqrt{\frac{p(1-p)}{n}}$ olur. O halde kitle oranı Π için nün $1-\alpha$ güven aralığı

$$p - \left[z_{1-\frac{\alpha}{2}}\, s_p\right] \leq \Pi \leq p + \left[z_{1-\frac{\alpha}{2}}\, s_p\right]$$

ile hesaplanır.

Örnek: Bir fabrikada üretilen ürünlerin içerisinden rasgele seçilen 80 ürünün 6 tanesinin bozuk olduğu gözlemlenmiştir. Fabrikada üretilen ürünlerin bozuk oranı için %95 güven aralığını bulunuz.

Çözüm: $n = 80$, $r = 6$ olup $p = \frac{r}{n} = \frac{6}{80} = 0.075$ ve $s_p = \sqrt{\frac{0.075(1-0.075)}{80}} = 0.0294$ dur. Burada $np = 80 * 0.075 = 6 > 5$ olup Π nin güven aralığı $p \mp \left[z_{1-\frac{\alpha}{2}}\, s_p\right]$ ile hesaplanabilir. $1-\alpha = 0.95$ ise $\alpha = 0.05$ dir. Buradan $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ olup

$$0.075 - [1.96 * 0.0294] \leq \Pi \leq 0.075 + [1.96 * 0.0294] \Rightarrow 0.0174 \leq \Pi \leq 0.1326$$

Yani üretilen ürünlerin bozuk olma oranı %95 güvenle (ihtimalle) %1.74 ile %13.26 arasındadır.

Örnek: Seçim öncesi yapılan bir ankette 1000 kişiden 200 ünün bir siyasi partiyi desteklediklerі tespit edilmiştir. Bu siyasi partinin oy oranı için %99 güven aralığını bulunuz.

Çözüm: $n = 1000$, $r = 200$ olup $p = \frac{r}{n} = \frac{200}{1000} = 0.20$ ve $s_p = \sqrt{\frac{0.20(1-0.20)}{1000}} = 0.0126$ dur. Burada $np = 200 > 5$ olup Π nin güven aralığı $p \mp \left[z_{1-\frac{\alpha}{2}}\, s_p\right]$ ile hesaplanabilir.

$1-\alpha = 0.99$ olup $\alpha = 0.01$ dir. Buradan $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.01}{2}} = Z_{0.995} = 2.58$

$$0.20 - [2.58 * 0.0126] \leq \Pi \leq 0.20 + [2.58 * 0.0126] \Rightarrow 0.1675 \leq \Pi \leq 0.2325$$

Yani bu siyasi partinin oy oranı %99 ihtimalle %17 ile %23 arasındadır.

Örnek: Bir bölgede sigara içen insanlar arasında kanser hastası olanların oranının belirlenmesi istenmektedir. Bu kişiler arasından rasgele seçilen 1500 kişinin 375 i kanser hastası olduğu tespit edilmiştir. Bu bölgedeki sigara içen insanların kansere yakalanması oranı için %95 güven aralığını bulunuz.

Çözüm: $n = 1500$, $r = 375$ olup $p = \frac{375}{1500} = 0.25$ ve $s_p = \sqrt{\frac{0.25(1-0.25)}{1500}} = 0.0112$ dur.

$\alpha = 0.05$ olup $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ dır. Π nin güven aralığı

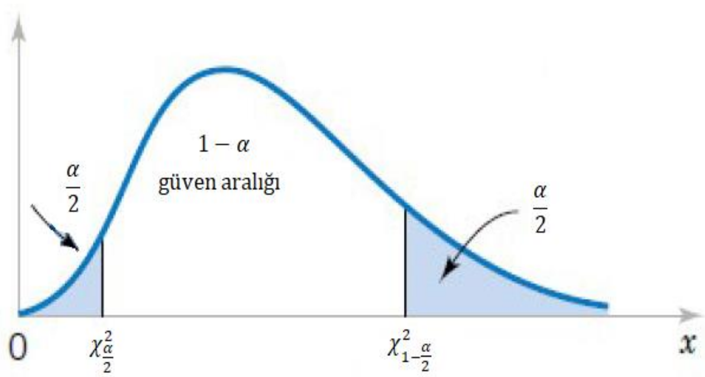
$$0.25 - [1.96 * 0.0112] \leq \Pi \leq 0.25 + [1.96 * 0.0112] \Rightarrow 0.2280 \leq \Pi \leq 0.2720$$

O halde bu bölgedeki sigara içen insanların kansere yakalanması oranı %95 ihtimalle %23 ile %27 arasındadır.

10.2.3 Kitle varyansı (σ^2) için aralık tahmini: μ ortalamalı σ^2 varyanslı normal dağılıma sahip bir kitleden çekilen n büyüklüğündeki örneklem varyansı s^2 olmak üzere

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

değeri $(n-1)$ serbestlik dereceli χ^2 (Ki-Kare) dağılımına sahiptir. Burada kitle varyansı σ^2 için $1-\alpha$ güven aralığı



$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};\,(n-1)}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};\,(n-1)}}$$

ile hesaplanır.

Örnek: Normal dağılan bir kitleden çekilen 12 birimlik örneklemin ortalaması 20 ve varyansı 10 olarak hesaplanmıştır. Kitle varyansı için %90 güven aralığını belirleyiniz.

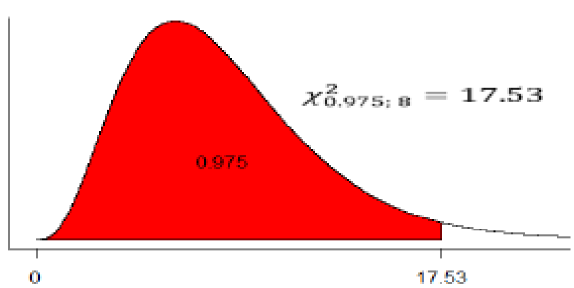
Çözüm: $s^2 = 10$, $n = 12$, $\alpha = 0.10$ olup $\chi^2_{\frac{\alpha}{2};\,(n-1)} = \chi^2_{\frac{0.10}{2};\,(12-1)} = \chi^2_{0.05;\,11} = 4.57$ ve $\chi^2_{1-\frac{\alpha}{2};\,(n-1)} = \chi^2_{1-\frac{0.10}{2};\,(12-1)} = \chi^2_{0.95;\,11} = 19.68$ olup σ^2 için %90 güven aralığı

$$\frac{(12-1)10}{19.68} \leq \sigma^2 \leq \frac{(12-1)10}{4.57} \Rightarrow 5.59 \leq \sigma^2 \leq 24.07$$

O halde %90 ihtimalle kitle varyansı σ^2 değeri 5.59 ile 24.07 arasındadır.

Birikimli (Kümülatif) Ki Kare Dağılım Tablosu

s.d	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.25}$	$\chi^2_{0.50}$	$\chi^2_{0.75}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17



10.3 Örneklem büyüklüğünün ($\boldsymbol{n}$) belirlenmesi: Örneklem büyüklüğü arttıkça kitle parametrelerine ilişkin bilgilerimizin de artacağı kesindir. Fakat bu durumun maliyet ve zaman açısından dezavantaja sahiptir. O halde optimum örneklem büyüklüğünün belirlenmesi önemlidir. Kitle parametresi için belirlenen maksimum hata değeri e olmak üzere $\left[\hat{\theta} - e \ \leq \theta \leq \hat{\theta} + e\right]$ güven aralığına düşen kitle parametresi θ için örneklem sayısını belirleyelim.

10.3.1 Kitle ortalamasının tahmini için örneklem büyüklüğü: Kitle parametresi μ için e hata değerine sahip güven aralığı

$$\bar{x} - e \ \leq \mu \leq \bar{x} + e$$

olarak yazılabilir. Kitle varyansının bilindiği ve kitlenin normal dağıldığı durumda μ nün $1 - \alpha$ güven aralığı

$$\bar{x} - \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \leq \mu \ \leq \bar{x} + \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

olup burada hata değerini $e = \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ alınabilir. Bu eşitlikten n değeri çekilirse $n = \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{e}\right]^2$ bulunur. O halde μ nün e hata payına sahip $1 - \alpha$ güven aralığında minimum örneklem büyüklüğü $n = \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{e}\right]^2$ alınabilir. Ancak burada kitle varyansının bilindiği bilgisine ihtiyaç vardır. Bu bilgi önceki çalışmalar yardımıyla temin edilebilir ya da σ yerine bir ön örnekleme yapılıp örneklemin standart sapması s kullanılabilir.

Örnek: Bir fabrikada üretilen çelik kabloların ortalama kaç tonluk bir gerilime dayanabilecekleri araştırılmaktadır. Yöneticiler, kabloların dayanma güçlerinin standart sapması 1 ton olduğunu ve normal dağıldıklarını geçmiş verilerden bilmektedir. Örneklemden hesaplanacak ortalama değer ile gerçek ortalama arasındaki farkın en fazla 0.5 ton olmasını istediklerine göre %90 güven aralığı oluşturmak için kaç kablonun test edilmesi gerekmektedir.

Çözüm: $\alpha = 0.10 \Rightarrow z_{1-\frac{\alpha}{2}} = Z_{0.95} = 1.64$, $e = |\mu - \bar{x}| = 0.5$, $\sigma = 1$ olup

$$n = \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{e}\right]^2 = \left[1.64 * \frac{1}{0.5}\right]^2 = 10.76 \cong 11$$

10.3.2 Kitle oranının tahmini için örneklem büyüklüğü: Kitle parametresi Π nin $1-\alpha$ güven aralığı formülünde e hata değeri olmak üzere $e = \left[z_{1-\frac{\alpha}{2}}\, s_p\right] = \left[z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$ alınabilir. Bu eşitlikten n değeri çekilirse $n = \left[z_{1-\frac{\alpha}{2}}\frac{\sqrt{p(1-p)}}{e}\right]^2$ bulunur. O halde Π nin maksimum e hata payına sahip $1-\alpha$ güven aralığında $n = p(1-p)\left[\frac{z_{1-\frac{\alpha}{2}}}{e}\right]^2$ örneklem büyüklüğü olarak alınabilir. Ancak burada örneklem oranı p bilinmediği için geçmiş verilere ilişkin bilgiler ya da bir ön örnekleme yapılıp p tahmin edilebilir. Herhangi bir bilgiye sahip değilsek $p = 1/2$ alınırsa $p(1-p) = 1/4$ değeri maksimum olacağından n nin maksimum değerini bulmuş oluruz. Böylece $n = \frac{1}{4}\left[\frac{z_{1-\frac{\alpha}{2}}}{e}\right]^2$ alınabilir.

Örnek: Bir fabrikada üretilen ampullerin defoluluk yüzdesi tahmin edilmek isteniyor. En fazla %3 lük bir hata payı ile %95 güven aralığında

a) Defoluluk oranının geçmiş bilgilere göre yaklaşık %10 bilgisiyle
b) Defoluluk oranına ilişkin her hangi bir bilginin olmadığında

durumlar için örneklem büyüklüğünü belirleyiniz.

Çözüm: $\alpha = 0.05$ olup $z_{1-\frac{\alpha}{2}} = 1.96$ ve $e = 0.03$ dür.

a) Geçmiş bilgilere göre $p = 0.10$ kullanılırsa örneklem büyüklüğü

$$n = p(1-p)\left[\frac{z_{1-\frac{\alpha}{2}}}{e}\right]^2 = 0.10(1-0.10)\left[\frac{1.96}{0.03}\right]^2 = 384.16 \cong 384$$

b) Herhangi bir bilgiye sahip değilsek örneklem büyüklüğü

$$n = \frac{1}{4}\left[\frac{z_{1-\frac{\alpha}{2}}}{e}\right]^2 = \frac{1}{4}\left[\frac{1.96}{0.03}\right]^2 = 1067.11 \cong 1067$$

11. HİPOTEZ TESTİ

Kitlenin bilinmeyen parametrelerinin (μ, σ^2, Π) değerleri hakkındaki varsayıma *istatistiksel hipotez* denir.

Örneğin; iplik üretimi yapan bir firma iplik mukavemetinin 5kg olduğunu iddia etsin. Bu iddianın doğruluğu için tüm üretilen ipliklerin mukavemetini ölçmek imkânsız olup belirlenen örneklem üzerinde iddia test edilir ve bir hata payı ile iddia ret ya da kabul edilir.

Kitle parametresi hakkındaki eşitlik hipotezine H_0 sıfır hipotezi denir. H_0 hipotezi test edilir ve iddia ret yada kabul edilir. H_0 hipotezi reddedildiğinde kabul edilecek olan farklılık (eşit değildir, azdır veya fazladır) hipotezine de H_a veya H_1 alternatif hipotez denir.

❖ Kitle ortalaması μ nün iddia edilen bir μ_0 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$

❖ Kitle oranı Π nün iddia edilen bir Π_0 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \Pi = \Pi_0$ $H_a: \Pi \neq \Pi_0$	$H_0: \Pi = \Pi_0$ $H_a: \Pi < \Pi_0$	$H_0: \Pi = \Pi_0$ $H_a: \Pi > \Pi_0$

❖ Kitle varyansı σ^2 nün iddia edilen bir σ_0^2 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$

Örnek: "PAÜ deki öğrencilerin aylık ortalama harcama miktarları 1000 TL den farklıdır" iddiası için $\mu_0 = 1000$ TL olup çift yönlü hipotezler

$$H_0: \mu = 1000\ TL$$
$$H_a: \mu \neq 1000\ TL$$

Örnek: "PAÜ deki öğrencilerin kütüphanede geçirdikleri süre 30 dk dan azdır" iddiası için $\mu_0 = 30$ dk olup tek yönlü hipotezler

$$H_0: \mu = 30\ dk$$
$$H_a: \mu < 30\ dk$$

Örnek: "PAÜ deki öğrencilerin okula gidiş dönüşte harcadıkları ortalama süre 80 dk dan fazladır" iddiası için $\mu_0 = 80$ dk olup tek yönlü hipotezler

$$H_0: \mu = 80\ dk$$
$$H_a: \mu > 80\ dk$$

Örnek: "PAÜ deki öğrencilerin %80 si üniversitenin sosyal ve kültürel imkânlarından faydalanmamaktadır" iddiası için $\Pi_0 = 0.80$ olup çift yönlü hipotezler

$$H_0: \Pi = 0.80$$
$$H_a: \Pi \neq 0.80$$

Örnek: "PAÜ deki öğrencilerin aylık harcama miktarlarının varyansı 25 dir" iddiası için $\sigma_0^2 = 25$ olup çift yönlü hipotezler

$$H_0: \sigma^2 = 25$$
$$H_a: \sigma^2 \neq 25$$

- **Karar Kuralı:** H_0 hipotezinin ret ya da kabul edilmesine, kitleden çekilen örneklemden elde edilen istatistikler kullanılarak hesaplanan *test istatistiğiyle* karar verilir. Burada dikkat edilmesi gereken durum H_0 hipotezinin reddedilmesi hipotezin kesin olarak yanlış olması anlamına gelmez. Aslında H_0 hipotezini ret etmeye yönelik güçlü kanıtlarımızın olmasıdır. Benzer şekilde H_0 kabul edilmesinde aslında hipotezi ret etmeye yönelik kanıtlarımızın olmamasıdır.

  İstatistiksel hipotezde gerçek durum tam anlamıyla bilinmediği için testin sonucunda aşağıdaki 4 durumla karşılaşırız.

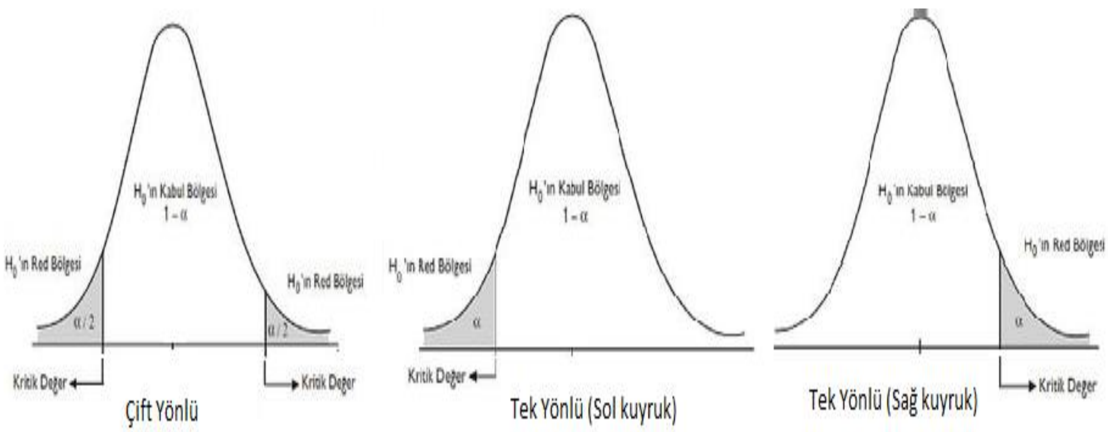
Testin Sonucu	Gerçek Durum	
	H_0 Doğru	**H_0 Yanlış**
H_0 Kabul edildi	Doğru Karar $(1-\alpha)$	Yanlış Karar (II. Tip Hata=β)
H_0 Ret edildi	Yanlış Karar (I. Tip Hata=α)	Doğru Karar $(1-\beta)$

Tabloda "**anlamlılık seviyesi**" olarak adlandırılan α değeri, gerçekte H_0 doğru iken H_0 ı reddetme olasılığı olup bu duruma **I. Tip Hata** denir. β ise gerçekte H_0 yanlış iken H_0 ı kabul edilmesi olasılığı olup bu duruma da **II. Tip Hata** denir. Burada $1-\alpha$ güven aralığımızı $1-\beta$ da testin gücünü gösterir.

İstatistiksel hipotez testlerinde sıkla karşılaşılan bir diğer kavramda ***p değeri*** (p-value, sig, significance) dir. p değeri H_0 hipotezinin reddedilmesine yol açacak en küçük anlamlılık seviyesi (α_{min}) dir. O zaman belirlenen bir α değerine göre tek yönlü hipotezlerde $\boldsymbol{p\ değeri > \alpha}$ çift yönlü hipotezlerde $\boldsymbol{p\ değeri > \alpha/2}$ ise H_0 reddedilemez.

❖ **Hipotez testinde uygulanacak adımlar:**

- H_0 ve H_a hipotezleri yazılır.
- H_a hipotezine bakılarak hipotezin türüne (çift yönlü ya da tek yönlü) karar verilir ve H_0 için ret ve kabul bölgeleri belirlenir.



- Kitleden çekilen n büyüklüğündeki örneklemin gerekli istatistikleri hesaplanır.
- Örneklem istatistikleri kullanılarak bir hesap değeri bulunur ve buna test istatistiği denir.
- Belirlenen α anlamlılık seviyesine göre kritik tablo değeri belirlenir.
- Bulunan hesap değeri kritik tablo değeri ile karşılaştırılır. Eğer hesap değeri ret bölgesine düşerse H_0 hipotezi reddedilir. Yani H_a kabul edilir. Aksi halde H_0 kabul edilir.

11.1 Tek kitle ortalaması (μ) için hipotez testi: Kitle ortalaması μ nün iddia edilen bir μ_0 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$
$H_a: \mu \neq \mu_0$	$H_a: \mu < \mu_0$	$H_a: \mu > \mu_0$

şeklinde kurulur.

Kitleden çekilen n büyüklüğündeki örneklemin ortalaması $\bar{x}$ ve varyansı s^2 olsun. α anlamlılık seviyesinde ($1-\alpha$ güven aralığında) kitle ortalamasının hipotez testi için, aralık tahmininde olduğu gibi varyansının bilinip bilinmediği ve kitlenin normal dağılıma sahip olup olmadığı durumlar için farklı formüller kullanılmaktadır.

❖ σ^2 biliniyor ve kitle normal ise n ne olursa olsun ya da kitle normal değil $n \geq 30$ ise hipotez testi için gerekli olan hesap değeri $z_h = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ ile hesaplanır.

Eğer σ^2 bilinmiyor ve kitle normal olsun olmasın $n \geq 30$ ise hipotez testi için gerekli olan hesap değeri $z_h = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ ile hesaplanır. Elde edilen z_h hesap değeri kritik z_t tablo değeri ile karşılaştırılır. Burada hipotez;

Çift yönlü ise $z_t = z_{1-\frac{\alpha}{2}}$ olup $z_h > z_t$ ya da $z_h < -z_t$ ise H_0 reddedilir.

Tek yönlü sağ kuyruk ise $z_t = z_{1-\alpha}$ olup $z_h > z_t$ ise H_0 reddedilir.

Tek yönlü sol kuyruk ise $z_t = z_{1-\alpha}$ olup $z_h < -z_t$ ise H_0 reddedilir.

❖ σ^2 bilinmiyor ve kitle normal olsun $n < 30$ ise hipotez testi için gerekli olan hesap değeri $t_h = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ ile hesaplanır. Elde edilen t_h hesap değeri kritik t_t tablo değeri ile karşılaştırılır. Burada hipotez;

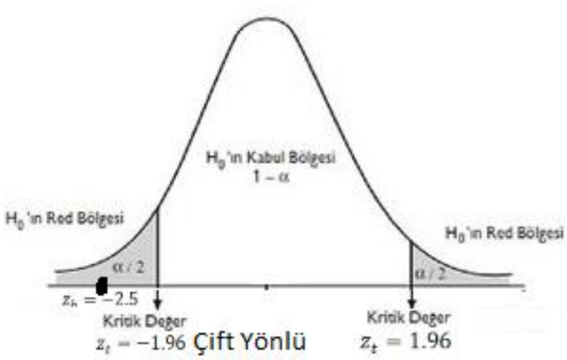
Çift yönlü ise $t_t = t_{1-\frac{\alpha}{2};(n-1)}$ olup $t_h > t_t$ ya da $t < -t_t$ ise H_0 reddedilir.

Tek yönlü sağ kuyruk ise $t_t = t_{1-\alpha;(n-1)}$ olup $t_h > t_t$ ise H_0 reddedilir.

Tek yönlü sol kuyruk ise $t_t = t_{1-\alpha;(n-1)}$ olup $t_h < -t_t$ ise H_0 reddedilir.

Örnek: Bir makinanın üretim hızının 10 m/sa olduğu iddia edilmektedir. Önceki bilgilere dayanarak üretim hızı verisinin normal olup varyansı 1 dir. Bu makinenin hızı rassal olarak 25 kez ölçülmüş ve bu ölçümlerinin ortalaması 9.5 m/sa çıkmıştır. $\alpha = 0.05$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: $\sigma^2 = 1$ biliniyor ve kitle normal olduğu için n ne olursa olsun test için hesap değeri olarak $z_h = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ alınır. İddia edilen ortalama $\mu_0 = 10m/sa$ olup hipotezler;



$$H_0: \mu = 10m/sa$$
$$H_a: \mu \neq 10m/sa$$

şeklinde kurulur.

Örneklem istatistikleri $n = 25$ ve $\bar{x} = 9.5$ olup $z_h = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{9.5-10}{1/\sqrt{25}} = -2.5$ olur.

Hipotez çift yönlü olup $\alpha = 0.05$ anlamlılık seviyesinde tablo değeri $z_t = z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ dır.

$z_h = -2.5 < -z_t = -1.96$ olup H_0 hipotezi reddedilir. O halde H_a alternatif hipotez kabul edilir. Yani makinanın ortalama üretim hızın 10m/sa den farklıdır. Ya da elde edilen verilere göre bu makinanın ortalama üretim hızının 10m/sa olduğu söylenemez.

Örnek: 30 km/sa hızla giden bir arabanın fren yapınca durabilmesi için gerekli ortalama mesafenin 7 metre olduğu iddia edilmektedir. 100 şoför tecrübe edilerek ortalama fren mesafesinin 7.5 metre ve standart sapmasının da 1.5 metre olduğu gözlemlenmiştir. 0.01 anlam seviyesinde iddiayı test ediniz.

Çözüm: σ^2 bilinmeyip $n = 100 \geq 30$ olduğu için hesap değeri olarak $z_h = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ alınır. İddia edilen ortalama $\mu_0 = 7$ olup hipotezler;

$$H_0: \mu = 7\ metre$$
$$H_a: \mu \neq 7\ metre$$

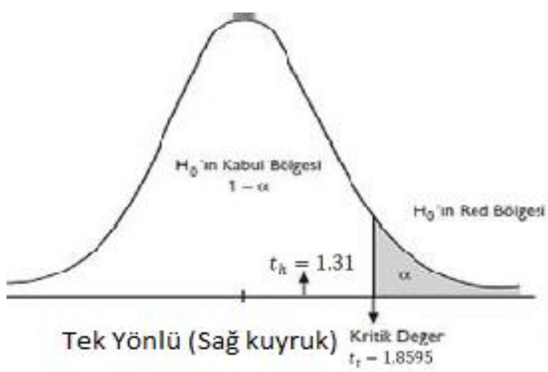
şeklinde kurulur. $n = 100$, $\bar{x} = 7.5$, $s = 1.5$ olup $z_h = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{7.5-7}{1.5/\sqrt{100}} = 3.33$.

Hipotez çift yönlü olup $\alpha = 0.01$ için tablo değeri $z_t = z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$ dir.

$z_h = 3.33 > z_t = 2.58$ olup H_0 hipotezi reddedilir. O halde H_a alternatif hipotez kabul edilmiş olup arabanın fren yapınca durabilmesi için gerekli ortalama mesafe 7 metreden farklıdır.

Örnek: Bir sınıftaki öğrencilerin cep telefonu ile bir haftada yaptıkları ortalama sosyal medya paylaşım sayısının 5 den fazla olduğu iddia edilmektedir. Bu iddiayı test etmek için sınıftan rasgele seçilen 9 kişinin paylaşım sayıları $\{6, 6, 7, 5, 8, 4, 6, 2, 10\}$ olarak belirlenmiştir. Verilerin normal dağıldığı varsayımı altında $\alpha = 0.05$ anlam seviyesinde iddiayı test ediniz.

Çözüm: σ^2 bilinmiyor veriler normal ve $n = 9 < 30$ olduğu için hesap değeri olarak $t_h = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ ve t_t kritik t tablosu değeri ile karşılaştırılır. İddia edilen ortalama $\mu_0 = 5$ olup tek yönlü hipotezler;



$$H_0: \mu = 5$$
$$H_a: \mu > 5$$

şeklindedir. $\bar{x} = \frac{6+\cdots+10}{9} = 6$

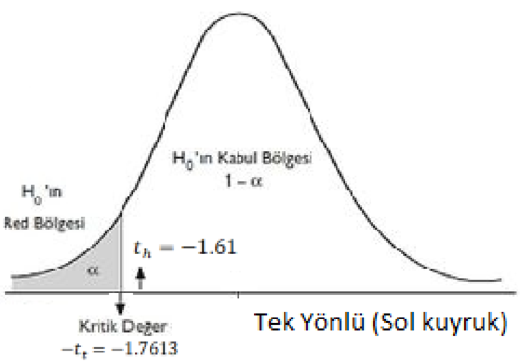
$$s^2 = \frac{(6-6)^2+\cdots+(10-6)^2}{9-1} = 5.25 \Rightarrow s = 2.29 \text{ olup}$$

$t_h = \frac{6-5}{2.29/\sqrt{9}} = 1.31$ olur.

Hipotez tek yönlü sağ kuyruk olup $\alpha = 0.05$ için
$t_t = t_{1-\alpha;(n-1)} = t_{1-0.05;(9-1)} = t_{0.95;8} = 1.8595$ dir. $t_h = 1.31 < t_t = 1.8595$ olup H_0 reddedilemez yani kabul edilir. Sınıftaki öğrencilerin ortalama paylaşım sayıları istatistiksel olarak 5 den fazla değildir.

Örnek: Bir elektrikli süpürge üreticisi ürettiği süpürgelerin yıllık ortalama 50 kWh (kilovat saat) az elektrik tükettiğini iddia etmektedir. Planlanan bir çalışma ile 15 evden oluşan rastgele bir örneklem seçilmiş ve elektrik süpürgelerinin 12 kWh bir standart sapma ile yıllık ortalama 45 kWh elektrik tükettikleri tespit edilmiştir. Verilerin normal dağıldığı varsayımı altında $\alpha = 0.05$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: İddia edilen ortalama $\mu_0 = 50$ kWh olup tek yönlü hipotezler;



$$H_0: \mu = 50 \text{ kWh}$$
$$H_a: \mu < 50 \text{ kWh}$$

şeklindedir.

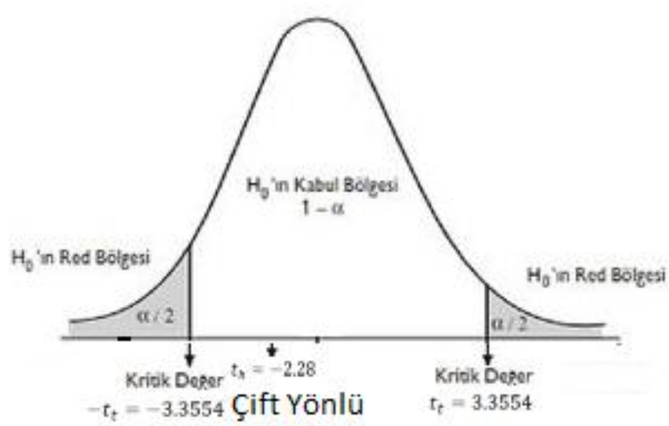
$\bar{x} = 45 \rightarrow s = 12 \rightarrow n = 15 < 30$ olup $t_h = \frac{45-50}{12/\sqrt{15}} = -1.61$ olur.

Hipotez tek yönlü sol kuyruk olup $\alpha = 0.05$ için $t_t = t_{1-\alpha;(n-1)} = t_{1-0.05;\,(15-1)} = t_{0.95;\,14} = 1.7613$ dir. $t_h = -1.61 > -t_t = -1.7613$ olup H_0 reddedilemez.

Yani elektrik süpürgelerinin harcadıkları yıllık kWh miktarının ortalamasının anlamlı olarak 50 den daha az değildir.

Örnek: Ortalama iyileşme süresinin 15 gün olduğu iddia edilen yeni bir ilaç rasgele seçilen 9 hastaya uygulanış ve iyileşme süreleri $\{5, 9, 10, 12, 12, 12, 14, 15, 19\}$ gün olarak tespit edilmiştir. Verilerin normal dağılımdan geldiği varsayımı altında $\alpha = 0.01$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: : İddia edilen ortalama $\mu_0 = 15$ gün olup hipotezler;



$$H_0: \mu = 15$$
$$H_a: \mu \neq 15$$

şeklinde kurulur. Örneklem istatistikleri

$$\bar{x} = \frac{5 + \cdots + 19}{9} = 12 \rightarrow$$
$$s^2 = 15.5 \rightarrow s = 3.94$$

ve $n = 9 < 30$ olup $t_h = \frac{12-15}{3.94/\sqrt{9}} = -2.28$ olur.

Hipotez çift yönlü olup $\alpha = 0.01$ anlamlılık seviyesinde tablo değeri $t_t = t_{1-\frac{\alpha}{2};(n-1)} = t_{1-\frac{0.01}{2};(9-1)} = t_{0.995;8} = 3.3554$ olup $t_h = -2.28 > -t_t = -3.3554$ olduğundan ise H_0 reddedilemez.

11.2 Tek kitle oranı (Π) için hipotez testi: Kitle oranı Π nün iddia edilen bir Π_0 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \Pi = \Pi_0$	$H_0: \Pi = \Pi_0$	$H_0: \Pi = \Pi_0$
$H_a: \Pi \neq \Pi_0$	$H_a: \Pi < \Pi_0$	$H_a: \Pi > \Pi_0$

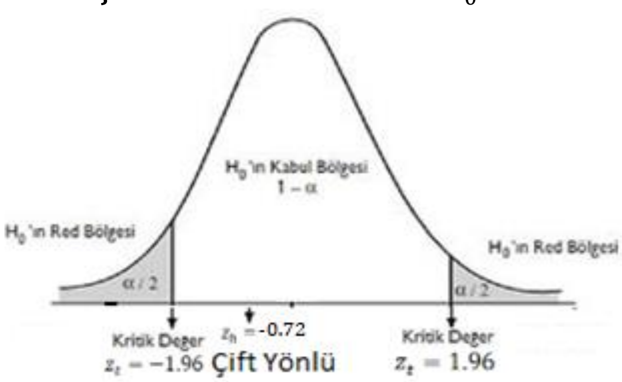
şeklinde kurulur.

Kitleden çekilen n büyüklüğündeki örneklemin oranı p ve örnekleme dağılımının standart sapması $\sigma_p = \sqrt{\Pi_0(1-\Pi_0)/n}$ olmak üzere α anlamlılık seviyesinde ($1-\alpha$ güven aralığında) kitle oranının hipotez testi için gerekli olan hesap değeri $z_h = \frac{p-\Pi_0}{\sigma_p}$ ile hesaplanır. Elde edilen z_h hesap değeri kritik z_t tablo değeri ile karşılaştırılır. Burada hipotez;

Çift yönlü ise $z_t = z_{1-\frac{\alpha}{2}}$ olup $z_h > z_t$ ya da $z_h < -z_t$ ise H_0 reddedilir.
Tek yönlü sağ kuyruk ise $z_t = z_{1-\alpha}$ olup $z_h > z_t$ ise H_0 reddedilir.
Tek yönlü sol kuyruk ise $z_t = z_{1-\alpha}$ olup $z_h < -z_t$ ise H_0 reddedilir.

Örnek: Günde bir paket sigara içenlerin akciğer kanserine yakalanma oranının %61 olduğu iddia edilmektedir. Bu iddiayı test etmek için akciğer kanserine yakalanan 50 kişinin 28 i günde 1 paket sigara içtiği tespit edilmiştir. $\alpha = 0.05$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: İddia edilen oran $\Pi_0 = 0.61$ olup hipotezler;



$$H_0: \Pi = 0.61$$
$$H_a: \Pi \neq 0.61$$

şeklinde kurulur.

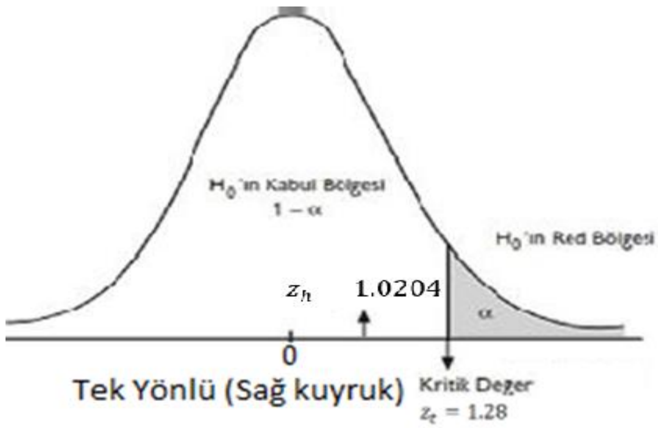
$n = 50, r = 28$ olup $p = \frac{28}{50} = 0.56$ ve

$\sigma_p = \sqrt{\frac{0.61(1-0.61)}{50}} = 0.069$dir. Buradan

$z_h = \frac{p-\Pi_0}{\sigma_p} = \frac{0.56-0.61}{0.069} = -0.7246$ olur.

Hipotez çift yönlü olup $\alpha = 0.05$ anlamlılık seviyesinde tablo değeri $z_t = z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ dır. $z_h = -0.7246 > -z_t = -1.96$ olup H_0 hipotezi reddedilemez. O halde günde bir paket sigara içenlerin akciğer kanserine yakalanma oranını %61 den farklı değildir.

Örnek: Yaygın olarak reçete edilen sinirsel gerginliği düşürücü bir ilacın %60 dan daha fazla etkin olduğuna inanılmaktadır. Rasgele bir örnekleme ile sinirsel gerilimi olan 100 yetişkine uygulanan bu ilacın deneysel sonuçları, bunların 65 inde rahatlama kaydedildiğini göstermiştir. Bu ilacın etkin oranının iddiasını doğrulayacak yeterli kanıt var mıdır? $\alpha = 0.10$ anlamlılık seviyesinde test ediniz.
Çözüm: İddia edilen oran $\Pi_0 = 0.60$ olup hipotezler;



$$H_0: \Pi = 0.60$$
$$H_a: \Pi > 0.60$$

şeklindedir. $n = 100$, $p = \frac{65}{100} = 0.65$ ve $\sigma_p = \sqrt{\frac{0.60(1-0.60)}{100}} = 0.049$ dir. Buradan

$$z_h = \frac{p-\Pi_0}{\sigma_p} = \frac{0.65-0.60}{0.049} = 1.0204 \text{ olur.}$$

Hipotez tek yönlü sağ kuyruk olup $\alpha = 0.10$ anlamlılık seviyesinde tablo değeri $z_t = z_{1-\alpha} = z_{0.90} = 1.28$ dır.

$z_h = 1.0204 < z_t = 1.28$ olup H_0 hipotezi reddedilemez. Yani bu ilacın %60 dan daha fazla etkin olduğunu gösteren yeterli kanıt yoktur.

Örnek: Üniversiteyi kazanan öğrencilerin %70 den azının yurtlarda kaldığı iddia edilmektedir. Rasgele seçilen 400 öğrenciden 260 ının yurtlarda kaldığı tespit edildiğine göre iddiayı %95 güven aralığında test ediniz.
Çözüm: İddia edilen oran $\Pi_0 = 0.70$ olup hipotezler;

$$H_0: \Pi = 0.70$$
$$H_a: \Pi < 0.70$$

şeklindedir. $n = 400$, $p = \frac{260}{400} = 0.65$ ve $\sigma_p = \sqrt{\frac{0.70(1-0.70)}{400}} = 0.0229$ dir.
Buradan $z_h = \frac{p-\Pi_0}{\sigma_p} = \frac{0.65-0.70}{0.0229} = -2.1834$ olur. Hipotez tek yönlü sol kuyruk olup $\alpha = 0.05$ anlamlılık seviyesinde tablo değeri $z_t = z_{1-\alpha} = z_{0.95} = 1.64$ dır. $z_h = -2.1834 < -z_t = 1.64$ olup H_0 hipotezi reddedilir.

Yani öğrencilerin %70 dan azının yurtlarda kaldığı iddiası kabul edilir.

11.3 Tek kitle varyansı (σ^2) için hipotez testi:

11.3 Tek kitle varyansı (σ^2) için hipotez testi: Kitle varyansı σ^2 nün iddia edilen bir σ_0^2 değeri için hipotezler;

Eşit değil (Çift yönlü)	Azdır (Tek yönlü)	Fazladır (Tek yönlü)
$H_0: \sigma^2 = \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$
$H_a: \sigma^2 \neq \sigma_0^2$	$H_a: \sigma^2 < \sigma_0^2$	$H_a: \sigma^2 > \sigma_0^2$

Normal dağılıma sahip bir kitleden çekilen n büyüklüğündeki örneklemin varyansı s^2 olmak üzere α anlamlılık seviyesinde ($1 - \alpha$ güven aralığında) kitle varyansının hipotez testi için gerekli olan hesap değeri $\chi_h^2 = \frac{(n-1)s^2}{\sigma_0^2}$ ile hesaplanır. Elde edilen χ_h^2 hesap değeri kritik tablo değeri ile karşılaştırılır. Burada hipotez;

<u>Çift yönlü</u> $\chi_h^2 < \chi^2_{\frac{\alpha}{2};\,(n-1)}$ ise $\chi_h^2 > \chi^2_{1-\frac{\alpha}{2};\,(n-1)}$ ya da H_0 reddedilir.

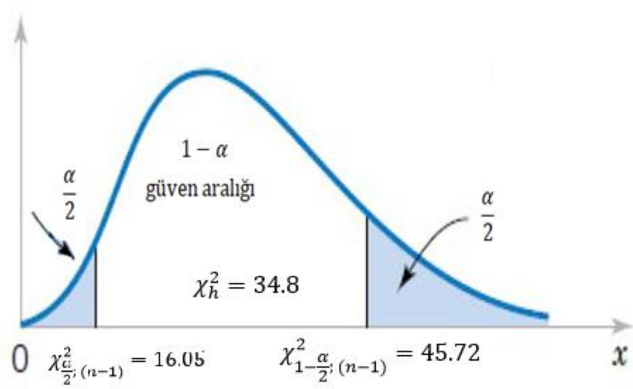
<u>Tek yönlü sol kuyruk</u> $\chi_h^2 < \chi^2_{\alpha;(n-1)}$ ise H_0 reddedilir.

<u>Tek yönlü sağ kuyruk</u> $\chi_h^2 > \chi^2_{1-\alpha;\,(n-1)}$ ise H_0 reddedilir.

Örnek: Bir cins pilin ömür süresinin varyansının 500 saat den farklı olduğu varsayılmaktadır. Bu durumu test etmek için 30 pilden oluşan bir örneklem ortalaması 1000 saat ve varyansı 600 saat olarak bulunmuştur. $\alpha = 0.05$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: İddia edilen oran $\sigma_0^2 = 500$ olup hipotezler;

$$H_0: \sigma^2 = 500$$
$$H_a: \sigma^2 \neq 500$$



şeklindedir. $n = 30$, $\bar{x} = 1000$ ve $s^2 = 600$ olup $\chi_h^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)*600}{500} = 34.8$ olur. Hipotez çift yönlü olup $\alpha = 0.05$ anlamlılık seviyesinde tablo değeri

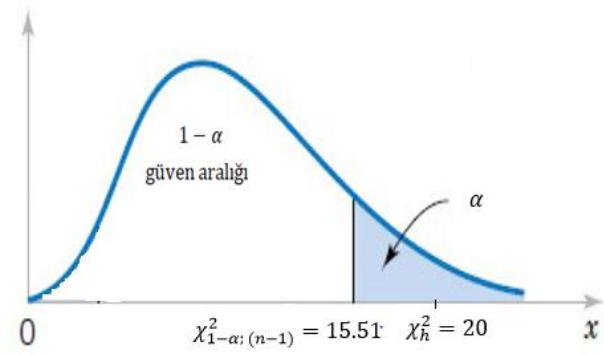
$\chi^2_{\frac{\alpha}{2};\,(n-1)} = \chi^2_{\frac{0.05}{2};\,(30-1)} = \chi^2_{0.025;29} = 16.05$ ve $\chi^2_{1-\frac{\alpha}{2};\,(n-1)} = \chi^2_{0.975;29} = 45.72$ dir.

$\chi^2_{\frac{\alpha}{2};\,(n-1)} = 16.05 < \chi_h^2 = 34.8 < \chi^2_{1-\frac{\alpha}{2};\,(n-1)} = 45.72$ olduğundan H_0 hipotezi reddedilemez. Yani kitlenin varyansı 500 değerinden farklı değidir.

Örnek: Denizli de aylık kültürel harcamaya ilişkin varyansın 400 den fazla olduğu iddia edilmektedir. Rasgele seçilen 9 örnek birimin ortalaması 600 ve varyansı 1000 olarak hesaplanmıştır. Kitle dağılımının normal olduğu bilgisiyle $\alpha = 0.05$ anlamlılık seviyesinde iddiayı test ediniz.

Çözüm: İddia edilen oran $\sigma_0^2 = 400$ olup hipotezler;

$$H_0: \sigma^2 = 400$$
$$H_a: \sigma^2 > 400$$



şeklindedir. $n = 9$, $\bar{x} = 600$ ve $s^2 = 1000$ olarak hesaplanmıştır.

$\chi_h^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(9-1)*1000}{400} = 20$ olur.

Hipotez tek yönlü olup $\alpha = 0.05$ anlamlılık seviyesinde tablo değeri $\chi^2_{1-\alpha;\,(n-1)} = \chi^2_{1-0.05;\,(9-1)} = \chi^2_{0.95;\,8} = 15.51$ dir. $\chi_h^2 = 20 > \chi^2_{0.95;8} = 15.51$ olduğundan H_0 hipotezi reddedilir. Yani kitlenin varyansının 400 değerinden fazladır.

12. İLİŞKİ KATSAYILARI

İki değişken arasındaki ilişkinin belirlenmesinde ve yorumlanmasında ilişki katsayıları sıklıkla kullanılır. Örneğin, Cinsiyet ile gözlük kullanımı, ülkenin gelişmişlik sıralaması ile çevre kirliliği sıralaması, öğrencinin sınav puanı ile sınava çalışma süresi arasındaki ilişkiler gibi.

Yaygın olarak kategorik değişkenler arasındaki ilişkiyi çapraz tablolar yardımıyla Cramer's v katsayısı ile, sayısal değişkenler arasındaki ilişkiyi Spearman's ρ katsayısı ve Pearson's r katsayısı ile incelenir.

İlişki katsayılarının mutlak değerce alabileceği değerlerin genel yorumu için aşağıdaki tablodan yararlanılır.

Katsayı	Yorum
0.00-0.19	Çok zayıf ilişki
0.20-0.39	Zayıf ilişki
0.40-0.69	Orta düzeyde ilişki
0.70-0.89	Güçlü ilişki
0.90-1.00	Çok güçlü ilişki

12.1 Cramer's v katsayısı: Kategorik değişkenler arasındaki ilişkilerin gösterilmesinde çapraz tablolar ya da kontenjans tabloları kullanılır. İki ya da daha fazla gruba sahip iki kategorik değişken için çapraz tabloların genel gösterimi aşağıdaki gibidir.

		DEĞİŞKEN II				
		Grup 1	Grup 2	...	Grup C	Toplam
DEĞİŞKEN I	Grup 1	G_{11}	G_{12}	...	G_{1C}	R_1
	Grup 2	G_{21}	G_{22}	...	G_{2C}	R_2
	⋮	⋮	⋮	⋮	⋮	⋮
	Grup C	G_{R1}	G_{R2}	...	G_{RC}	R_R
	Toplam	C_1	C_2	...	C_C	n

- ❖ $i = 1,2,\dots,R$ olup Değişken I de R tane grup vardır.
- ❖ $j = 1,2,\dots,C$ olup Değişken II de C tane grup vardır
- ❖ G_{ij}: i inci satır ve j inci sütundaki gözlenen frekanslar.
- ❖ R_i: i inci satır toplamı.
- ❖ C_j: j inci sütun toplamı.

- n: verideki toplam gözlem sayısı olmak üzere
- $B_{ij} = \frac{R_i * C_j}{n}$ ile hesaplanan B_{ij} ye i inci satır ve j inci sütun beklenen frekans denir
- $\chi_h^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(G_{ij} - B_{ij})^2}{B_{ij}}$ değerine de ki-kare hesap değeri denir.

RxC türündeki çapraz tablo ile ifade edilen iki kategorik değişken arasındaki ilişkiyi için Cramer's in v katsayısı

$$v = \sqrt{\frac{\chi_h^2}{n * min\{(R-1),(C-1)\}}}$$

şeklinde hesaplanır. v katsayısı sıfır ile bir arasında değerler alır $(0 \leq v \leq 1)$. $v = 0$ olması iki değişken arasında ilişkinin olmadığı, $v = 1$ olması da iki değişken arasında tam ilişki olduğu anlamına gelir. Sıfıra yakın değerler zayıf ilişkiyi, bire yakın değerlerde güçlü ilişkiyi gösterir. Katsayının yorumu için verilen tablodan yararlanılabilir.

Örnek: Matematik bölümü öğrencilerinin cinsiyetleri ile istatistik dersinde başarılı olmaları arasında bir ilişki olup olmadığını verilen çapraz tablo yardımıyla inceleyiniz ve yorumlayınız.

	Başarı Durumu	
Cinsiyet	*Başarılı*	*Başarısız*
Erkek	30	40
Kadın	10	20

Çözüm: Değişken I: Cinsiyet(Erkek-Kadın) değişkeni $R = 2$ gruba sahip, Değişken II: Başarı Durumu (Başarılı-Başarısız) değişkeni de C= 2 gruba sahiptir. Tabloda yer alan gözlenen frekanslar ile satır ve sütun toplamlarını kullanarak beklenen frekanslar B_{ij};

	Başarı Durumu		
Cinsiyet	*Başarılı*	*Başarısız*	*Toplam*
Erkek	$G_{11} = 30$	$G_{12} = 40$	$R_1 = 70$
Kadın	$G_{21} = 10$	$G_{22} = 20$	$R_2 = 30$
Toplam	$C_1 = 40$	$C_2 = 60$	$n = 100$

$$B_{11} = \frac{R_1 * C_1}{n} = \frac{70 * 40}{100} = 28$$
$$B_{12} = \frac{R_1 * C_2}{n} = \frac{70 * 60}{100} = 42$$
$$B_{21} = \frac{R_2 * C_1}{n} = \frac{30 * 40}{100} = 12$$
$$B_{22} = \frac{R_2 * C_2}{n} = \frac{30 * 60}{100} = 18$$

$$\chi_h^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(G_{ij} - B_{ij})^2}{B_{ij}} = \frac{(30-28)^2}{28} + \frac{(40-42)^2}{42} + \frac{(10-12)^2}{12} + \frac{(20-18)^2}{18} = 0.79$$

$R = 2$ ve $C = 2$ olup $min\{(R-1),(C-1)\} = min\{(2-1),(2-1)\} = min\{1,1\} = 1$

$$v = \sqrt{\frac{\chi_h^2}{n * min\{(R-1),(C-1)\}}} = \sqrt{\frac{0.79}{100 * 1}} = 0.09$$

O halde cinsiyet ile başarı durumu arasındaki ilişki katsayısı $v = 0.09$ olup sıfıra çok yakın olduğu için iki değişken arasında ilişki yok denecek kadar azdır. Ya da ilişki yoktur şeklinde yorumlanır.

Örnek: Bir kurumdaki 200 çalışandan eğitim düzeyi ve en çok izlenen TV program türü değişkenleri için veri derlenmiştir. Verilere göre iki değişken arasındaki ilişki katsayısını hesaplayınız ve yorumlayınız.

	TV program türü				
Eğitim Düzeyi	*Dizi*	*Müzik*	*Haber*	*Diğer*	*Toplam*
İlköğretim	20	10	5	5	40
Lise	10	40	15	35	100
Üniversite	10	10	30	10	60
Toplam	40	60	50	50	200

Çözüm: Çapraz tabloya göre $R = 3$ ve $C = 4$ olup gözlenen frekanslar

$G_{11} = 20, G_{12} = 10, G_{13} = 5, G_{14} = 5,$
$G_{21} = 10, G_{22} = 40, G_{23} = 15, G_{24} = 35,$
$G_{31} = 10, G_{32} = 10, G_{33} = 30, G_{34} = 10$
dir. Satır toplamları $R_1 = 40, R_2 = 100, R_3 = 60$ ve
sütun toplamları $C_1 = 40, C_2 = 60, C_3 = 50, C_4 = 50$ olup $n = 200$ dir.
Beklenen frekanslar;

$B_{11} = \frac{40*40}{200} = 8, B_{12} = \frac{40*60}{200} = 12, B_{13} = \frac{40*50}{200} = 10, B_{14} = \frac{40*50}{200} = 10$
$B_{21} = \frac{100*40}{200} = 20, B_{22} = \frac{100*60}{200} = 30, B_{23} = \frac{100*50}{200} = 25, B_{24} = \frac{100*50}{200} = 25$
$B_{31} = \frac{60*40}{200} = 12, B_{32} = \frac{60*60}{200} = 18, B_{33} = \frac{60*50}{200} = 15, B_{34} = \frac{60*50}{200} = 15$
olup ki-kare hesap değeri

$$\chi_h^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{(G_{ij} - B_{ij})^2}{B_{ij}} = \frac{(20-8)^2}{8} + \frac{(10-12)^2}{12} + \frac{(5-10)^2}{10} + \frac{(5-10)^2}{10}$$
$$+ \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30} + \frac{(15-25)^2}{25} + \frac{(35-25)^2}{25}$$
$$+ \frac{(10-12)^2}{12} + \frac{(10-18)^2}{18} + \frac{(30-15)^2}{15} + \frac{(10-15)^2}{15} = 60.22$$

$R = 3$ ve $C = 4$ olup $min\{(R-1), (C-1)\} = min\{(3-1), (4-1)\} = min\{2,3\} = 2$

$$v = \sqrt{\frac{\chi_h^2}{n * min\{(R-1), (C-1)\}}} = \sqrt{\frac{60.22}{200 * 2}} = 0.39$$

elde edilir. Bu sonuca göre eğitim düzeyi ile en çok izlenen TV program türü arasında zayıf ilişki olduğu söylenir.

12.2 Spearman's ρ katsayısı: Verilerin belirli bir kritere göre sıralanarak elde edilmiş sıralama ya da oran ölçeğiyle ölçülmüş iki değişken arasındaki ilişkinin derecesini (zayıf/güçlü) ve yönünü (aynı yönlü/ters yönlü) belirlenmek için kullanılır.

n gözleme sahip iki sayısal değişken X ve Y olsun. x_i i inci gözlemin X değişkenine göre sıra puanı, y_i de i inci gözlemin Y değişkenine göre sıra puanı olmak üzere Spearman ilişki katsayısı

$$\rho = 1 - \frac{6 * \sum(x_i - y_i)^2}{n * (n^2 - 1)}$$

ile hesaplanır. Burada $-1 \leq \rho \leq +1$ olup, pozitif değerler aynı yönlü ilişkiyi (değişkenlerden birinin değeri artarken diğerinin de artması), negatif değerler ise ters yönlü (değişkenlerden birinin değeri artarken diğerinin azalması) ilişkiyi gösterir. Katsayının 0 ya da yakın değerler alması ilişkinin olmadığı -1 ya da +1 e yakın değerler alması da güçlü ilişkiyi gösterir.

Örnek: Aşağıdaki tabloda 10 kente ait yaşanılabilirlik değişkenine ve sosyokültürel aktivite zenginliği değişkenine göre sıralamaları verilmiştir. İki değişken arasındaki ilişki katsayısını hesaplayınız ve yorumlayınız.

Kentler	Yaşanılabilirlik	Sosyokültürel	Kentler	Yaşanılabilirlik	Sosyokültürel
Kent1	4	7	**Kent6**	9	9
Kent2	6	5	**Kent7**	10	8
Kent3	1	2	**Kent8**	2	3
Kent4	3	1	**Kent9**	5	4
Kent5	7	6	**Kent10**	8	10

Çözüm: Sıra puanlarına sahip değişkenler için Spearman ilişki katsayısı hesaplanır.

Kentler	Yaşanılabilirlik	Sosyokültürel	$(x_i - y_i)^2$
Kent1	4	7	$(4-7)^2 = 9$
Kent2	6	5	$(6-5)^2 = 1$
Kent3	1	2	$(1-2)^2 = 1$
Kent4	3	1	$(3-1)^2 = 4$
Kent5	7	6	$(7-6)^2 = 1$
Kent6	9	9	$(9-9)^2 = 0$
Kent7	10	8	$(10-8)^2 = 4$
Kent8	2	3	$(2-3)^2 = 1$
Kent9	5	4	$(5-4)^2 = 1$
Kent10	8	10	$(8-2)^2 = 4$

$$\begin{aligned} \rho &= 1 - \frac{6 * \sum(x_i - y_i)^2}{n * (n^2 - 1)} \\ &= 1 - \frac{6 * 26}{10 * (10^2 - 1)} \\ &= 0.84 \end{aligned}$$

olup kentlerdeki yaşanılabilirlik değişkeni ile sosyokültürel aktivite değişkeni arasında pozitif (aynı yönlü) ve güçlü (0.84) bir ilişki vardır. Yani yaşanılabilirlikte üst sıralarda olan kent, sosyokültürel aktivite zenginliği bakımından da üst sıralardadır.

Örnek: Yedi şehre ait futbol ve basketbol takımlarının sezon sonu sıralamaları tabloda verilmiştir. Şehirdeki futbol takımının başarısı ile basketbol takımının başarısı arasındaki ilişkiyi ilişki katsayısını bularak yorumlayınız.

Kentler	Futbol	Basketbol
A	1	5
B	2	6
C	3	7
D	4	3
E	5	1
F	6	4
G	7	2

Çözüm: Sıra puanlarına sahip değişkenler için Spearman ilişki katsayısı hesaplanır.

Kentler	Futbol	Basketbol	$(\boldsymbol{x_i - y_i})^2$
A	1	5	$(1-5)^2 = 16$
B	2	6	$(2-6)^2 = 16$
C	3	7	$(3-7)^2 = 16$
D	4	3	$(4-3)^2 = 1$
E	5	1	$(5-1)^2 = 16$
F	6	4	$(6-4)^2 = 4$
G	7	2	$(7-2)^2 = 25$
Toplam			94

$$\rho = 1 - \frac{6 * 94}{7 * (7^2 - 1)} = -68$$

Katsayı negatif olup değişkenler arasında ters yönlü bir ilişki vardır. Yani futbolda iyi olan şehir basketbolda düşük sıralardadır. Katsayının değeri de $\rho = -0.68$ olup orta düzeyde ilişki olduğunu gösterir.

12.3 Pearson's r katsayısı: Sayısal iki değişken arasındaki doğrusal ilişkinin derecesini (zayıf/güçlü) ve yönünü (aynı yönlü/ters yönlü) belirlenmek için kullanılır.

Rasgele seçilen n gözleme sahip iki sayısal değişken X ve Y olsun. X değişkeni için i inci gözlem değeri x_i, aritmetik ortalaması $\bar{x} = \frac{\sum x_i}{n}$ ve standart sapması $s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$ olup benzer şekilde Y değişkeni için i inci gözlem değeri y_i, aritmetik ortalaması $\bar{y} = \frac{\sum y_i}{n}$ ve standart sapması $s_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{n-1}}$ olsun. X ve Y arasındaki kovaryans değeri $s_{xy} = \frac{\sum(x_i-\bar{x})*(y_i-\bar{y})}{n-1}$ olmak üzere Pearson ilişki katsayısı;

$$r = \frac{s_{xy}}{s_x * s_y}$$

ile hesaplanır. Burada $-1 \leq r \leq +1$ olup r nin 0 olması ilişkinin olmadığı, +1 e yakın değerler aynı yönlü (değişkenlerden birinin değeri artarken diğerinin de artması), -1 e yakın değerlerde ters yönlü (değişkenlerden birinin değeri artarken diğerinin azalması) ilişkinin olduğu anlamına gelir.

Örnek: Rasgele seçilen 8 öğrencinin sınava hazırlanma süreleri (saat) ve sınavdan aldıkları puana ilişkin veriler tabloda verilmiştir. Öğrencilerin sınava hazırlanma süreleri ile sınavdan aldıkları puanlar arasındaki ilişkiyi ilişki katsayısını hesaplayarak yorumlayınız.

Öğrenci	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Süre (X)	30	30	25	20	18	16	12	9
Puan (Y)	80	70	70	65	60	50	50	35

Çözüm: Sayısal değişkenler arasındaki ilişki için Pearson ilişki katsayısı kullanılır.

Öğrenci	**X**	**Y**	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})*(y_i-\bar{y})$
1	30	80	10	100	20	400	200
2	30	70	10	100	10	100	100
3	25	70	5	25	10	100	50
4	20	65	0	0	5	25	0
5	18	60	-2	4	0	0	0
6	16	50	-4	16	-10	100	40
7	12	50	-8	64	-10	100	80
8	9	35	-11	121	-25	625	275
Toplam	**160**	**480**	**0**	**430**	**0**	**1450**	**745**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{160}{8} = 20 \rightarrow s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{430}{8-1}} = 7.84$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{480}{8} = 60 \rightarrow s_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{1450}{8-1}} = 14.39$$

$$s_{xy} = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{n-1} = \frac{745}{8-1} = 106.43$$

$$r = \frac{s_{xy}}{s_x * s_y} = \frac{106.43}{7.84 * 14.39} = 0.94$$

İlişki katsayısı $r = 0.94$ pozitif ve bire çok yakın değer aldığı için iki değişken arasında aynı yönü ve çok güçlü bir ilişki vardır. Yani sınava çalışma süresi artan öğrencilerin sınav puanları da artmaktadır.

Örnek: Rasgele seçilen 8 sporcunun haftalık antrenman süreleri (X:saat) ile 100 metreyi koşma süreleri (Y:saniye) tabloda verilmiştr. İki değişken arasındaki ilişkiyi ilişki katsayısını hesaplayarak yorumlayınız.

Sporcu	X(saat)	Y (saniye)
1	10	14
2	12	14
3	14	14
4	16	13
5	18	13
6	20	12
7	22	12
8	24	12

Çözüm: Sayısal değişkenler arasındaki ilişki için Pearson ilişki katsayısı kullanılır.

Sporcu	X	Y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
1	10	14	-7	49	1	1	-7
2	12	14	-5	25	1	1	-5
3	14	14	-3	9	1	1	-3
4	16	13	-1	1	0	0	0
5	18	13	1	1	0	0	0
6	20	12	3	9	-1	1	-3
7	22	12	5	25	-1	1	-5
8	24	12	7	49	-1	1	-7
Toplam	**136**	**104**	**0**	**168**	**0**	**6**	**-30**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{136}{8} = 17 \rightarrow s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{168}{8-1}} = 4.90$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{104}{8} = 13 \rightarrow s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{6}{8-1}} = 0.93$$

$s_{xy} = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{n-1} = \frac{-30}{8-1} = -4.29$ olup $r = \frac{s_{xy}}{s_x * s_y} = \frac{-4.29}{4.90 * 0.93} = -0.94$

İlişki katsayısı $r = -0.94$ negatif ve bire çok yakın değer aldığı için iki değişken arasında ters yönü ve çok güçlü bir ilişki vardır. Yani antrenman süresi artan sporcunun 100 metreyi koşma süresi azalmaktadır.

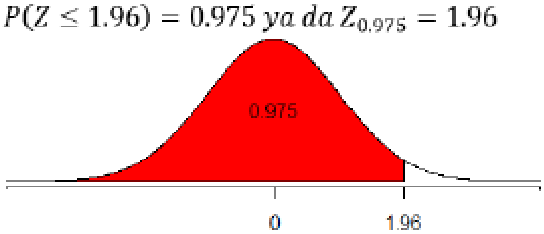
13. TABLOLAR

13.1 Birikimli Standart Normal Dağılım Tablosu

Birikimli (Kümülatif) Standart Normal Dağılım Tablosu

Z_t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

$$P(Z \leq 1.96) = 0.975 \ ya \ da \ Z_{0.975} = 1.96$$



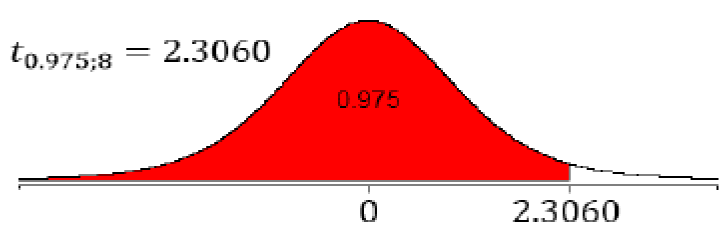
Bazı birikimli olasılıklar için tablo değerleri

Olasılık	0.90	0.95	0.975	0.99	0.995	0.999
Z_t	1.28	1.64	1.96	2.33	2.58	3.09

13.2 Birikimli Student's t- Dağılım Tablosu

Birikimli (Kümülatif) Student's t - Dağılım Tablosu

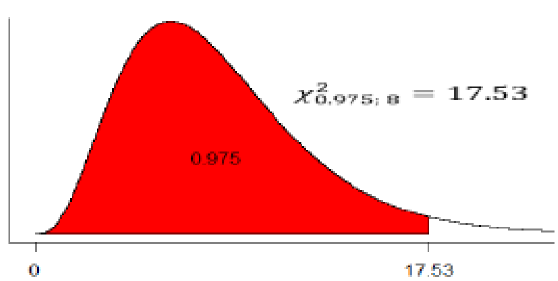
s.d.	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.999}$	$t_{0.9995}$
1	0.7265	1.3764	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	0.5325	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	0.5317	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.5312	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	0.5306	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.5302	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
35	0.5292	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400	3.5911
40	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
45	0.5281	0.8497	1.3006	1.6794	2.0141	2.4121	2.6896	3.2815	3.5203
50	0.5278	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960
60	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.5268	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.5263	0.8456	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
200	0.5258	0.8446	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
∞	0.5256	0.8442	1.2876	1.6558	1.9771	2.3533	2.6114	3.1495	3.3614



13.3 Birikimli Ki-Kare- Dağılım Tablosu

Birikimli (Kümülatif) Ki Kare Dağılım Tablosu

s.d	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.25}$	$\chi^2_{0.50}$	$\chi^2_{0.75}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17



14. FORMÜL KAĞIDI

- *Merkezi eğilim ve Dağılış Ölçüleri*

$AO = \bar{x} = \frac{\sum x_i}{n} = \frac{\sum(x_j * f_j)}{\sum f_j} = \frac{\sum(m_j * f_j)}{\sum f_j}$		$GO = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[\sum f_j]{\prod x_j^{f_j}} = \sqrt[\sum f_j]{\prod m_j^{f_j}}$
$KO = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{\sum(x_j^2 * f_j)}{\sum f_j}} = \sqrt{\frac{\sum(m_j^2 * f_j)}{\sum f_j}}$		$HO = \frac{n}{\sum \frac{1}{x_i}} = \frac{\sum f_j}{\sum \frac{f_j}{x_j}} = \frac{\sum f_j}{\sum \frac{f_j}{m_j}}$
$Mod = L_{Mod} + h_{Mod}\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$		$Medyan(M) = L_M + \frac{h_M}{f_M}\left(\frac{n}{2} - f_B\right)$
$Q_1 = L_1 + \frac{h_1}{f_1}\left(\frac{n}{4} - f_{B1}\right)$		$Q_3 = L_3 + \frac{h_3}{f_3}\left(\frac{3n}{4} - f_{B3}\right)$
$DA = EB - EK$	$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$	$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{\sum(x_j - \bar{x})^2 * f_j}{n-1} = \frac{\sum(m_j - \bar{x})^2 * f_j}{n-1}$
$DK = \frac{s}{\bar{x}}$ veye $DK = \frac{\sigma}{\mu}$	$KA = Q_3 - Q_1$	*Aykırı gözlem sınırları* $(Q_1 - 1.5 * KA)$ ile $(Q_3 + 1.5 * KA)$

- *Olasılık ve Koşullu Olasılık*

$P(A) = \frac{m}{n}$ veya $P(A) = \frac{f}{n}$	$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$	$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$
$P(A \cap B) = P(A) * P(B \mid A) = P(B) * P(A \mid B)$	$P(B_r \mid A) = \frac{P(B_r \cap A)}{P(A)} = \frac{P(B_r) * P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i) * P(A \mid B_i)}$	

- *Rasgele Değişkenler*

$\mu = E(X) = \sum_{i=1}^{N} x_i f(x_i)$	$\sigma^2 = V(X) = \sum_{i=1}^{N} (x_i - \mu)^2 f(x_i) = E(X^2) - (E(X))^2$
$\mu = E(X) = \int_{D_x} x f(x) dx$	$\sigma^2 = V(X) = \int_{D_x} (x - \mu)^2 f(x) dx = E(X^2) - (E(X))^2$

- *Rasgele Değişkenler ve Bazı Önemli Dağılımlar*

Binom	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; $x = 0,1,\dots,n$ olup $E(X) = n * p$ ve $\sigma^2 = n * p * (1-p)$
Geometrik	$f(x) = p * (1-p)^{x-1}$; $x = 1,2,3,\dots$ olup $E(X) = 1/p$ ve $V(X) = \frac{(1-p)}{p^2}$
Poisson	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$; $x = 0,1,2,\dots$ olup $E(X) = V(X) = \lambda$
Düzgün	$f(x) = \frac{1}{b-a}$; $a \le x \le b$ olup $E(X) = \frac{a+b}{2}$ ve $V(X) = \frac{(b-a)^2}{12}$
Üstel	$f(t) = \frac{1}{\alpha} e^{-\frac{t}{\alpha}}$; $t > 0$ olup $E(T) = \alpha$ ve $V(T) = \alpha^2$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $-\infty < x, \mu < \infty$; $\sigma > 0$ olup $E(X) = \mu$ ve $V(X) = \sigma^2$ Standartlaştırma: $X \sim N(\mu, \sigma^2)$ ve $Z = \frac{X-\mu}{\sigma}$ ya da $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ise $Z \sim N(0,1)$
Standart Normal	$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $-\infty < z < \infty$ olup $E(Z) = 0$ ve $V(Z) = 1$

- *Parametrelerin* $(1-\alpha)$ *aralık tahminleri*

$\bar{x} - \left[z_{1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}}\right] < \mu < \bar{x} + \left[z_{1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}}\right]$	$\bar{x} - \left[t_{1-\frac{\alpha}{2};(n-1)} * \frac{s}{\sqrt{n}}\right] < \mu < \bar{x} + \left[t_{1-\frac{\alpha}{2};(n-1)} * \frac{s}{\sqrt{n}}\right]$		
$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};(n-1)}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};(n-1)}}$	$p - \left[z_{1-\frac{\alpha}{2}} * s_p\right] \le \Pi \le p + \left[z_{1-\frac{\alpha}{2}} * s_p\right]$	$p = r/n$	$s_p = \sqrt{\frac{p(1-p)}{n}}$

- *Parametrelerin "e" hata payına sahip* $(1-\alpha)$ *güven aralığında minimum örneklem büyüklüğü*

$n = \left[z_{1-\frac{\alpha}{2}} \frac{\sigma}{e}\right]^2$ veya $n = \left[\frac{N\sigma^2 \left(z_{1-\frac{\alpha}{2}}\right)^2}{(N-1)e^2 + \sigma^2 \left(z_{1-\frac{\alpha}{2}}\right)^2}\right]$	$n = p(1-p)\left[\frac{z_{1-\frac{\alpha}{2}}}{e}\right]^2$ veya $n = \left[\frac{Np(1-p)\left(z_{1-\frac{\alpha}{2}}\right)^2}{(N-1)e^2 + p(1-p)\left(z_{1-\frac{\alpha}{2}}\right)^2}\right]$

- *Hipotez testi test istatistikleri*

$z_h = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t_h = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$z_h = \dfrac{p - \Pi_0}{\sigma_p}$	$\sigma_p = \sqrt{\dfrac{\Pi_0(1 - \Pi_0)}{n}}$	$\chi_h^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

$z_h = \dfrac{\bar{D}}{s_D/\sqrt{n}}$	$t_h = \dfrac{\bar{D}}{s_D/\sqrt{n}}$	$t_h = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$sd \cong \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{(n_1 - 1)} + \dfrac{(s_2^2/n_2)^2}{(n_2 - 1)}}$

$z_h = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$t_h = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} * \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$sd = (n_1 - 1) + (n_2 - 1)$

- *İlişki katsayıları*

$\chi_h^2 = \sum\sum \dfrac{\left(G_{ij} - B_{ij}\right)^2}{B_{ij}}$	$B_{ij} = \dfrac{r_i * c_j}{n}$	$\vartheta = \sqrt{\dfrac{\chi_h^2}{n * \min\{(r-1),(c-1)\}}}$	$\rho = 1 - \dfrac{6 * \sum(x_i - y_i)^2}{n * (n^2 - 1)}$

$r = \dfrac{s_{xy}}{s_x * s_y}$ $\quad s_{xy} = \dfrac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{n - 1}$	$r = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i{}^2 - n(\bar{x})^2)(\sum y_i{}^2 - n(\bar{y})^2)}}$

- *Basit Doğrusal Regresyon Analizi*

$\hat{Y} = b_0 + b_1 X$	$b_1 = \dfrac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i{}^2 - (\sum x_i)^2}$	$b_0 = \bar{y} - b_1\bar{x}$	$R^2 = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$