



Benzetim

6- Benzetimde İstatistik ve Olasılık

Dr. Öğr. Üyesi Caner Erden, cerden@subu.edu.tr

Uluslararası Ticaret ve Finansman Bölümü

İstatistik, verilerin toplanması, analiz edilmesi, raporlanması ve yorumlaması ile ilgilenden bir matematiğin bir dalıdır. İstatistik deneysel olaylar ile ilgilenir.

Deney, sonuçların kesin olmadığı durumlar demektir. Deneyin olabilecek tüm sonuçları örneklem uzayını gösterir. S ile gösterilir. Deneysel olaylar çok sayıda faktörden etkilenir ve birkaç deneme ile olayların karakteri belirlenemez.

Örneğin paranın havaya atılmasında iki durum söz konusudur.

$$S = \{Yazı, Tura\}$$

Bir zarın atılması olayında 6 durum söz konusudur.

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

$$S = \{(1, 1), (1, 2), \dots, (6, 6)\}$$

İki zarın atılması olayında zarların yüzlerine gelen sayıların toplamına rassal değişken(X) dersek, örneğin zarların yüzeyine 2,3 geldiğinde $X = 5$ değerini alır.

$$S = \{(Y, T), (T, Y), (Y, Y), (T, T)\}$$

İki paranın atılması olayında üst yüze gelen «Tura» sayısına rassal değişken (X) dersek, örneğin iki para da tura gelirse $X = 2$ değerini alır.

Değişkenleri 2'ye ayırmak mümkündür. Örneğin sıcaklık değeri nümerik(kantitatif, sayılabilen) bir veri iken rüzgar durumu kategorik(kalitatif, sayılamayan) bir veridir. Kantitatif verileri de sürekli ve kesikli olarak 2'ye ayırabiliriz. Kesikli veriler sonlu sayıda elemanı olan verilerdir. Örneğin ailedeki çocuk sayısı, satışı yapılan TV adedi.

Başka bir veri türü sınıflandırması:

- **Nominal**
- **ordinal**
- **interval**
- **ratio**

1. **Nominal veriler:** Daha fazla ifadesi ile kullanılmazlar. ID, göz rengi vb.
2. **Sıralama Ölçeği (Ordinal):** Daha fazla ifadesi ile kullanılabilir. Uzunluk (uzun, orta boy, kısa) vb.
3. **Aralıklı (interval):** Sayısal değerler ile ölçülür. Sıcaklık vb.
4. **Oran (Ratio):** Başlangıç noktası o'dır ve o noktası yokluk ifade eder. Kelvin, uzunluk (cm), zaman gibi.

Rassal değişken sürekli veya kesikli olabilir. Eğer rassal değişken belirli sayıda değerler içerisinde bir değer alıyorsa değişkene kesikli rassal değişken denir. Örneğin tam sayılar $\{1,2,3,\dots\}$ gibi sayılar arasından atanacak rassal değişkenlere kesikli rassal değişken denir. Bir parayı 3 kere havaya attığımızda yazı gelme sayısı rassal kesikli bir değişkene örnektir. Bu sayı her seferinde 0,1,2,3 arasında değişecektir. Veya diğer bir örnekte de 3 kere havaya atılan paranın kaçınıcı kerede tura geleceğinin sayısı olabilir. Bu durumda da değişken 1,2 ve 3 değerleri arasından bir değer alır. Bu ilişkilerle ilgili çizilecek grafiğe olasılık dağılım fonksiyonu adı verilir.

Paranın 3 sefer atılmasında gelebilecek tüm durumlar $2^3=8$ adettir. Bunlar YYY, YYT, YTY, TYY, TTT, TTY, TYT, YTT.

Dağılım fonksiyonu;

Yazı Gelme Sayısı (x)	Olasılık
$x_1 = 0$	$P(x_1) = \frac{1}{8}$
$x_1 = 1$	$P(x_2) = \frac{3}{8}$
$x_1 = 2$	$P(x_3) = \frac{3}{8}$
$x_1 = 3$	$P(x_4) = \frac{1}{8}$

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

$$\sum P(x_i) = 1$$

$$\mu_x = E(X) = \sum x_i \times P(x_i)$$

$$\sigma^2 = \sum (x_i - \mu_x)^2 \times P(x_i)$$

$$\sigma = \sqrt{\sigma^2}$$

3 kere para atılması örneği

$$\mu_x = E(X) = \sum x_i \times P(x_i) = 0 \left(\frac{1}{8}\right) + 1 \left(\frac{3}{8}\right) + 2 \left(\frac{3}{8}\right) + 3 \left(\frac{1}{8}\right) = 1.50$$

Varyansı

$$\sigma^2 = \sum (x_i - \mu_x)^2 \times P(x_i) = (0 - 1.5)^2 \left(\frac{1}{8}\right) + (1 - 1.5)^2 \left(\frac{3}{8}\right) + (2 - 1.5)^2 \left(\frac{3}{8}\right) + (3 - 1.5)^2 \left(\frac{1}{8}\right) = 0.75$$

Ana kütle standart sapması ya da ortalamasının tahmin edilmesi için örneklem dağılımlarından yararlanılır.

Dağılım: Tüm durumları olasılıkları ile belirlemek

Örneğin elimizde 1-2-3-4 sayılarından oluşan bir zar olsun. Olaya ilişkin olasılık dağılımı;

X	$P(x)$
$x_1 = 1$	0.25
$x_2 = 2$	0.25
$x_3 = 3$	0.25
$x_4 = 4$	0.25

Ortalama ve Standart Sapma

$$\mu_x = \sum x_i(P(x_i)) = 1 \left(\frac{1}{4}\right) + 2 \left(\frac{1}{4}\right) + 3 \left(\frac{1}{4}\right) + 4 \left(\frac{1}{4}\right) = 2.5$$

$$\begin{aligned}\sigma_x &= \sqrt{(x_i - \mu_x)^2 \times P(x_i)} \\ &= \sqrt{(1 - 2.5)^2(0.25) + \dots + (4 - 0.25)^2} = 1.12\end{aligned}$$

Aşağıdaki gibi 2 adet örneklem alalım.

$\{1, 1\} \{2, 1\} \{3, 1\} \{4, 1\}$

$\{1, 2\} \{2, 2\} \{3, 2\} \{4, 2\}$

$\{1, 3\} \{2, 3\} \{3, 3\} \{4, 3\}$

$\{1, 4\} \{2, 4\} \{3, 4\} \{4, 4\}$

Bu örnekleme tanımlayan yeni bir rassal değişken (iki sayının ortalaması) tanımlayalım ve olasılık dağılımını belirleyelim.

Yeni değişkenin ortalama değerleri aşağıdaki gibi olur. Şimdi yeni değişken(örneklem) için olasılık dağılım tablosunu yazalım.

$$\frac{1+1}{2} = 1$$

$$\frac{2+1}{2} = 1.5$$

$$\frac{3+1}{2} = 2$$

$$\frac{4+1}{2} = 2.5$$

$$\frac{1+2}{2} = 1.5$$

$$\frac{2+2}{2} = 2$$

$$\frac{3+2}{2} = 2.5$$

$$\frac{4+2}{2} = 3$$

$$\frac{1+3}{2} = 2$$

$$\frac{2+3}{2} = 2.5$$

$$\frac{3+3}{2} = 3$$

$$\frac{4+3}{2} = 3.5$$

$$\frac{1+4}{2} = 2.5$$

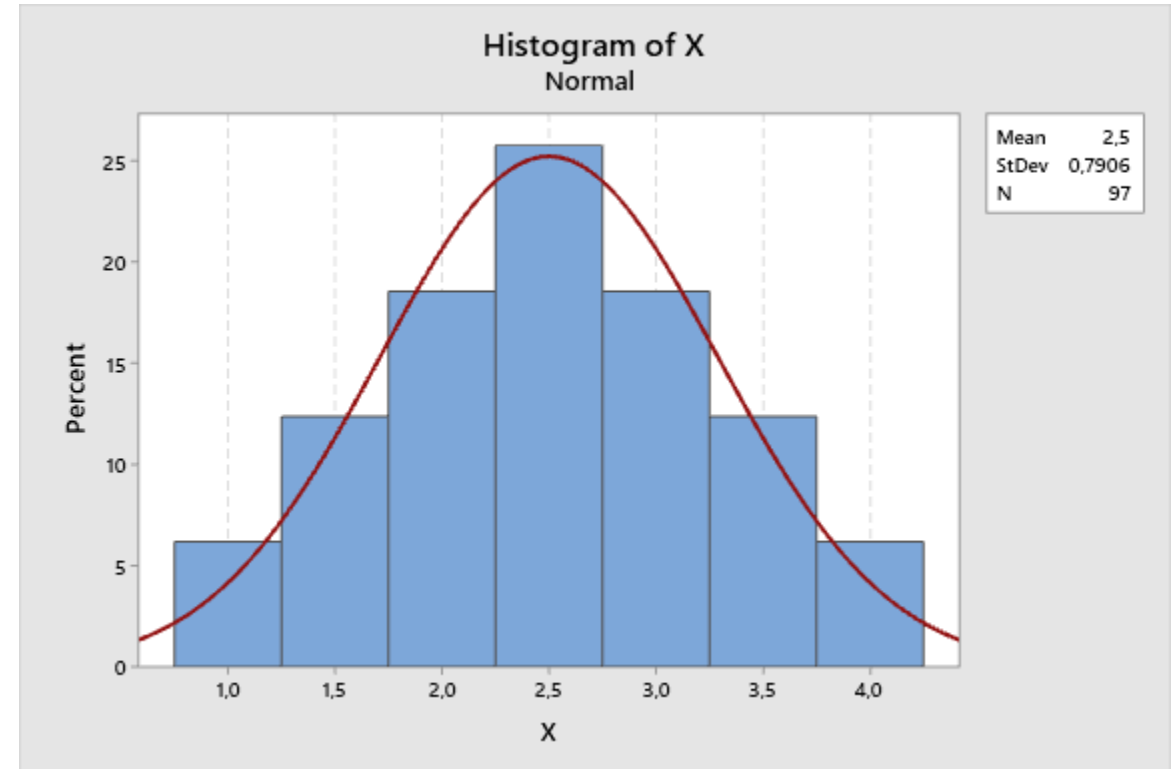
$$\frac{2+4}{2} = 3$$

$$\frac{3+4}{2} = 3.5$$

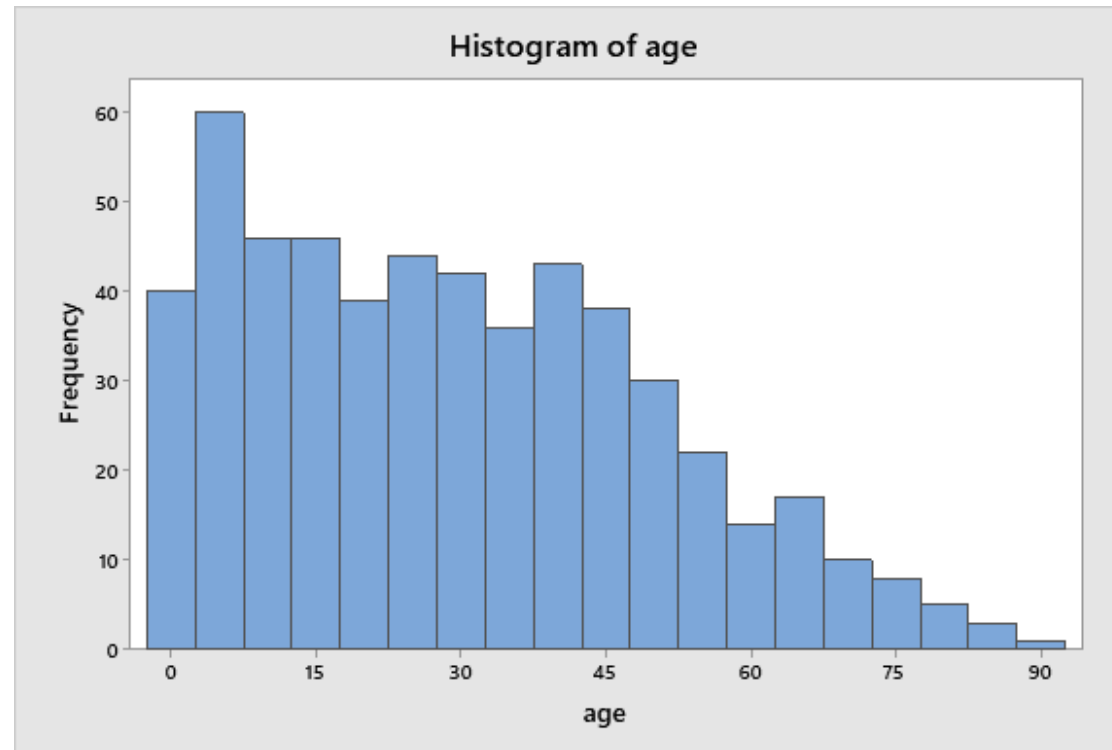
$$\frac{4+4}{2} = 4$$

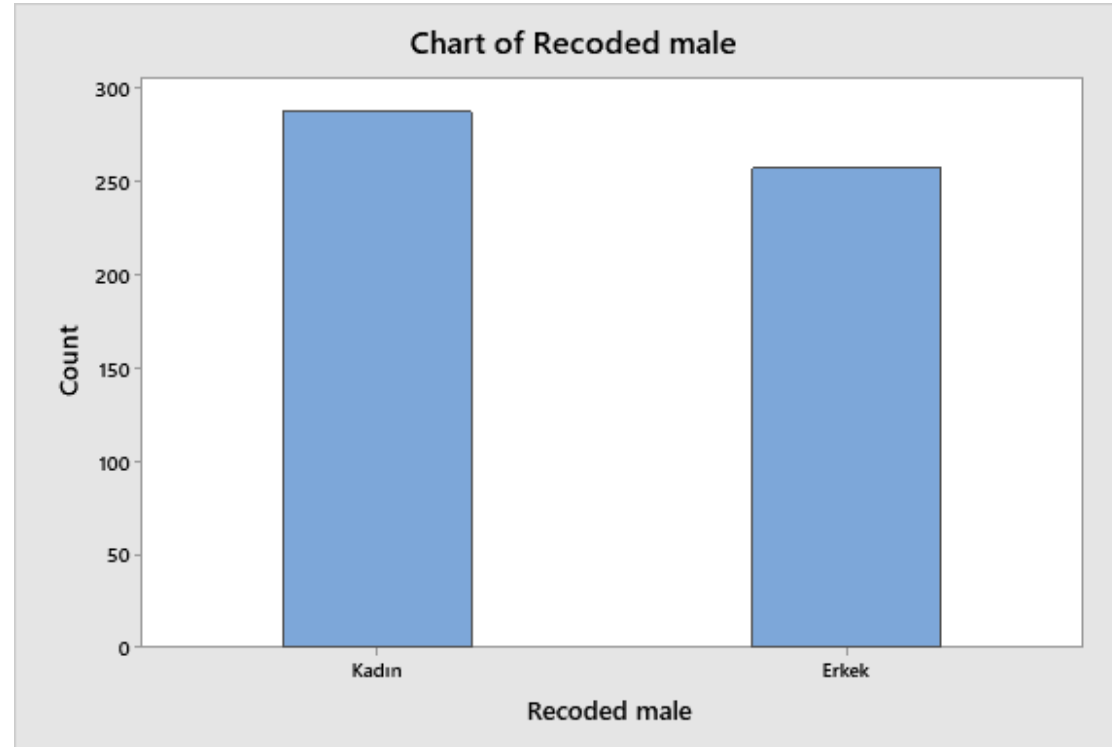
$\bar{x}$	$P(\bar{x})$
1	$\frac{1}{16} = 0.0625$
1.5	$\frac{2}{16} = 0.125$
2	$\frac{3}{16} = 0.1875$
2.5	$\frac{4}{16} = 0.250$
3	$\frac{3}{16} = 0.1875$
3.5	$\frac{2}{16} = 0.1250$
4	$\frac{1}{16} = 0.0625$

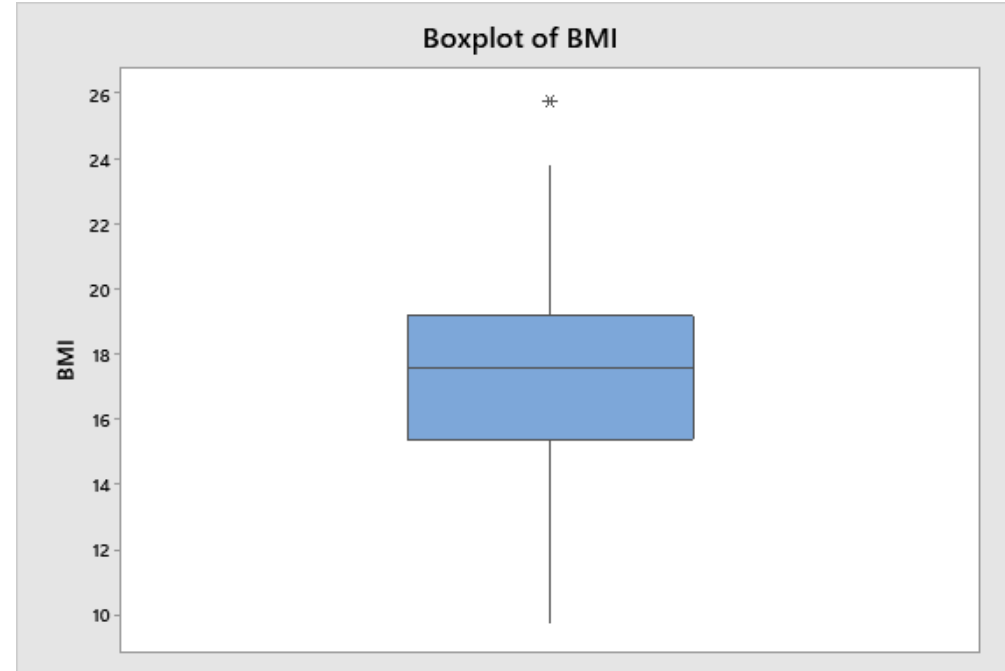
1. Bütün olasılıklar toplamı 1 olmalı
2. Olasılıklar 0 ile 1 arasında olmalı.



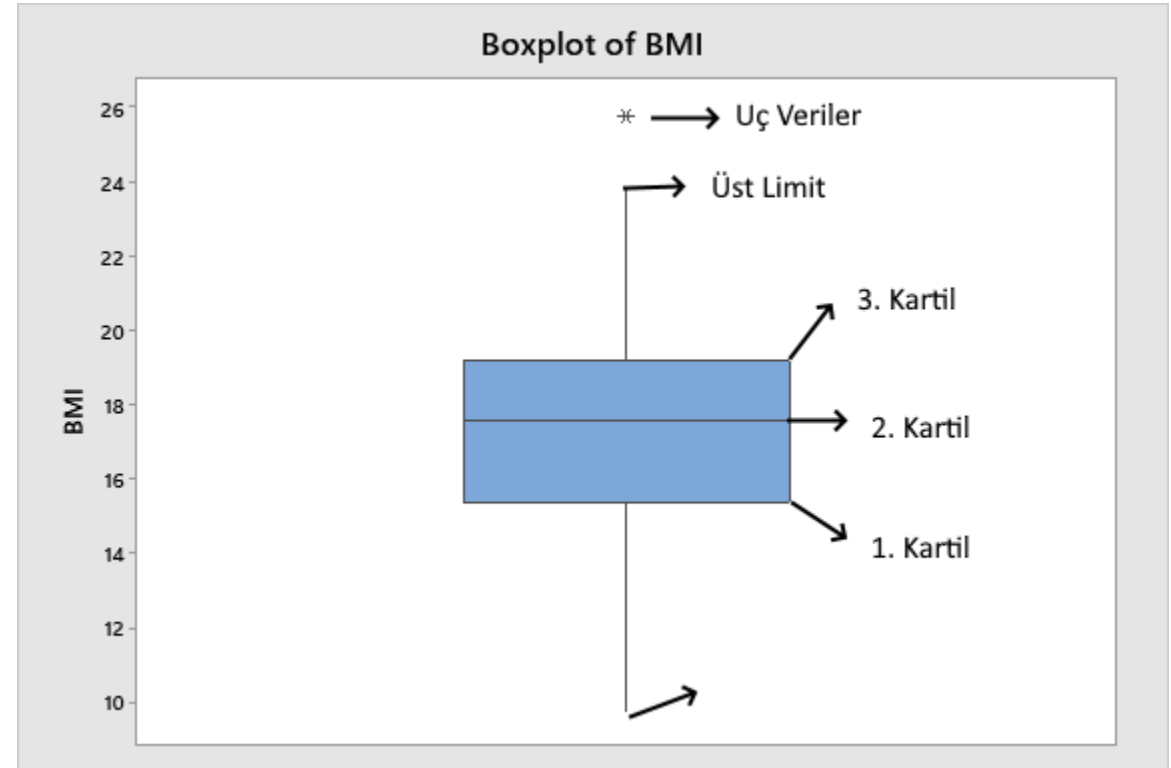
Değişkenlerin dağılımını gösterir.



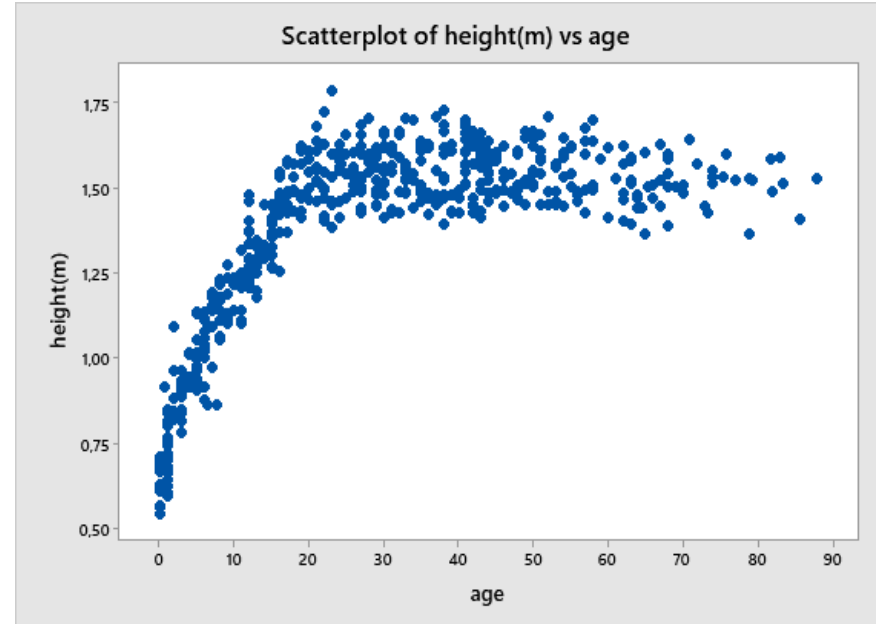


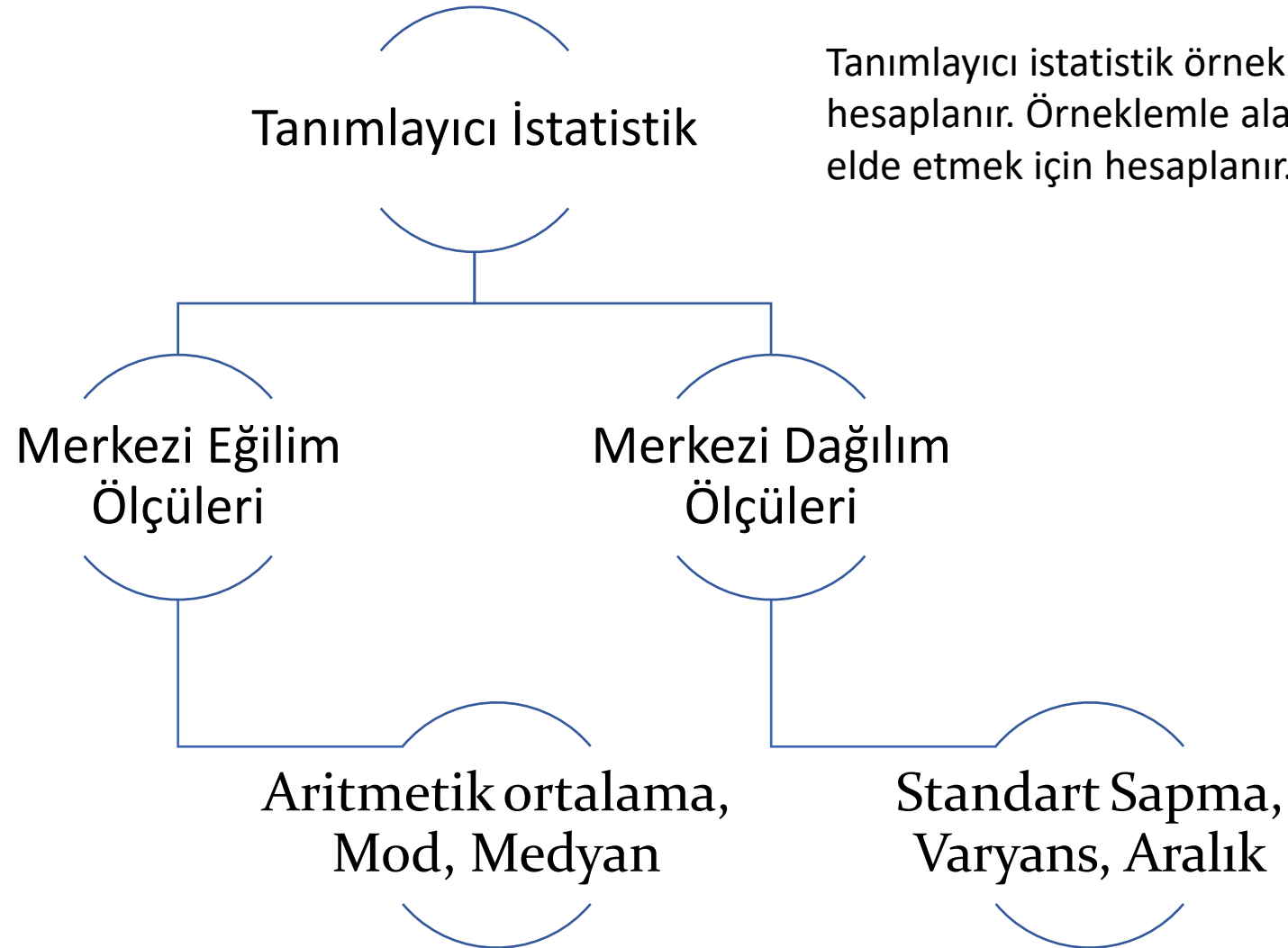


- Kartil Hesabı
- Q_1 Birinci Kartil (ilk yarıdaki verilerin medyan değeri)
- Q_3 Üçüncü Kartil (ikinci yarıdaki verilerin medyan değeri)
- IQR(Orta yarıdaki değerlerin aralığı)
- $IQR = Q_3 - Q_1$



Bir veri için dağılımı göstermek için kullanılır. x ve y koordinatları üzerindeki noktalardan oluşur.





Tanımlayıcı istatistik örneklemden yola çıkılacak hesaplanır. Örnekleme alakalı işe yarar bilgiler elde etmek için hesaplanır.

Örneklemdaki orta nokta ile ilgilenir.

Örneklemdaki bir noktaya ya da değere karşılık gelir. Örneğin aritmetik ortalama aşağıdaki formülde olduğu gibi hesaplanır.

$$\bar{x} = \frac{\sum_i^n x_i}{n} \quad n \text{ gözlem sayısı, } x_1, x_2, \dots, x_n \text{ bireysel gözlemler}$$

Bir dizi ölçümün ya da gözlem sonucunun aritmetik ortalaması, dizideki ölçümlerin toplanıp toplamın ölçüm sayısına bölünmesiyle elde edilir (Arııcı, 2006).

Hesaplanmasında veri setindeki tüm ölçümler kullanılır.

En az eşit aralık düzeyindeki (sürekli) veriler için uygundur.

Dağılımdaki tüm puanlar dikkate alınarak hesaplandığı ve ileri matematiksel işlemler için de uygun olduğu için en kararlı ve tutarlı merkezi eğilim ölçüsüdür.

Puan dağılımında aşırı uç değerlerin olması durumunda ya da dağılımın çarpık olması yanıltıcı sonuçlara neden olabilir. Bu durumlarda ortancanın kullanılması daha uygundur.

Örneğin 5 öğrenciye ait not ortalamaları 2.56, 3.21, 3.56, 2.10, 1.87 olursa

ortalaması $\bar{x} = \frac{\sum_i^n x_i}{n} = \frac{13.30}{5} = 2.66$ olarak bulunur.

Puan	Frekans	Puan x Frekans
2	1	$2 \times 1 = 2$
3	4	$3 \times 4 = 12$
4	8	$4 \times 8 = 32$
5	5	$5 \times 5 = 25$
Toplam	18	$\bar{x} = \frac{71}{18} = 3.94$

Aşağıdaki tabloda öğrencilerin yaratıcılık testi puanlarına ilişkin ölçümler ve frekansları verilmiştir. Öğrencilerin ortalama puanlarını hesaplayalım

Ölçüm	f
85	5
70	4
65	6
50	4
45	3

Veriler sınıflandırılmış ise aritmetik ortalama;

Doğru Sayısı	Frekans	Puan Sınıfı Ara Değeri	Frekans x Ara değer
1-5	2	$(1+5) / 2 = 3$	$2 \times 3 = 6$
6-10	3	$(6+10) / 2 = 8$	$3 \times 8 = 24$
11-15	11	$(11+15) / 2 = 13$	$11 \times 13 = 143$
16-20	6	$(16+20) / 2 = 18$	$6 \times 18 = 108$
21-25	10	$(21+25) / 2 = 23$	$10 \times 23 = 230$
26-30	4	$(26+30) / 2 = 28$	$4 \times 28 = 112$
31-35	14	$(31+35) / 2 = 33$	$14 \times 33 = 462$
Toplam	50		

Küçükten büyüğe dizilmiş ölçümlerin orta puanını gösterir. Üst yarısını alt yarısından ayıran noktadır.

Dağılımın normalden uzak ya da çarpık olduğu durumlarda kullanılır.

Küçük örneklerde;

n tek ise $(n+1)/2$. değere karşılık gelir.

n çift ise $n/2$. ve $(n/2)+1$. sıradaki değerlerin ortalamasına karşılık gelir.

Ortanca (Medyan)

Değerler sıraya konulduğunda tam ortada yer alan değerdir.

5, 2, 3, 5, 3, 3, 4, 5, 4, 4, 5, 5, 2, 4, 5, 5, 4, 5, 3

Sıraladığımızda

2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5,

Ortanca değer: $\frac{n+1}{2} = \frac{19+1}{2} = 10. \text{değer}$ Ortanca: 4

n sayısı çift olduğunda tam ortada bir değer yoktur.

Örneğin: 2, 2, 3, 4, 5, 5, 6, 7, 9, 11

$$\frac{10 + 1}{2} = 5.5$$

Olduğundan 5. ve 6. değer aritmetik ortalaması alınır. $(5+5)/2 = 5$

Birikimli Frekanstan Yararlanarak Ortanca Hesaplama

Puan	Frekans	Birikimli Frekans
2	2	2
3	1	3
4	3	6
5	2	8
Toplam	8	

4. ve 5.
değer bu
satırda

$$\frac{8+1}{2} = 4.5$$

Ortanca= 4

Bir değişkenle ilgili bir dizi ölçümden en çok tekrarlanan ölçümdür.

Dağılımın normal olmadığı durumlarda birden fazla mod olabilir.

Bu durumda dağılım çok modludur.

Doğru Sayısı	Frekans
1-5	2
6-10	4
11-15	6
16-20	7
21-25	6
26-30	2
Toplam	25

Tepe Değeri:

- Frekansı en yüksek değer,
- Sınıflandırılmış verilerde frekansı en yüksek sınıfın ara değeri.

$$\text{Tepe Değeri (Mod)} = (16 + 20) / 2 = 18$$

Dağılımda en çok tekrar eden değerdir. Öğrencilerin notları sıralandığında:

2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5

En çok tekrar eden değer:4

Tepe Değeri (Mod)= 4



Betimsel İstatistikler (Merkezi Dağılım Ölçüleri)

Bir dağılımda veriler arasındaki değişimden kaynaklanan farklılıkların istatistiksel ölçüleridir.

Dağılımda bir aralığın değerini verir.

Bir merkezi eğilim ölçüsünden olan gösterir.

Ranj, Standart sapma, Varyans, Çeyrek sapma, Değişim katsayısı, Çarpıklık katsayısı, Basıklık katsayısı

Ölçümlerin en büyüğü ile en küçüğü arasındaki farktır.

Ortalamalar eşit ve kişi sayıları aynı olan iki grubun kabaca karşılaştırılmasında kullanılır.

5, 2, 4, 2, 4, 7, 2

En küçük değer=2, En büyük değer=7,

Ranj = en büyük - en küçük **Ranj** = $7 - 2 = 5$

Kartiller Arası Mesafe (Interquartile Range)(IQR)

Veriler 4'e bölünerek kartillere (Q_1, Q_2, Q_3, Q_4) ayrılır. 3. Q_3 ve Q_1 1. kartil arası mesafeye kartiller arası mesafe (IQR) denir. Aşağıdaki verilerin kartillerini hesaplayalım.

Sıra	1	2	3	4	5	6	7	8	9	10
Veri	2	4	5	7	9	13	15	17	28	30

Yukarıdaki tabloyu önce 4 eşit parçaya ayırırız. Veri sayısı 10 yani çift bir sayı olduğundan $\frac{1}{4}(10+1)=2.75$. Ancak 2.75 tam sayı olmadığı için aşağıdaki şekildeki gibi bir enterpolasyon yapılır.

$$Q_1 = x_2 + 0.75(x_3 - x_2) = 4 + 0.75(5 - 4) = 4.75 \text{ olarak bulunur.}$$

Kartiller Arası Mesafe (IQR)

Aynı şekilde Q_3 değeri de $\frac{3}{4}(10+1)$ 'den 8.25 olarak bulunur. Aşağıdaki enterpolasyon ile

$$Q_3 = x_8 + 0.25(x_9 - x_8) = 17 + 0.25(28 - 17) = 19.75$$

Dolayısıyla IQR (Kartiller arası mesafe) şu şekilde hesaplanır.

$$IQR = Q_3 - Q_1 = 19.75 - 4.75 = 15$$

Kartiller Arası Mesafe de sadece bazı verileri kullandığı için tam dağılım ya da sapma ile ilgili bilgi vermez. Tüm verilerin hesaba katıldığı sapma ölçüsü için standart sapma hesaplaması yapılır.

Dağılımdaki ölçümlerin tümünü işleme katan, en güvenilir ve kararlı değişim ölçüsüdür. Dağılımdaki değerlerin dağılımın orta noktasına olan uzaklıklarının standart bir ölçüsüdür. Birimi yoktur. Varyans aşağıdaki formüldeki gibi hesaplanır.

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

Standart sapma ise varyans'ın kareköküdür.

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

Varyans ve Standart Sapma Hesabı

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
$x_1 = 2.56$	$2.56 - 2.66 = -0.10$	0.0100
$x_2 = 3.21$	$3.21 - 2.66 = 0.55$	0.3025
$x_3 = 3.56$	$3.56 - 2.66 = 0.90$	0.81
$x_4 = 2.1$	$2.00 - 2.66 = -0.56$	0.3136
$x_5 = 1.87$	$1.87 - 2.66 = -0.79$	0.6240
		$\sum (x_i - \bar{x})^2 = 2.0602$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{2.0602}{5 - 1} = 0.5151$$

Standart sapma ise;

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{0.5151} = 0.7177$$

Eğer popülasyonun sapması hesaplanacaksa

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \text{ Varyans}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \text{ Standart Sapma}$$

Olarak hesaplanır.

Grup aralığının orta noktası ve grup frekanslarınıı temele alan formüller vardır.

$$s = \sqrt{\frac{\sum f(x_0 - \bar{x})^2}{n - 1}}$$

Gruplanmış Veriler İçin Standart Sapma Hesaplanması

Puan aralığı	f	x_0	$f(x_0 - \bar{x})^2$
0-4	5	2	125
5-9	3	7	0
10-14	2	12	50
15-19	1	17	100
Toplam	11		275

Ortalama 7'ye
yuvarlanmıştır.

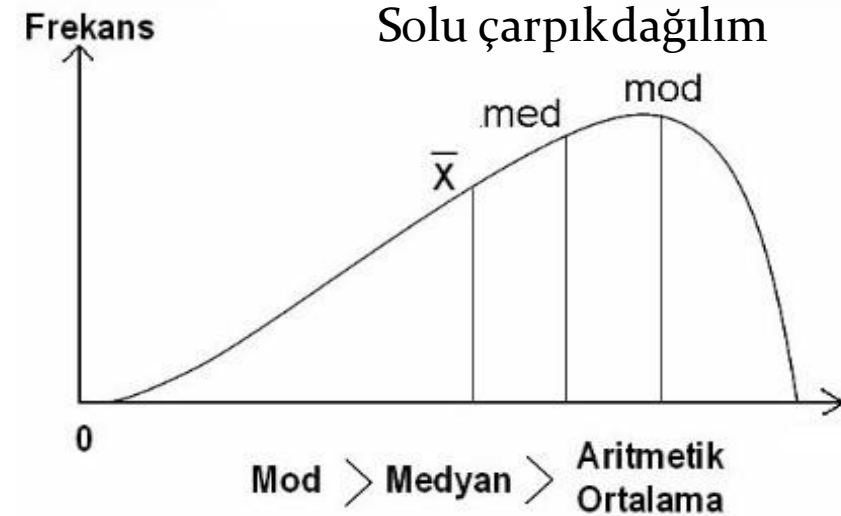
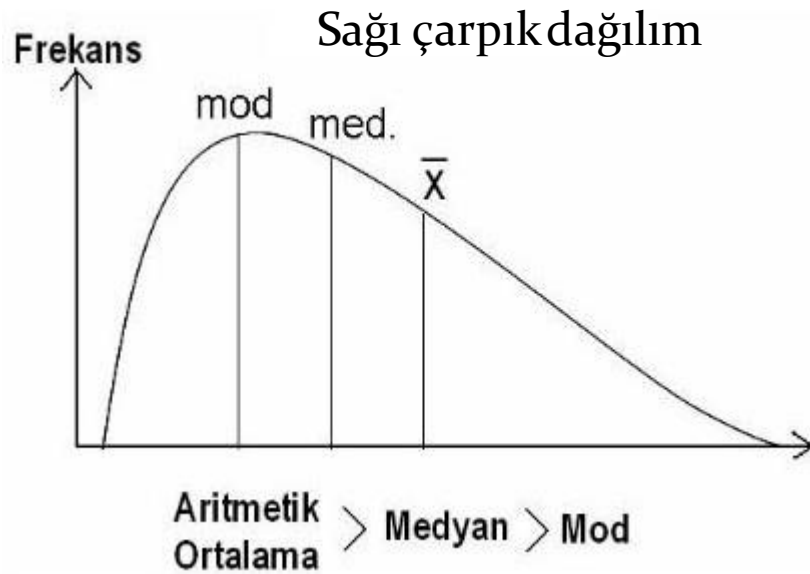
$$s = \sqrt{\frac{\sum f(x_0 - \bar{x})^2}{n - 1}}$$
$$= \sqrt{\frac{275}{10}} = 5.25$$

Çarpıklık Katsayısı: $= \frac{3(\bar{x} - \text{medyan})}{s}$

Bir dağılımda ortalama ve ortanca farklı noktalardaysa dağılım çarpıktır.

Çarpıklık katsayısı sıfırdan küçük ise dağılım sola (negatif) çarpık

Çarpıklık katsayısı sıfırdan büyük ise dağılım sağa (pozitif) çarpık



Basıklık Katsayısı

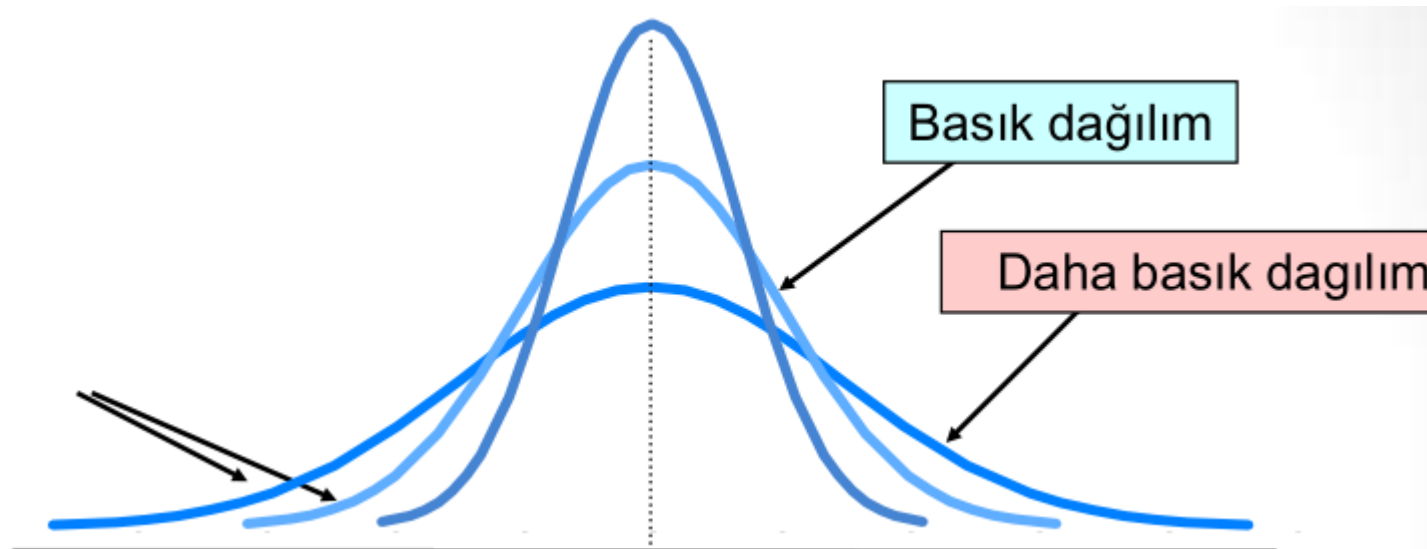
Dağılımın genişliğini yorumlamayı sağlar.

Basıklık katsayısı, sıfırdan küçük ise dağılım basık

Basıklık katsayısı sıfırdan büyük ise dağılım sivri

Basıklık katsayısı sıfıra eşit ise dağılım normal

$$\mathbf{BK} = \frac{\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})^4}{\mathbf{S}^4}$$



Cintas, Pedro Grima, Lluís Marco-Almagro, ve Javier Tort-Martorell Llabres. *Industrial statistics with Minitab*. Wiley Online Library, 2012.

Erhardt, Erik B., Edward J. Bedrick, ve Ronald M. Schrader. “Advanced Data Analysis-Lecture Notes”, 2016.

Karagöz, Murat. *İstatistik Yöntemleri*. 9. bs. Ekin Kitabevi Yayınları, 2015.

Khan, Rehman M. *Problem solving and data analysis using minitab: A clear and easy guide to six sigma methodology*. John Wiley & Sons, 2013.

Lesik, Sally A. *Applied statistical inference with MINITAB®*. CRC Press, 2018.

Newton, Isaac. *Minitab cookbook*. Packt Publishing Ltd, 2014.

“Support | Minitab”. Erişim 29 Mart 2020. <https://www.minitab.com/en-us/support/>.