



HACETTEPE  
ÜNİVERSİTESİ  
İSTATİSTİK BÖLÜMÜ

## İST155 İSTATİSTİĞE GİRİŞ I

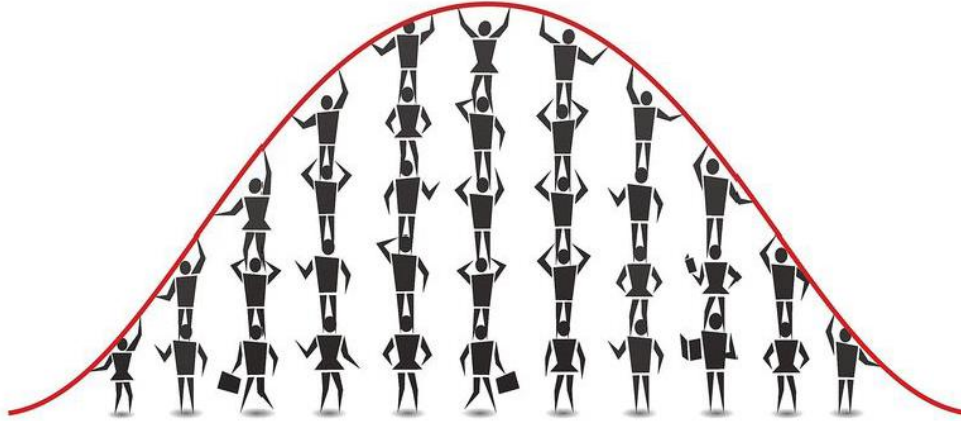
### DERS 9

### Normal Dağılım ve Simetri- Asimetri Ölçüleri

**Ders sorumluları: Prof.Dr.Serpil AKTAŞ ALTUNAY (01 Şubesi)  
Doç.Dr. Ayten YİĞİTER (02 Şubesi)**

# 1. NORMAL DAĞILIM VE STANDART NORMAL DAĞILIM (NORMAL DISTRIBUTION AND STANDARD NORMAL DISTRIBUTION)

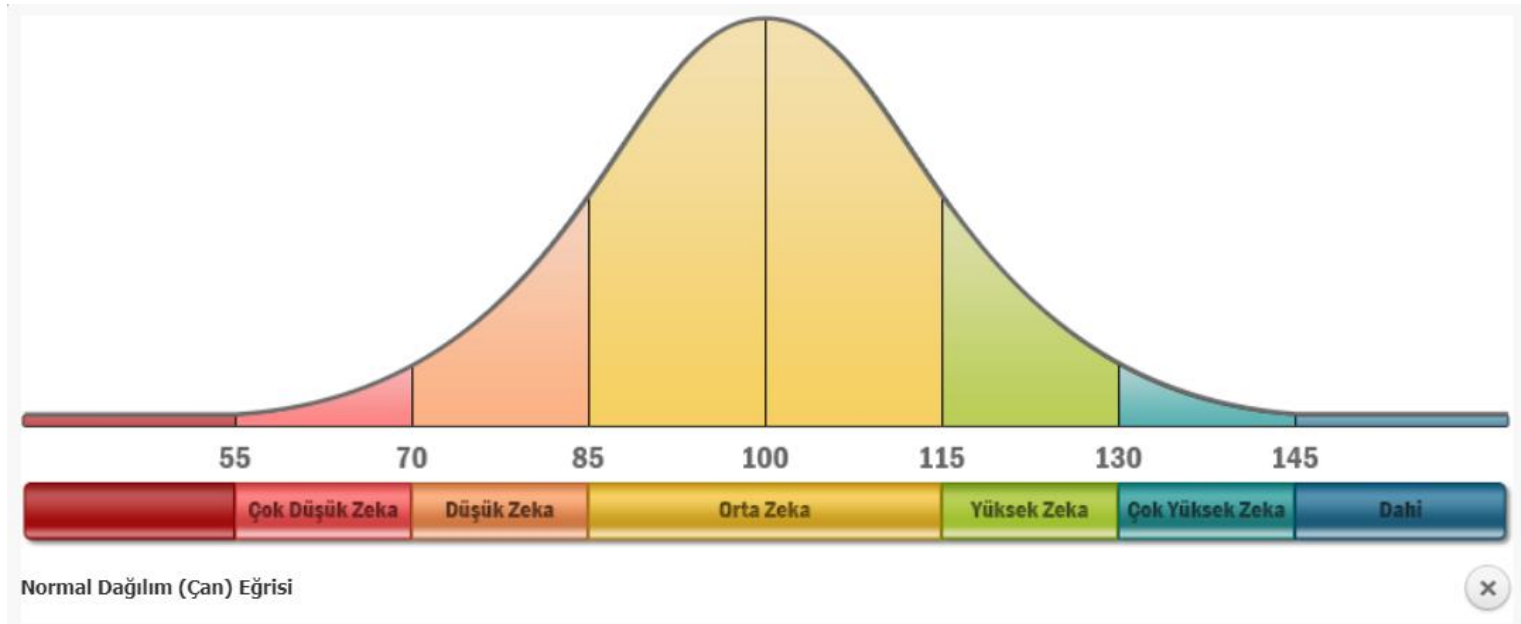
Normal dağılım ilk olarak 1733'te Fransız matematikçi Abraham de Moivre tarafından bulunmuş; daha sonra Pierre Laplace ve Karl Gauss tarafından çalışılmıştır. Eğrinin biçimi çana benzediği için «ÇAN EĞRİSİ» olarak da adlandırılır.



<https://www.tezyardimplatformu.com>

Bireylerin fiziksel  
özellikleri:

- Boy
- Kilo
- Zeka
- Sınav notları
- Ölçüm hataları



Şekil 1: IQ puanlarına ait Normal Dağılım eğrisi

Bu dağılıma normal dağılım denir. Normal dağılım  $N(\mu, \sigma^2)$  ile gösterilir.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \begin{array}{l} -\infty < x < \infty \\ x \in \mathbb{R}_x \end{array}$$

Burada  $\mu$ , kitlenin ortalamasını;  $\sigma^2$  ise varyansını gösterir.

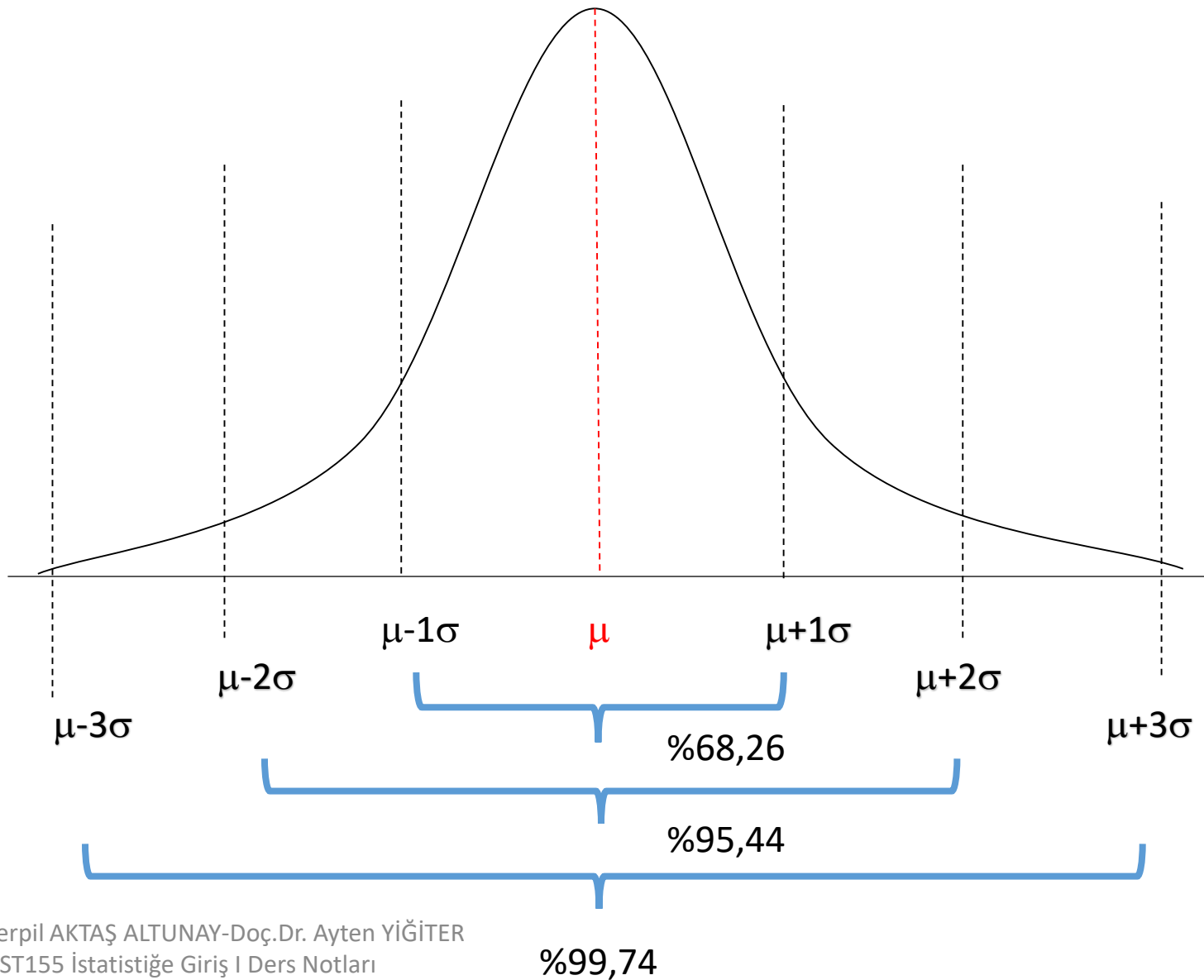
Normal dağılımda,

Verilerin

%68,26'sı	$\mu \pm 1\sigma$ sınırları içinde,
%95,44'ü	$\mu \pm 2\sigma$ sınırları içinde,
%99,74'ü	$\mu \pm 3\sigma$ sınırları içinde,

yer alır.

# Normal Dağılım

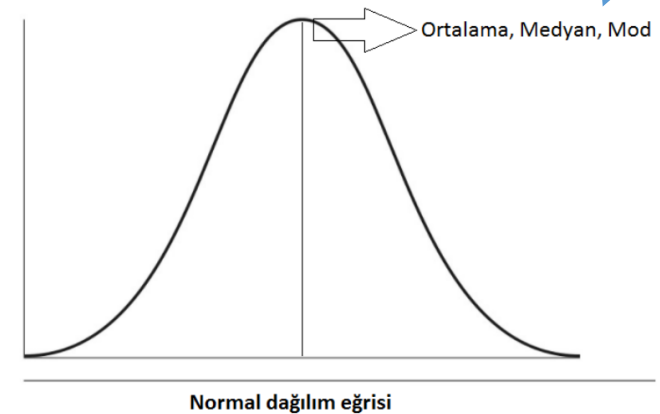


# Normal Dağılımın Özellikleri

Ortalamaya göre simetriktir

Ortalama=Medyan=Mod

Tek tepelidir



$$n \rightarrow \infty \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Bu özellik istatistikte «Merkezi Limit Teoremi» dir.

Normal dağılımlı kitleden çekilen bir örneklem olsun. Bu örnekte yer alan her değer için,

$$Z_i = \frac{(X_i - \mu)}{\sigma}$$

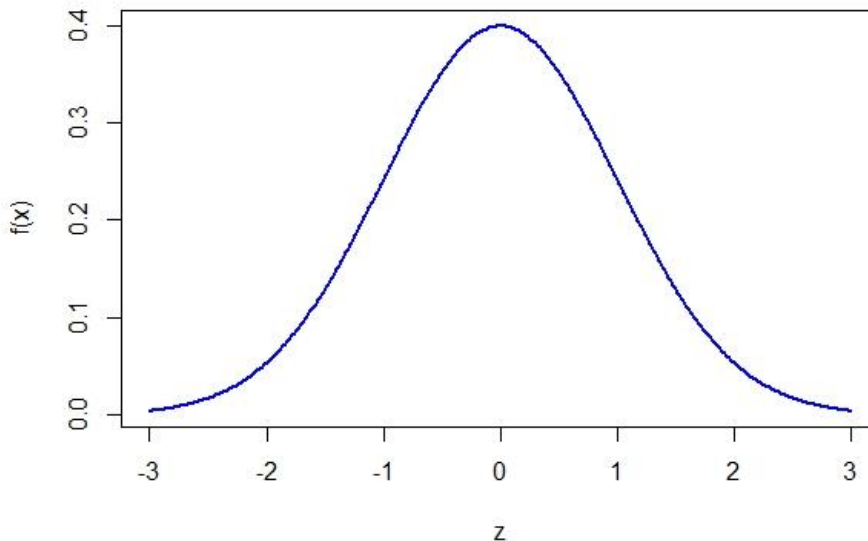
← Kitleye göre standartlaştırma

$$Z_i = \frac{(X_i - \bar{X})}{S}$$

← Örnekleme göre standartlaştırma

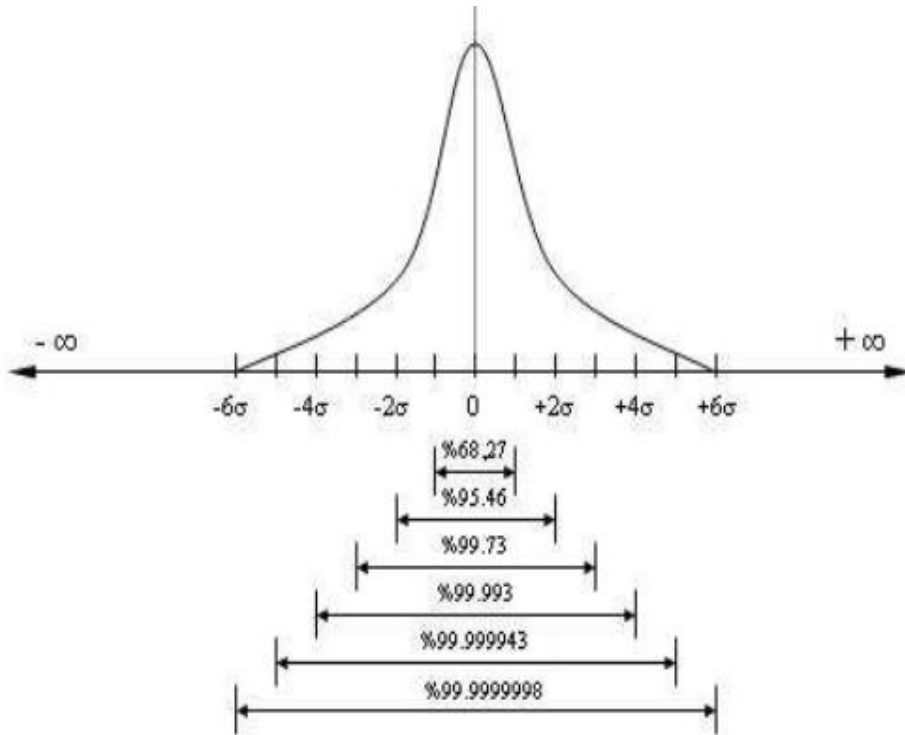
dönüşümü yapıldığında, ortalaması 0 ve varyansı 1 olan bir dağılım elde edilir. Bu dağılıma standart normal dağılım denir ve  $N(0, 1)$  ile gösterilir.

Standart Normal Dağılım



$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad \begin{matrix} z \in \mathbb{R} \\ -\infty < z < \infty \end{matrix}$$

# Standart normal dağılımın $\pm 6\sigma$ kadar olan dağılımı



%68,27'si	1 $\sigma$ sınırları içinde	
%95,46'sı	2 $\sigma$ sınırları içinde	
%99,73'ü	3 $\sigma$ sınırları içinde	
%99,993'ü	4 $\sigma$ sınırları içinde	
%99,999943'ü	5 $\sigma$ sınırları içinde	
%99,9999998'ü	6 $\sigma$ sınırları içinde	

Örneklem ile çalıştığımız için  $\sigma$  yerine örneklem standart sapması  $S'$ yi kullanacağız.

<https://studylibr.com/doc/970580/normal-da%C4%9F%C4%B1l%C4%B1m-e%C4%9Frisi-alt%C4%B1daki-alan%C4%B1n-standart-sapmaya-b...>

## Kalite kavramında 6 Sigma (6 $\sigma$ ) ne anlama gelir?

6 Sigma bir kalite iyileştirme sürecidir.



<u>Sigma Seviyesi</u>	<u>Milyonda Olası Hata Adedi</u>
1 Sigma	697700
2 Sigma	308700
3 Sigma	66810
4 Sigma	6210
5 Sigma	233
6 Sigma	3,4

DÜNYA'DA VE TÜRKİYE'DE ALTI SİGMA (SIX SIGMA IN THE WORLD AND TURKİYE) Bugün Dünyada bu metodolojiyi kullanan firmaların başında Motorola, GE, Ford, Citibank, Quantum, Pirelli, Toshiba, Samsung, Ericsson, Hyundai, Sony, Kodak, Shell, Jaguar, Volvo, Fiat, Dupont, Xerox, LG, Siemens gelmektedir. Tahmin edilen sigma seviyeleri Amerikan firmaları için 3-4 sigma seviyesindedir.

Türkiye'de, TEI, Arçelik, DupontSa, Vitra, Kordsa, Profilo, Çimtaş, Kalekim, Ford, Borusan, Bos, Teba, Vestel, Fırat Plastik, Bosch, Çalık Tekstil, yeşim Tekstil Altı Sigmayı başarı ile uygulayan firmalar arasındadır. Tahmin edilen sigma seviyeleri Türk firmaları için 2,5-3,5 sigma seviyesindedir

<https://dergipark.org.tr/en/download/article-file/187044>



### QUALITATIVE STUDIES

Received: July 2008

Accepted: January 2009

Series : E

ISSN : 1308-724X

© 2009 [www.newwsa.com](http://www.newwsa.com)

Okşan Kansoy

Esra Dirgar

University of Ege

[oksan.kansoy@ege.edu.tr](mailto:oksan.kansoy@ege.edu.tr)

Elazığ-Türkiye

### **ALTI SİGMA NEDİR?**

4 Sigma seviyesinde:

Her saat 20.000 mektubun kaybolması

Her gün hemen hemen 15 dakika güvenli olmayan içme suyunun akması

Haftada 5.000 hatalı ameliyat yapılması

Her gün büyük havaalanlarına 2 hatalı inişin yapılması

Her yıl 200.000 hatalı reçetenin yazılması

Her ay hemen hemen 7 saat elektriğin kesilmesi.

6 Sigma seviyesinde ise:

Her saat 7 mektubun kaybolması

Her yedi ayda 1 dakika güvenli olmayan içme suyunun akması

Haftada 1,7 hatalı ameliyat yapılması

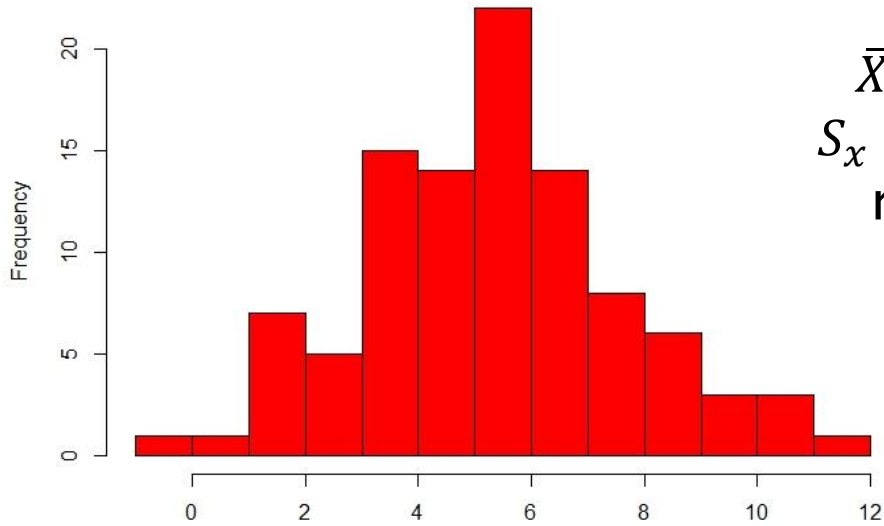
Her beş yılda bir büyük havaalanlarına 2 hatalı inişin yapılması

Her yıl 68 hatalı reçetenin yazılması

Her 34 yılda 1 saat elektriğin kesilmesi söz konusudur

<https://dergipark.org.tr/en/download/article-file/187044>

Histogram of x



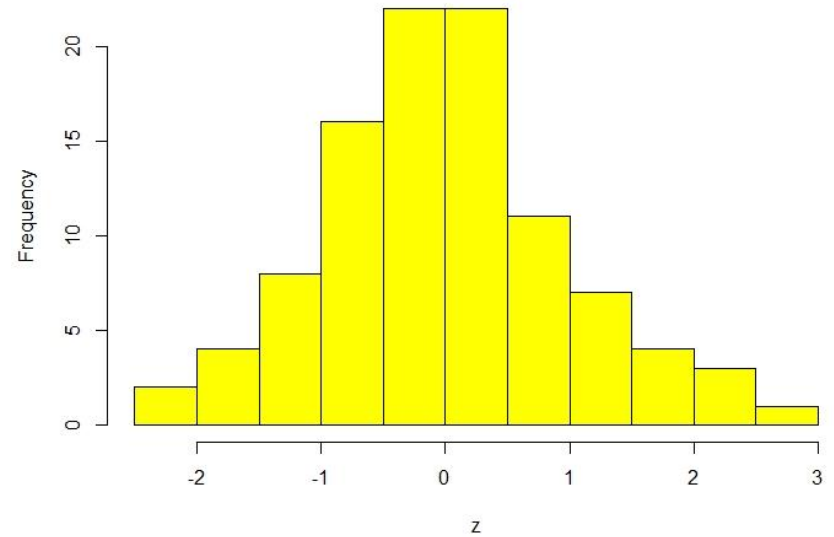
$$\bar{X} = 5,394$$

$$S_x = 2,27031$$

$$n=100$$

$x_i$	$z_i = \frac{x_i - \bar{X}}{S}$
3.7	-0.74615
8.8	1.500236
0.5	-2.15565
4.4	-0.43783
6.4	0.443111
5.1	-0.1295
4.5	-0.39378
4.1	-0.56997
5.9	0.222877
$\vdots$	$\vdots$

Histogram of z

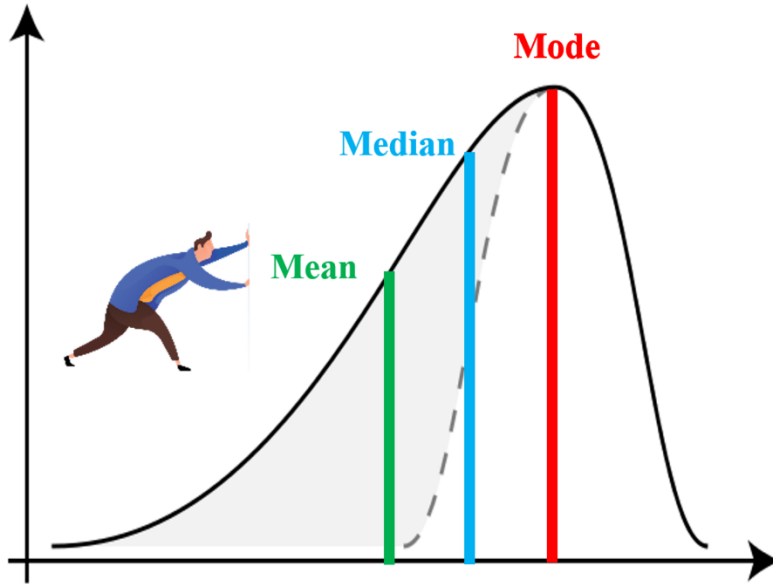


$$\bar{Z} \sim 0$$

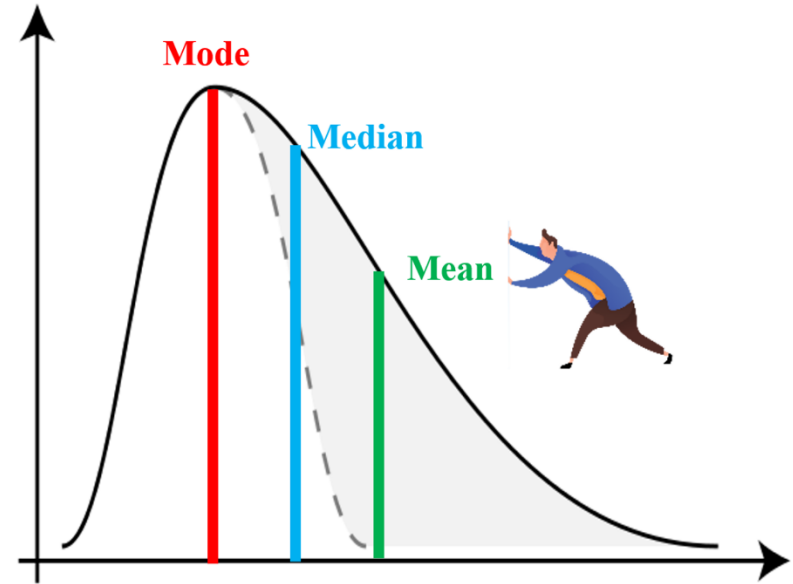
$$S_z = 1$$

## 2. Çarpıklık Katsayısı (Skewness)

Verinin dağılımı, normal/SND dağılımlı ile karşılaştırmak için, verinin simetriden ayrılışının bir ölçüsü olan çarpıklık (skewness) katsayısı (ÇK) kullanılır.



Sola Çarpık



Sağa Çarpık

Bu çarpıklığı bir katsayı hesaplayarak ifade edebilir miyiz ?

Ham verilerde çarpıklık katsayısı,

$$\zeta K = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{S^3} = \frac{\sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3}{n}$$

$$\zeta K = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3$$

Burada  $\bar{X}$ , aritmetik ortalama; S, standart sapmadır. Bu katsayıya göre,

- $\zeta K \sim 0$  ise, dağılım ortalamaya göre simetrik,
- $\zeta K < 0$  ise, dağılım sola doğru çarpık yani (–) yöne eğilimli,
- $\zeta K > 0$  ise, dağılım sağa doğru çarpık yani (+) yöne eğilimlidir.

Sıklık çizelgesi düzenlenmiş verilerde çarpıklık katsayısı,

$$\zeta K = \frac{\sum_{i=1}^k f_i (S_i - \bar{X})^3 / n}{S^3} = \frac{\sum_{i=1}^k f_i \left( \frac{S_i - \bar{X}}{S} \right)^3}{n}$$

$$\zeta K = \frac{n}{(n-1)(n-2)} \sum_{i=1}^k f_i \left( \frac{S_i - \bar{X}}{S} \right)^3$$

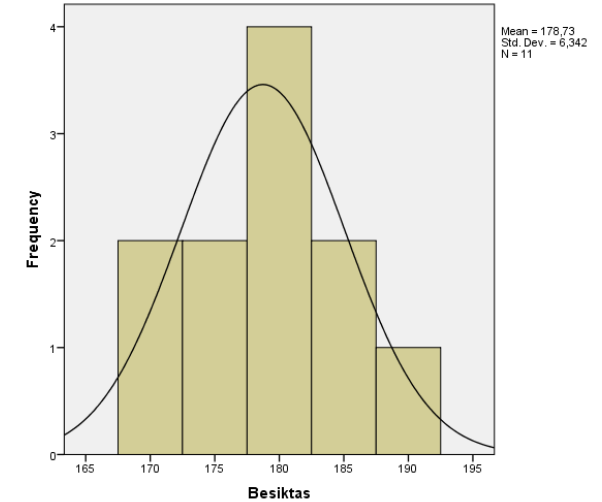
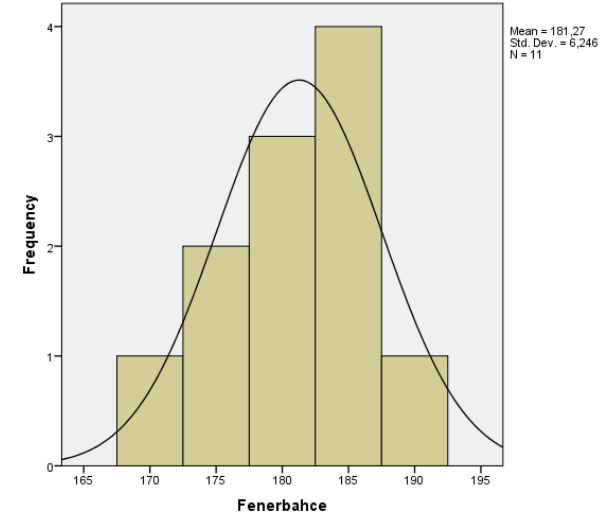
Burada  $\bar{X}$ , aritmetik ortalama; S, standart sapmadır. Bu katsayıya göre,

- $\zeta K \sim 0$  ise, dağılım ortalamaya göre simetrik,
- $\zeta K < 0$  ise, dağılım sola doğru çarpık yani (–) yöne eğilimli,
- $\zeta K > 0$  ise, dağılım sağa doğru çarpık yani (+) yöne eğilimlidir.

**Örnek 1:** Aşağıda takımlarda yer alan oyuncuların boy uzunlukları verilmiştir. Çarpıklık katsayısını her iki takım için de bulunuz ve yorumlayınız.

Fenerbahçe(X)	Beşiktaş(Y)
191	192
181	184
187	180
183	170
179	173
187	179
170	176
184	178
173	171
177	180
182	183
$\bar{X}=181,2727$	$\bar{Y}=178,72727$
$S=6,24645$	$S=6,34178$

$(X_i - \bar{X})^3$	$(Y_i - \bar{Y})^3$
920,39294	2338,19384
-0,0202855	146,590533
187,86401	2,06160781
5,1532682	-664,71525
-11,739294	-187,86401
187,86401	0,0202855
-1432,4748	-20,2855
20,2855	-0,3846732
-566,16905	-461,4012
-78,003757	2,06160781
0,3846732	78,0037566
<b>Toplam=-766,46281</b>	<b>Toplam=1232,28099</b>



$$\zeta K_{Fe} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{S^3} = \frac{-766,46281/11}{6,24645^3} = -0,2858898$$

$$\zeta K_{Be} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{S^3} = \frac{1232,28099/11}{6,34178^3} = 5,986704e-08$$

**Örnek 2:** A marka pilin kullanım süresi saat olarak verilmiştir. Çarpıklık katsayısını bulunuz.

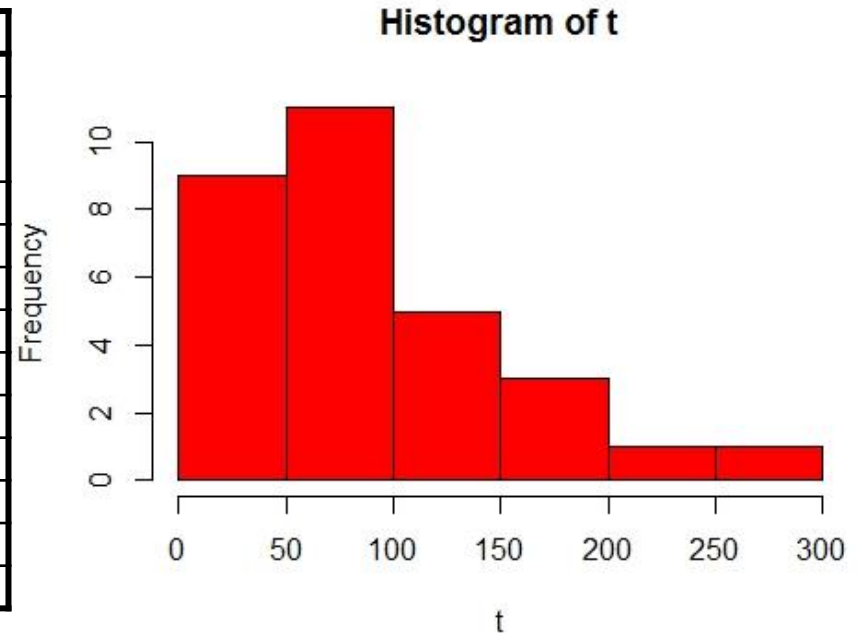
51,2	24,46	71,34
131,83	74,42	95,05
42,85	166,18	60,67
6,78	101,47	142,66
205,07	70,24	0,3
52,66	21,86	63,56
61,05	39,01	117,21
36,5	53,14	186,77
37,91	56,65	178,64
3,37	113,83	268,24

$\bar{X}$ =	84,49733333
S=	65,30520853
$\sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3$	29,33669163

$$\text{ÇK} = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{S^3} = \frac{\sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3}{n} = \frac{29,33669163}{30} \cong 0,97 > 0$$

**Örnek 2:** A marka pilin kullanım süresi saat olarak verilmiştir. Çarpıklık katsayısını bulunuz.

Descriptives		Statistic	Std. Error
omur	Mean	84,4973	11,92305
	95% Confidence Interval for Mean	60,1120	
		108,8827	
	5% Trimmed Mean	80,0800	
	Median	62,3050	
	Variance	4264,770	
	Std. Deviation	65,30521	
	Minimum	,30	
	Maximum	268,24	
	Range	267,94	
	Interquartile Range	82,13	
	Skewness	1,084	,427
	Kurtosis	,820	,833

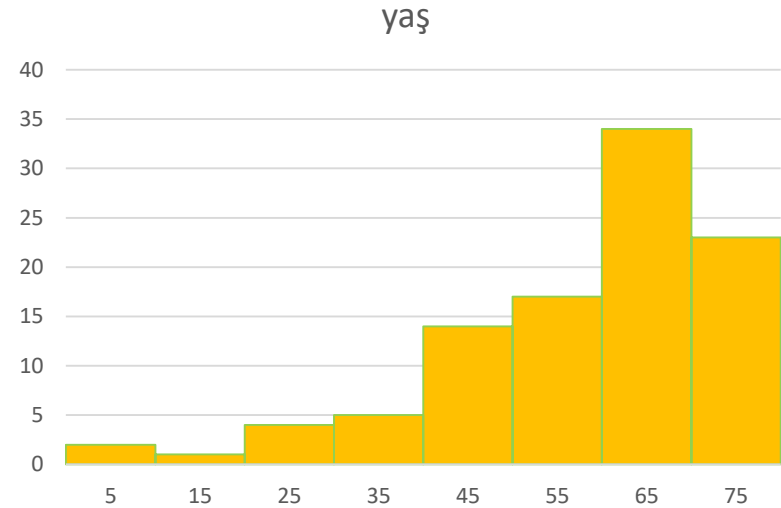


$$ÇK = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s} \right)^3 = \frac{30}{(29 \times 28)} \times 29,33669163 \cong 1,084 > 0$$

SPSS bu formülü kullanmaktadır.

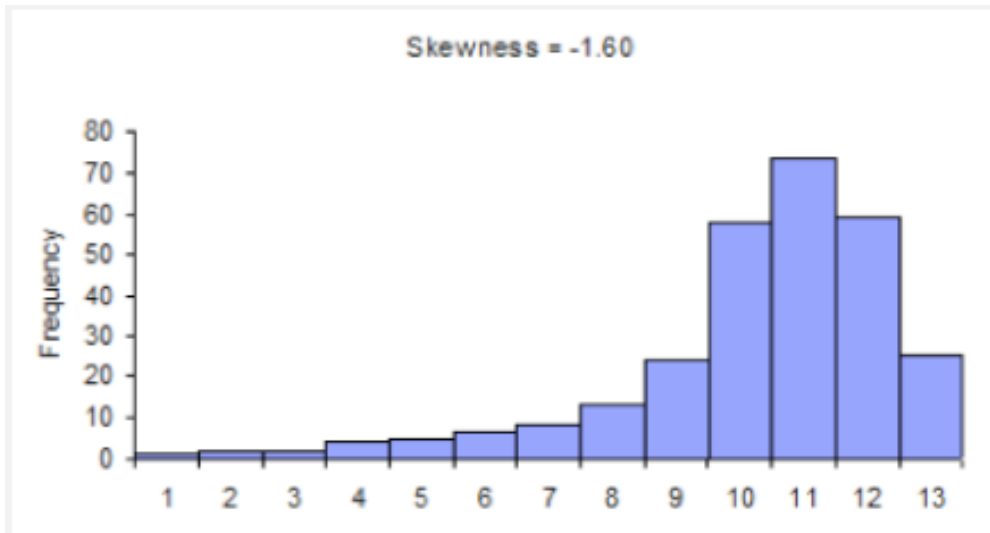
**Örnek 3:** Covid-19'dan en çok etkilenen kişilerin yaş dağılımı aşağıdaki sıklık çizelgesinde verilmiştir. Çarpıklık katsayısını hesaplayınız.

yaş	$f_i$	$f_i \times S_i$	$f_i (S_i - \bar{X})^3$
5	2	10	-297754
15	1	15	-79507
25	4	100	-143748
35	5	175	-60835
45	14	630	-30758
55	17	935	-459
65	34	2210	11662
75	23	1725	112999
toplam	100	5800	-488400

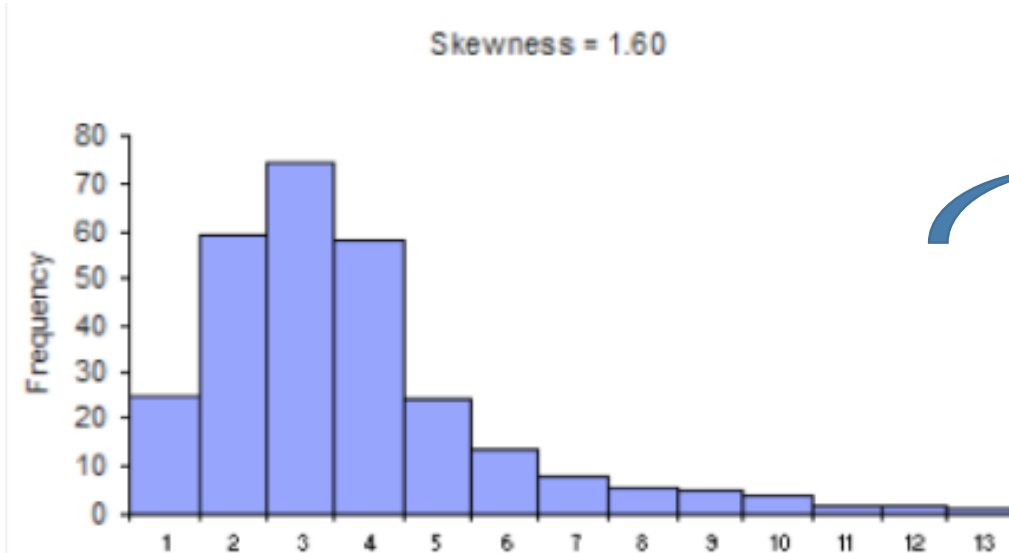


$$S^2 = 255,556 \quad S \cong 16$$

$$\zeta K = \frac{\sum_{i=1}^k f_i (S_i - \bar{X})^3 / n}{S^3} = \frac{-488400 / 100}{16^3} = -1,19 < 0$$



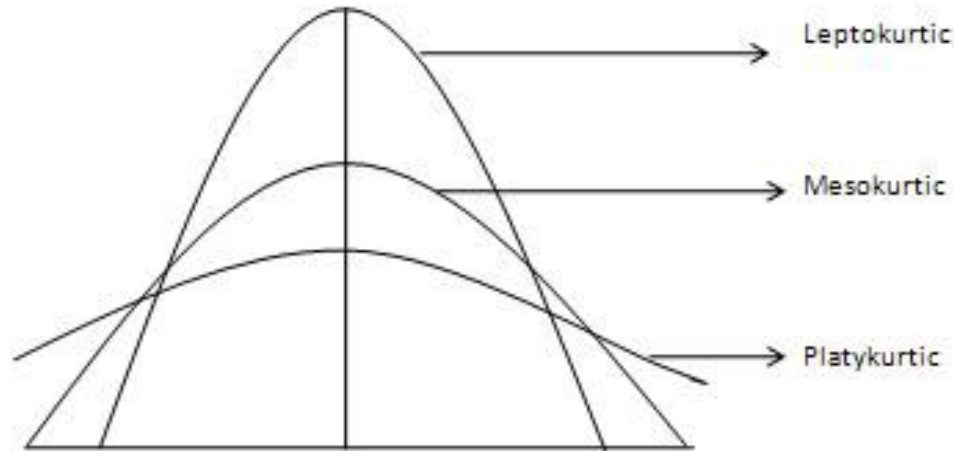
$\text{ÇK} = -1.60$   
Sola çarpık



$\text{ÇK} = 1.60$   
Sağa çarpık

### 3. Basıklık Katsayısı (Kurtosis)

Dağılım tek tepeli ise, dağılımın diklik ya da daha basıklık ölçüsü olan basıklık katsayısı (BK) hesaplanır.



**Mesokurtic** : Standart normal dağılımın Basıklık Katsayısı 3'e eşittir. Bu durumda verilerin dağılımı SND'ye uygun demektir.

**Lepokurtic** :  $BK > 3$  demektir. Dağılım SND'ye göre daha sivridir.

**Playkurtic** :  $BK < 3$  demektir. Dağılım SND'ye göre daha basıktır.

Bu basıklığı bir katsayı ile ifade edebilir miyiz ?

Ham verilerde basıklık Katsayısı,

$$BK = \frac{\sum_{i=1}^n (X_i - \bar{X})^4 / n}{S^4} - 3 = \frac{\sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4}{n} - 3$$

$$BK = \left[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Burada  $\bar{X}$ , aritmetik ortalama; S, standart sapmadır. Bu katsayıya göre,

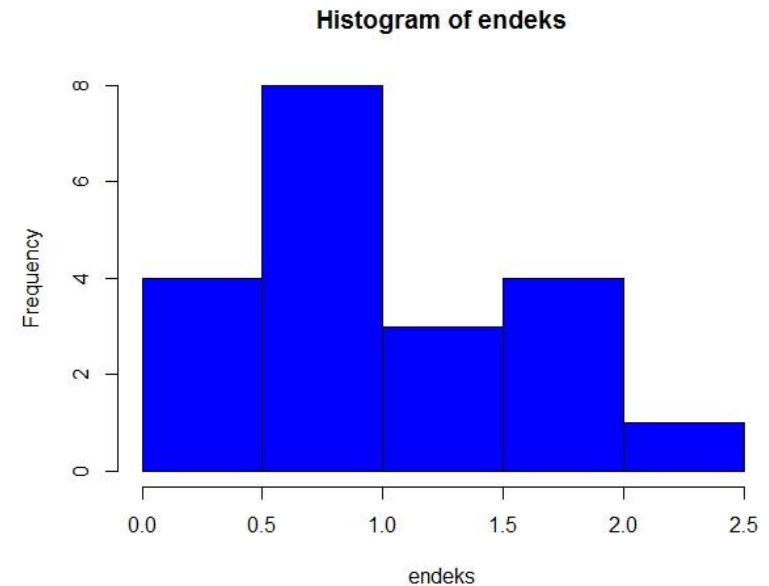
- $BK \sim 0$  ise, dağılımın yüksekliği standart normal dağılıma uygun,
- $BK < 0$  ise, dağılımın yüksekliği standart normal dağılıma göre daha basık,
- $BK > 0$  ise, dağılımın yüksekliği standart normal dağılıma göre daha diktir.

Not: Formüller 3'e göre düzeltildiği için karşılaştırmalar sıfır'a göre yapılmaktadır

**Örnek 4:** Belirli 20 ilin kitap okuma düzeylerine ilişkin endeks değerleri aşağıda verilmiştir. (en düşük=0 (kötü) en yüksek=2 (çok iyi)) Basıklık katsayısını hesaplayınız.

Endeks	1,03	0,82	0,37	1,16	0,96	0,37	0,61	2,0	1,99	0,77	$\bar{X} = 1$
	0,76	0,65	1,74	1,7	0,55	1,77	1,33	0,47	0,33	0,95	$S = 0,56$

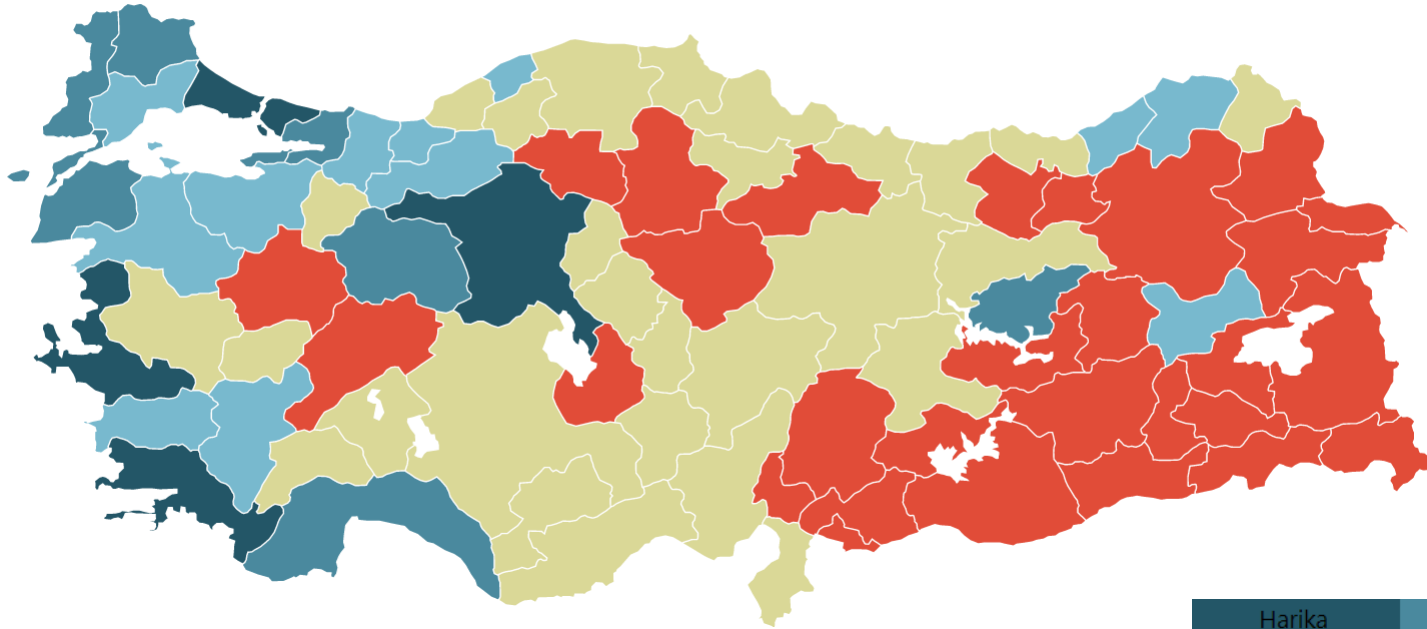
$$BK = \frac{\sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4}{n} - 3 = \frac{36,06}{20} - 3 = -1,197$$



$$BK = \left[ \frac{n(n+1)}{\underbrace{(n-1)(n-2)(n-3)}_{0,0722}} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4 \right] - \frac{3(n-1)^2}{\underbrace{(n-2)(n-3)}_{3,54}} = 0,0722 \times 36,06 - 3,54 = -0,94$$

Türkiye’de en çok hangi kitaplar okunuyor? İl il en çok okuyan şehirler hangileri?  
<https://www.idefix.com/turkiye-ne-okuyor>, 07.12.2020

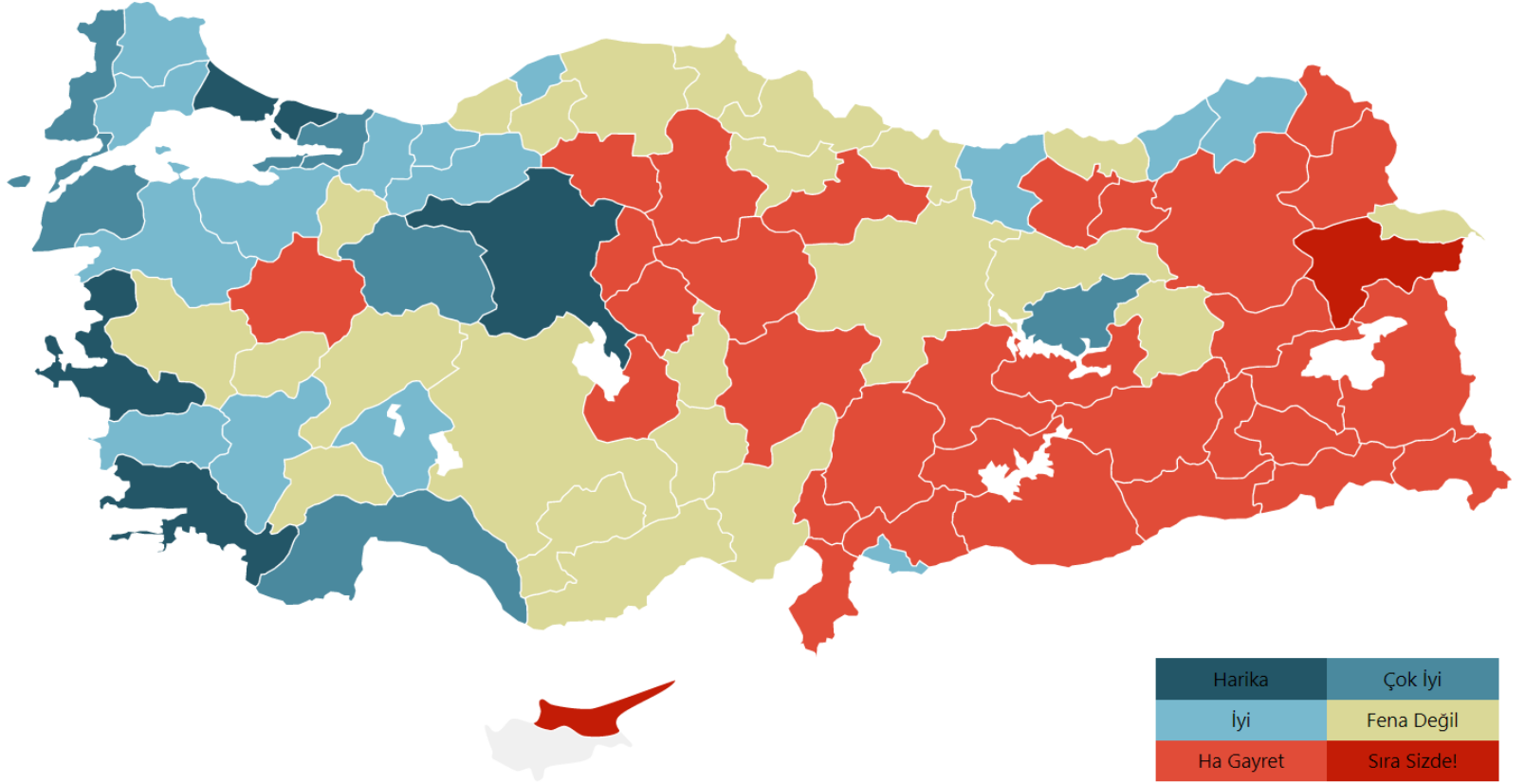
### Türkiye Ne Okuyor – Türkiye’de En Çok Okunan Kitaplar



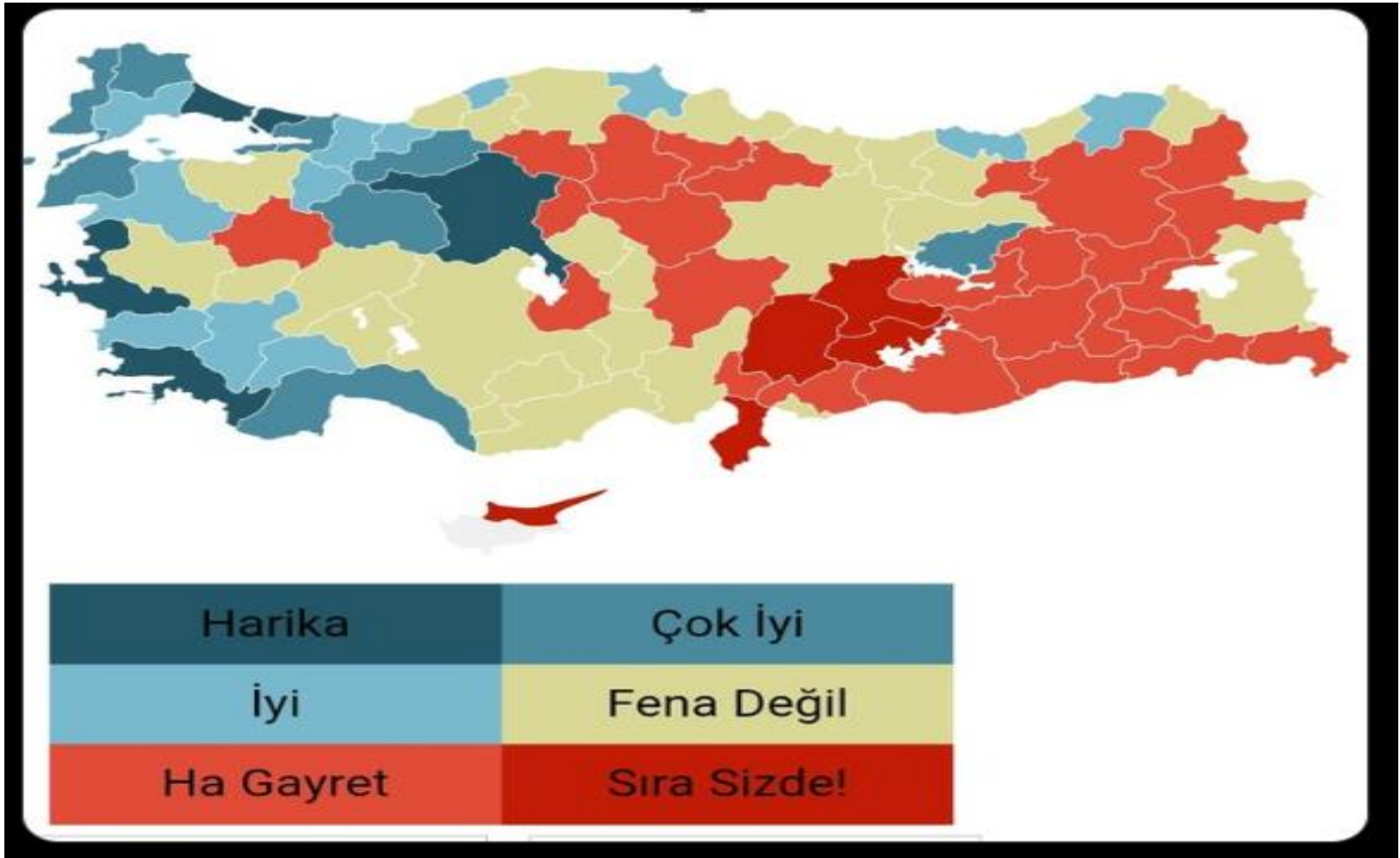
<https://www.idefix.com/turkiye-ne-okuyor>

Harika	Çok İyi
İyi	Fena Değil
Ha Gayret	Sıra Sizde!

## Türkiye Ne Okuyor – Türkiye’de En Çok Okunan Kitaplar



<https://www.idefix.com/turkiye-ne-okuyor>, 05.12.2022



<https://x.com/IsKitaplari/status/1661938383302676480>,

02.12.2024

Sıklık çizelgesi düzenlenmiş verilerde Basıklık katsayısı,

$$BK = \frac{\sum_{i=1}^k f_i (S_i - \bar{X})^4 / n}{S^4} - 3$$

$$BK = \left[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^k f_i \left( \frac{S_i - \bar{X}}{S} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Burada  $\bar{X}$ , aritmetik ortalama; S, standart sapmadır. Bu katsayıya göre,

- $BK \sim 0$  ise, dağılımın yüksekliği standart normal dağılıma uygun,
- $BK < 0$  ise, dağılımın yüksekliği standart normal dağılıma göre daha basık,
- $BK > 0$  ise, dağılımın yüksekliği standart normal dağılıma göre daha diktir.

**Örnek 5:** BEB dersinin ara sınavından alınan notlar verilmiştir. Basıklık katsayısını hesaplayalım.

Not	fi	Si
54-57	5	55,5
58-61	7	59,5
62-65	10	63,5
66-69	12	67,5
70-73	6	71,5
74-77	5	75,5
78-81	4	79,5
82-85	1	83,5

$$\bar{X} = \frac{\sum_{i=1}^k f_i S_i}{n} = \frac{3347}{50} = 66,94$$

$$S^2 = \frac{\sum_{i=1}^k f_i (S_i - \bar{X})^2}{n - 1} = \frac{2560,32}{50 - 1} = 51,206$$

$$S = \sqrt{51,206} = 7,1558$$

$$BK = \frac{\sum_{i=1}^k f_i (S_i - \bar{X})^4 / n}{S^4} - 3 = \frac{\frac{312677,37}{50}}{7,1558^4} - 3 = 2,39 - 3 = -0,61 < 0$$

## 4. Asimetri Ölçütleri

Bowley'in asimetri ölçüsü,

$$BÇ = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Pearson'ın asimetri ölçüsü,

$$PÇ = \frac{\bar{X} - mod}{S}$$

$$PÇ = \frac{3(\bar{X} - medyan)}{S}$$

$BÇ \sim 0$  ise dağılım simetrik

$BÇ < 0$  ise sola çarpık

$BÇ > 0$  ise sağa çarpık

$PÇ \sim 0$  ise dağılım simetrik

$PÇ < 0$  ise sola çarpık

$PÇ > 0$  ise sağa çarpık

**Örnek 6:** Öğrencilerin bilgisayar başındaki geçirdikleri süreye ilişkin aşağıdaki istatistikler verilmiştir. Buna göre Bowley ve Pearson çarpıklık katsayılarını bulunuz.

$$BÇ = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \frac{6,4 + 4,6 - 2 \times 5,6}{6,4 - 4,6} = -0,1111$$

$$PÇ = \frac{\bar{X} - mod}{S} = \frac{5,37 - 4,6}{1,69} = 0,456$$

$$PÇ = \frac{3(\bar{X} - medyan)}{S} = \frac{3(5,37 - 5,6)}{1,69} = -0,41$$

Statistics		
süre		
N	Valid	50
	Missing	0
Mean		5,3680
Std. Error of Mean		,23941
Median		5,6000
Mode		4,60
Std. Deviation		1,69290
Variance		2,866
Range		9,00
Minimum		,90
Maximum		9,90
Sum		268,40
Percentiles	25	4,6000
	50	5,6000
	75	6,4000

**Örnek 7:** 24 ailenin aylık gıda harcaması aşağıda verilmiştir. Bowley ve Pearson çarpıklık katsayılarını hesaplayınız.

1050	1100	1200	1200	1200	1230	1300	1300	1420	1450	1450	1460
1500	1500	1500	1520	1550	1580	1600	1600	1700	1700	2000	2500

$$n = 24 \quad i = \frac{n}{4} = 6 \quad Q_1 = \frac{x_6 + x_7}{2} = \frac{1230 + 1300}{2} = 1265$$

$$i = \frac{n}{2} = 12 \quad Q_2 = \frac{x_{12} + x_{13}}{2} = \frac{1460 + 1500}{2} = 1480$$

$$i = \frac{3n}{4} = 18 \quad Q_3 = \frac{x_{18} + x_{19}}{2} = \frac{1580 + 1600}{2} = 1590$$

$$BÇ = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \frac{1590 + 1265 - 2 \times 1480}{1590 - 1265} = \frac{-105}{325} = -0,32$$

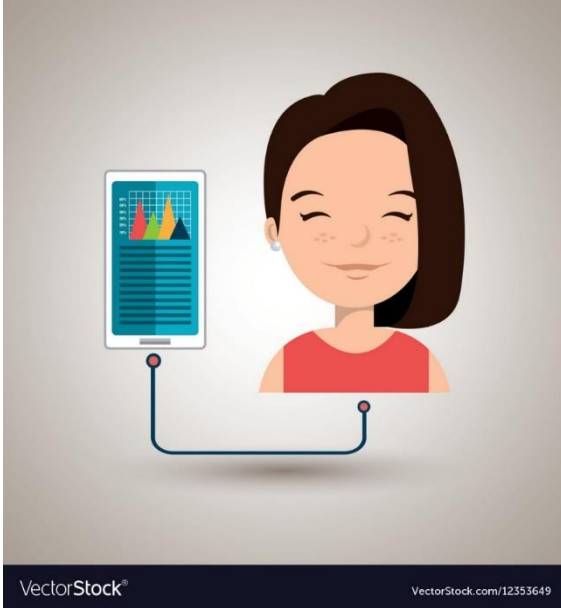
$$PÇ = \frac{\bar{X} - mod}{S} = \frac{1483,75 - 1200}{305,839} = 0.93$$

$$PÇ = \frac{3(\bar{X} - medyan)}{S} = \frac{3(1483,75 - 1480)}{305,839} = 0.036$$

$$\begin{aligned}\bar{X} &= 1483,75 \\ S &= 305,839 \\ mod &= 1200\end{aligned}$$







Bir sonraki derste «Aykırıdeğer» incelenecek.



## KAYNAKLAR

- 1-) H.Demirhan, C.Hamurkaroğlu ,“İstatistiksel Yöntemlere Giriş”, H.Ü.Yayınları, 2011.
- 2-) Serpil Cula, Zehra Muluk, “Temel İstatistik Yöntemler”, Başkent Üniversitesi yayınları,2006.
- 3-) Levent Özbek, Esin Köksal Babacan, “İstatistiğe Giriş”, TÜBİTAK e-kitap.
- 4-) Birdal Şenoğlu, Mehmet Yılmaz, Sibel Açık Kemaloğlu, İstatistiğe Giriş, TÜBİTAK e-kitap.