AKT102 İSTATİSTİK

BÖLÜM 2

BETİMLEYİCİ İSTATİSTİKLER II

§ 2.3

Merkezi Eğilim Ölçüleri

Ortalama

Merkezi eğilim ölçüsü, bir veri kümesinin tipik veya merkezi bir girişini temsil eden bir değerdir. En yaygın kullanılan üç merkezi eğilim ölçüsü ortalama, medyan ve moddur.

Bir veri kümesinin **ortalaması**, veri sayısına bölünen veri girişlerinin toplamıdır.

Kitle Ortalaması: $\mu = \frac{\sum x}{N}$ Örneklem Ortalaması: $\bar{x} = \frac{\sum x}{n}$

"mü"

"*x*-bar"

Ortalama

Örnek:

Aşağıdakiler küçük bir şirketin yedi çalışanının yaşlarıdır:

53 32 61 57 39 44 57

Kitle ortalamasını hesaplayın.

$$\mu = \frac{\sum \boldsymbol{x}}{\boldsymbol{N}} = \frac{343}{7}$$

Yaşları yoplayın ve 7'ye bölün.

$$= 49 \text{ years}$$

Çalışanların yaş ortalaması 49'dur.

Medyan (Ortanca)

Bir veri setinin ortancası, veri seti sıralandığında verilerin ortasında olan değerdir. Veri setinde tek sayıda veri varsa, medyan, ortada kalan veridir. Veri setinde çift sayıda veri varsa, medyan, ortadaki iki veri girişinin ortalamasıdır.

Örnek:

Yedi çalışanın ortanca yaşını hesaplayın.

53 32 61 57 39 44 57

Medyanı bulmak için verileri sıralayın.

32 39 44 $\boxed{53}$ 57 57 61

Çalışanların ortanca yaşı 53'tür.

Mod (Tepe Değeri)

Bir veri setinin modu, en yüksek frekansta gerçekleşen veri girişidir. Herhangi bir veri tekrar edilmezse, veri setinin modu yoktur. İki giriş (veri) aynı büyük frekansta gerçekleşirse, her giriş bir moddur ve veri kümesine çift modlu denir.

Örnek:
Yedi çalışanın yaş grubunu bulun..

53 32 61 $\boxed{57}$ 39 44 $\boxed{57}$

Mod 57, çünkü diğer veriler bir kez varken 57 iki kez tekrarlanıyor.

Aykırı değer, veri kümesindeki diğer girdilerden çok uzaktaki bir veri girişidir.

Ortalama-Mod-Medyan Karşılaştırılması

Örnek:

29 yaşında bir çalışan şirkete katılıyor ve çalışanların yaşları şimdi:

53 32 61 57 39 44 57 29

Ortalama, medyan ve modu yeniden hesaplayın. Bu yeni yaş eklendiğinde hangi merkezî eğilim ölçüsü etkilendi?

Mean = 46.5

Ortalama her değeri hesaba katar, ancak aykırı değerden etkilenir.

Median = 48.5

Ortanca ve mod uç değerlerden etkilenmez.

Mode = 57

Ağırlıklı Ortalama

Ağırlıklı ortalama, değişken ağırlıklara sahip olan bir veri kümesinin ortalamasıdır. Ağırlıklı bir ortalama;

$$\bar{x} = \frac{\sum(x \cdot w)}{\sum w}$$

buradaki w, her bir x girişin ağırlığıdır.

Örnek:

İstatistik sınıfındaki notlar aşağıdaki gibi ağırlıklandırılır:
Ara sınav notun% 50'sini, ödev notun% 30'unu, final% 20'sini etkilemektedir. Bir öğrenciye ara sınavda toplam 80 puan, ödevinde 100 puan ve finalde 85 puan verilir. Şu anki notu ne olur?

Ağırlıklı Ortalama

Verileri bir tabloda düzenleyerek başlayın.

Kaynak	**Not, *x***	**Ağırlık, *w***	***xw***
Ara sınav	80	0.50	40
Ödev	100	0.30	30
Final	85	0.20	17

$$\bar{x} = \frac{\sum(x \cdot w)}{\sum w} = \frac{87}{100} = 0.87$$

Öğrencinin mevcut notu 87'dir.

Frekans Dağılımının Ortalaması

Bir örneklem için bir frekans dağılımının ortalaması yaklaşık:

$$\bar{x} = \frac{\sum(x \cdot f)}{n} \qquad \text{Note that } n = \sum f$$

x ve f , sınıfların orta noktası ve frekans değerleridir.

Örnek:

Aşağıdaki sıklık dağılımı, bir istatistik sınıfındaki 30 öğrencinin yaşını temsil eder. Frekans dağılımının ortalamasını bulun.

Frekans Dağılımının Ortalaması

Sınıf orta noktası

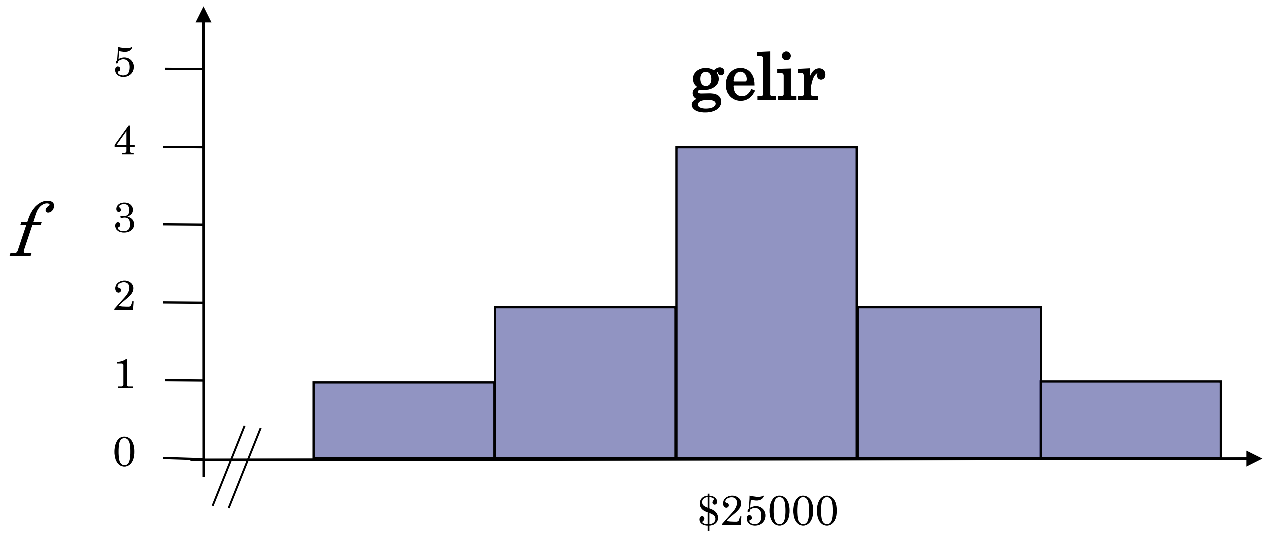
Class	x	f	$(x \cdot f)$
18 – 25	21.5	13	279.5
26 – 33	29.5	8	236.0
34 – 41	37.5	4	150.0
42 – 49	45.5	3	136.5
50 – 57	53.5	2	107.0
		$n = 30$	$\Sigma = 909.0$

$$\bar{x} = \frac{\sum(x \cdot f)}{n} = \frac{909}{30} = 30.3$$

Öğrencilerin yaş ortalaması 30,3'tür

Simetrik Dağılım

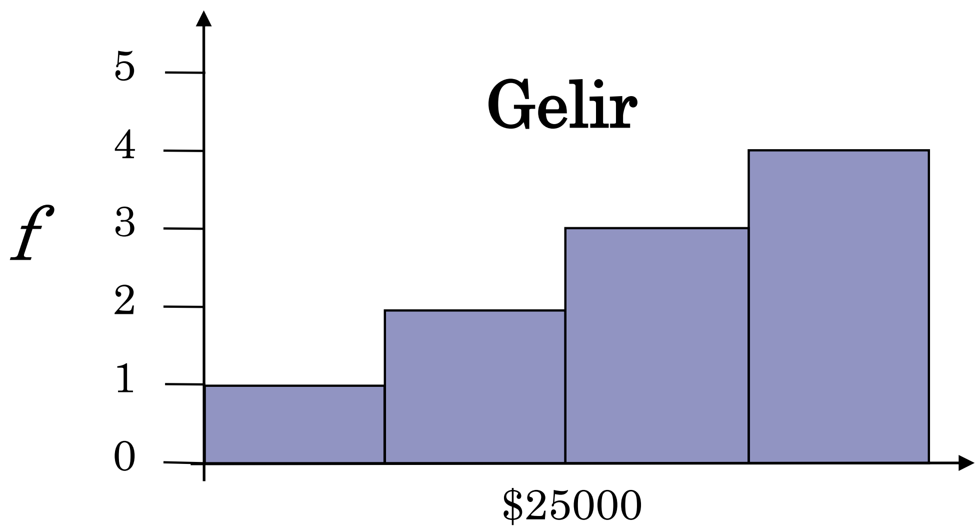
10 Yıllık Gelir
15,000
20,000
22,000
24,000
25,000
25,000
26,000
28,000
30,000
35,000



ortalama = medyan = mod
= $25,000

Sola Çarpık Dağılım

10 Yıllık Gelir
0
20,000
22,000
24,000
25,000
25,000
26,000
28,000
30,000
35,000

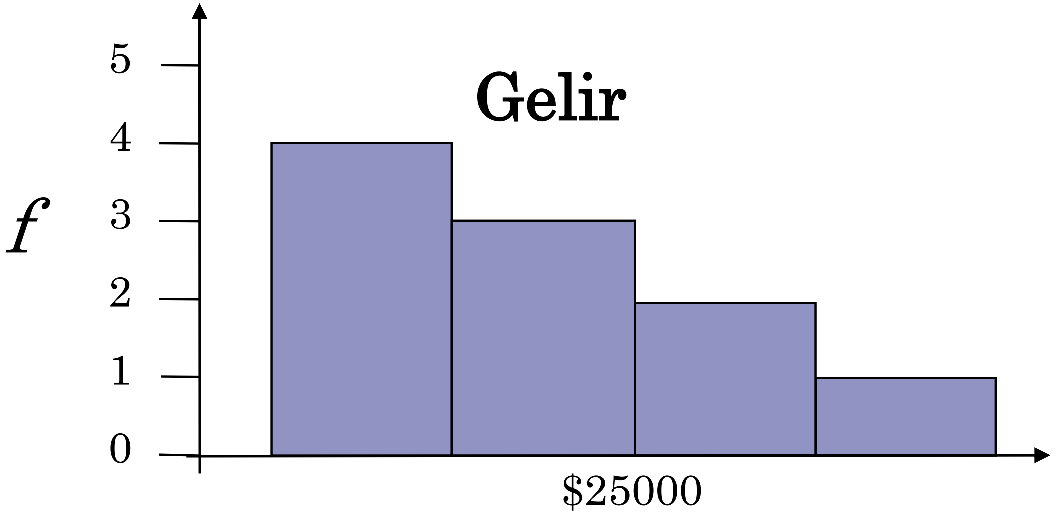


ortalama = $23,500
medyan = mod = $25,000

Ortalama < Medyan

Sağa Çarpık Dağılım

10 yıllık Gelir
15,000
20,000
22,000
24,000
25,000
25,000
26,000
28,000
30,000
1,000,000

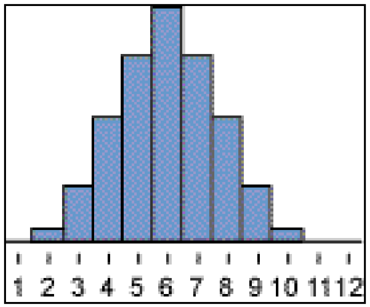


ortalama = $121,500
medyan = mod = $25,000

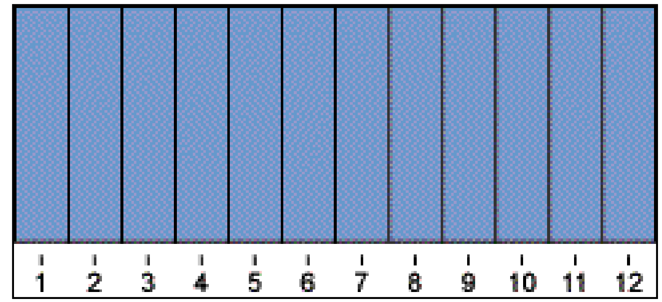
Ortalama > Medyan

Dağılım Şekillerinin Özeti

Simetrik

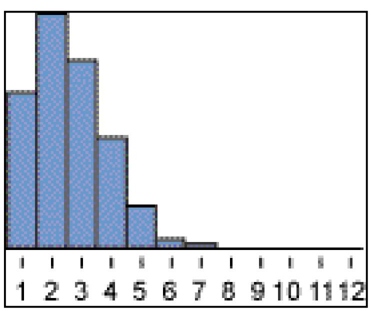


Uniform



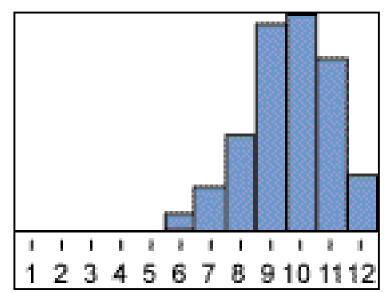
Ortalama = Medyan

Sağa Çarpık



Ortalama > Medyan

Sola Çarpık



Ortalama < Medyan

§ 2.4

Dağılma Ölçüleri

Açıklık

Bir veri kümesinin aralığı, kümedeki maksimum ve minimum veri girişleri arasındaki farktır.
Açıklık = (En büyük veri) – (En küçük veri)

Örnek:
Aşağıdaki veriler, ardışık on Cuma günü belirli bir hisse senedi için kapanış fiyatlarıdır. Aralığı bul.

Stock	56	56	57	58	61	63	63	67	67	67

Aralık : 67 – 56 = 11.

Sapma

Bir popülasyon veri setinde bir x girişinin sapması, giriş ile veri setinin ortalaması µ arasındaki farktır.

x in sapması: $x = x - \mu$

Örnek:
Yandaki veriler ardışık beş Cuma günü belirli bir hisse senedi için kapanış fiyatlarıdır. Her fiyatın sapmasını bulun.

Fiyat x	Sapma $x - \mu$
56	56 – 61 = – 5
58	58 – 61 = – 3
61	61 – 61 = 0
63	63 – 61 = 2
67	67 – 61 = 6
$\Sigma x = 305$	$\Sigma(x - \mu) = 0$

Ortalama hisse senedi fiyatı

$$\mu = 305/5 = 61.$$

Varyans ve Standart Sapma

N hacimli bir kitle verisi için kitle varyansı;

$$\text{Kitle varyansı} = \sigma^2 = \frac{\sum(x-\mu)^2}{N}.$$

“sigma kare”

N tane kitle veri setinin kitle standart sapması kitle varyansının

kareköküdür.

$$\text{Kitle standart sapması} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x-\mu)^2}{N}}.$$

“sigma”

Kitle Standart Sapmasının Bulunması

Kılavuz

Açıklama	*Gösterim*
1. Kitle veri setinin ortalamasını bulun.	$\mu = \frac{\sum x}{N}$
2. Her girişin sapmasını bulun.	$x - \mu$
3. Her sapmanın karesini alın.	$(x - \mu)^2$
4. Karelerin toplamını almak için toplayın.	$SS_x = \sum (x - \mu)^2$
5. Kitle varyansını elde etmek için N'ye bölün.	$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$
6. Kitle standart sapmasını elde etmek için varyansın karekökünü bulun.	$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$

Örneklem Standart Sapmasının Bulunması

Kılavuz

Açıklama	*Gösterim*
1. Örneklem veri setinin ortalamasını bulun.	$\bar{x} = \dfrac{\sum x}{n}$
2. Her girişin sapmasını bulun.	$x - \bar{x}$
3. Her sapmanın karesini alın.	$(x - \bar{x})^2$
4. Karelerin toplamını bulmak için toplayın.	$SS_x = \sum (x - \bar{x})^2$
5. Örneklem varyansını elde etmek için n - 1'e bölün.	$s^2 = \dfrac{\sum (x - \bar{x})^2}{n - 1}$
6. Örneklem standart sapmasını elde etmek için varyansın karekökünü bulun	$s = \sqrt{\dfrac{\sum (x - \bar{x})^2}{n - 1}}$

Kitle Standart Sapmasının Bulunması

Örnek:
Aşağıdaki veriler ardışık beş Cuma günü belirli bir hisse senedi için kapanış fiyatlarıdır. Kitle ortalaması 61'dir. Kitle standart sapmasını bulun.

Her zaman pozitif!

Fiyat x	Sapma $x - \mu$	Kareler $(x - \mu)^2$
56	– 5	25
58	– 3	9
61	0	0
63	2	4
67	6	36
$\Sigma x = 305$	$\Sigma(x - \mu) = 0$	$\Sigma(x - \mu)^2 = 74$

$$SS_2 = \Sigma(x - \mu)^2 = 74$$

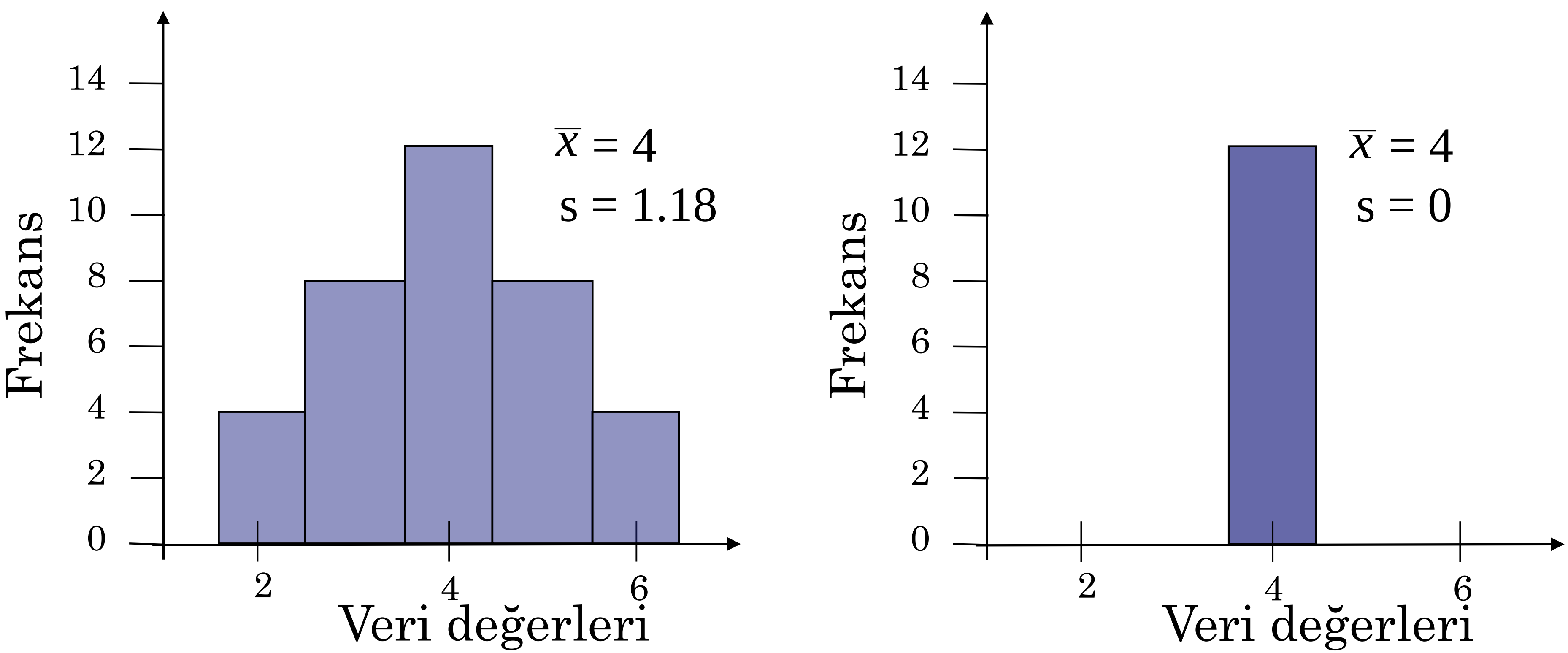
$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{74}{5} = 14.8$$

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} = \sqrt{14.8} \approx 3.85$$

$$\sigma \approx \$3.85$$

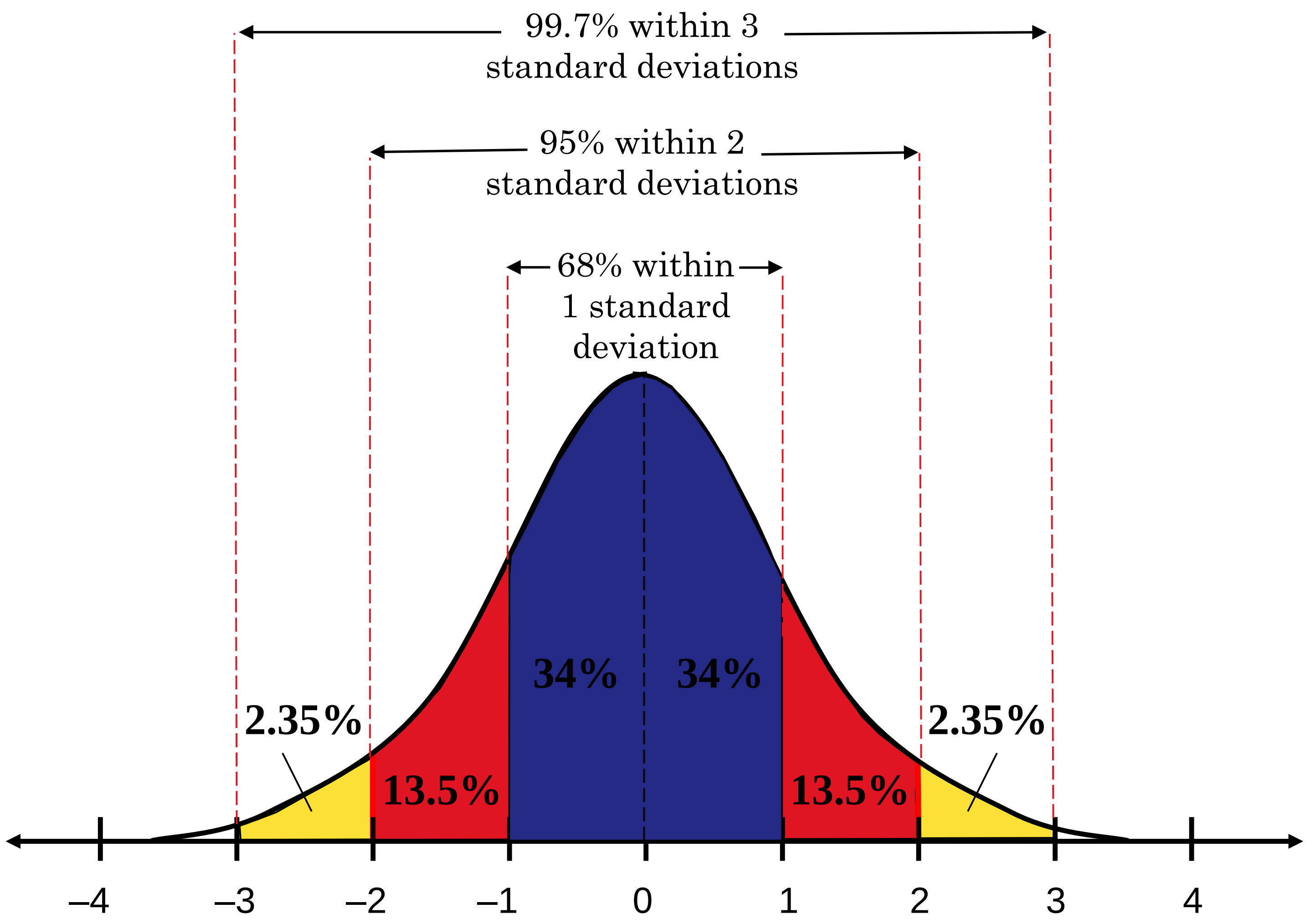
Standart Sapmanın Yorumlanması

Standart sapmayı yorumlarken, bunun bir verinin ortalamadan saptığı tipik miktarın bir ölçüsü olduğunu unutmayın. Veriler ne kadar çok yayılırsa, standart sapma o kadar büyük olur.



Simetrik (çan biçimli) bir dağılım gösteren veriler için standart sapma aşağıdaki özelliklere sahiptir.

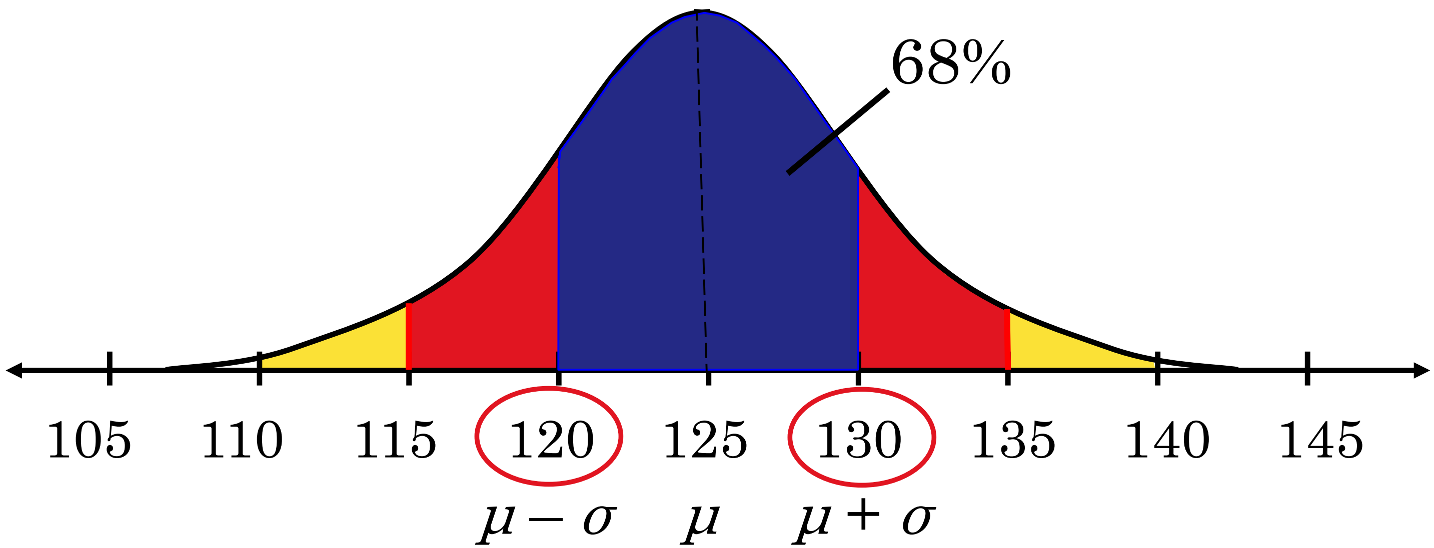
1. Verilerin yaklaşık% 68'i ortalamanın bir standart sapması içinde yer almaktadır.
2. Verilerin yaklaşık% 95'i ortalamanın iki standart sapması içinde yer almaktadır.
3. Verilerin yaklaşık% 99.7'si ortalamanın üç standart sapması içinde yer almaktadır.



99.7% within 3 standard deviations
95% within 2 standard deviations
68% within 1 standard deviation
2.35%
13.5%
34%
34%
13.5%
2.35%
–4
–3
–2
–1
0
1
2
3
4

Örnek:

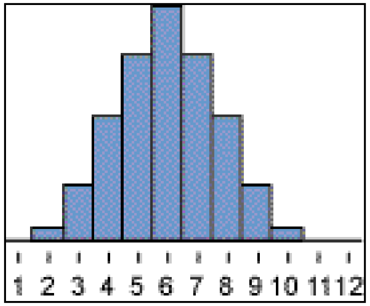
Sokaktaki evlerin ortalama değeri 5 bin dolarlık standart sapma ile 125 bin dolar. Veri seti çan şeklinde bir dağılıma sahiptir. 120 ila 130 bin dolar arasındaki evlerin yüzdesini tahmin et.



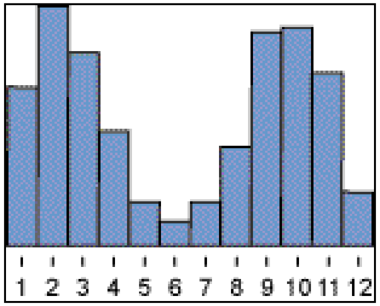
Evlerin% 68'i 120 ile 130 bin dolar arasında bir değere sahip.

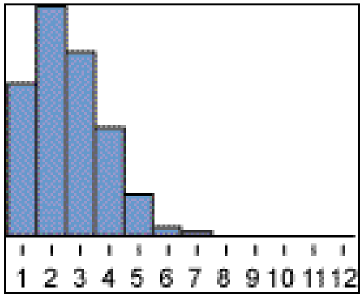
Chebyshev Teoremi

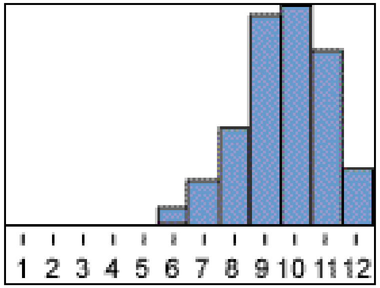
Ampirik kural sadece simetrik dağılımlar için kullanılır.



Chebyshev Teoremi, şekli ne olursa olsun, herhangi bir dağılım için kullanılabilir.







Chebychev Teoremi

Herhangi bir veri setinin ortalamanın k standart sapma (k> 1) içinde kalan kısmı en az;

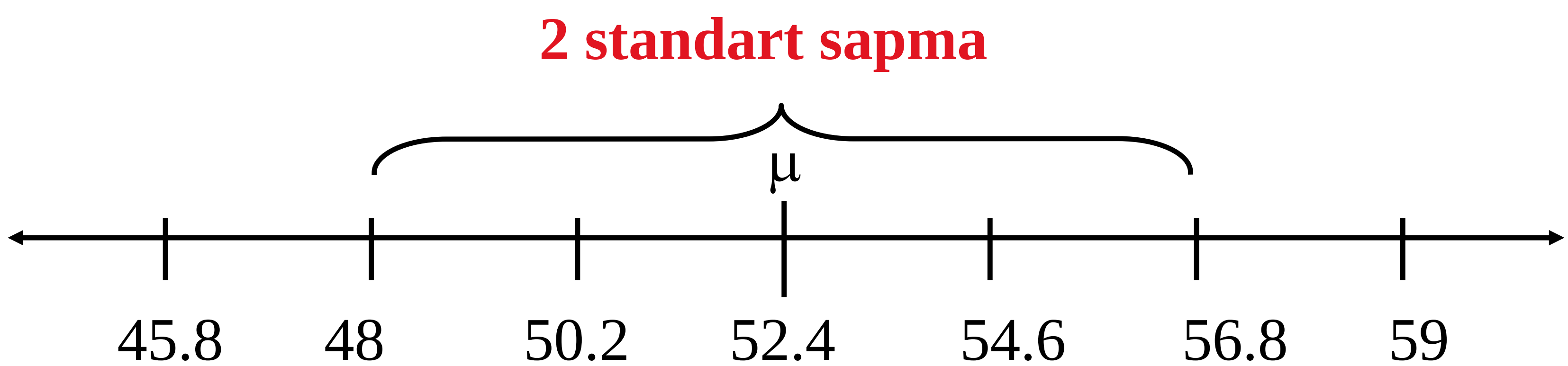
$$1-\frac{1}{k^2}.$$

$k = 2$ için: Herhangi veri setinde, en az $1-\frac{1}{2^2}=1-\frac{1}{4}=\frac{3}{4},$ ya da %75'i, ortalamanın 2 standart sapması içindedir.

$k = 3$ için: Herhangi veri setinde, en az $1-\frac{1}{3^2}=1-\frac{1}{9},=\frac{8}{9},$ ya da %88.9'u, ortalamanın 3 standart sapması içindedir..

Chebyshev Teoremi Kullanımı

Örnek:
Bir kadının 400 metrelik koşusunda ortalama süre, standart 2,2 saniyelik sapma ile 52,4 saniyedir. Kadınların en az% 75'i hangi iki değer arasında kalacak?



Kadınların 400 metrelik çizgi sürelerinin en az% 75'i 48 ile 56.8 saniye arasındadır.

Grup Verilerinde Standart Sapma

Örneklem standart sapması$= s = \sqrt{\dfrac{\sum (x-\overline{x})^2 f}{n-1}}$

$n = \Sigma f$ veri kümesindeki girişlerin sayısıdır ve x bir aralığın veri değeri veya orta noktasıdır.

Örnek:
Aşağıdaki sıklık dağılımı, bir istatistik sınıfındaki 30 öğrencinin yaşını temsil eder. Öğrencilerin yaş ortalaması 30,3'tür. Frekans dağılımının standart sapmasını bulun.

Grup Verilerinde Standart Sapma

Öğrencilerin yaş ortalaması 30,3.

Sınıf	x	f	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
18 – 25	21.5	13	– 8.8	77.44	1006.72
26 – 33	29.5	8	– 0.8	0.64	5.12
34 – 41	37.5	4	7.2	51.84	207.36
42 – 49	45.5	3	15.2	231.04	693.12
50 – 57	53.5	2	23.2	538.24	1076.48
		$n = 30$		$\sum = 2988.80$	

$$s = \sqrt{\frac{\sum(x-\bar{x})^2 f}{n-1}} = \sqrt{\frac{2988.8}{29}} = \sqrt{103.06} = 10.2$$

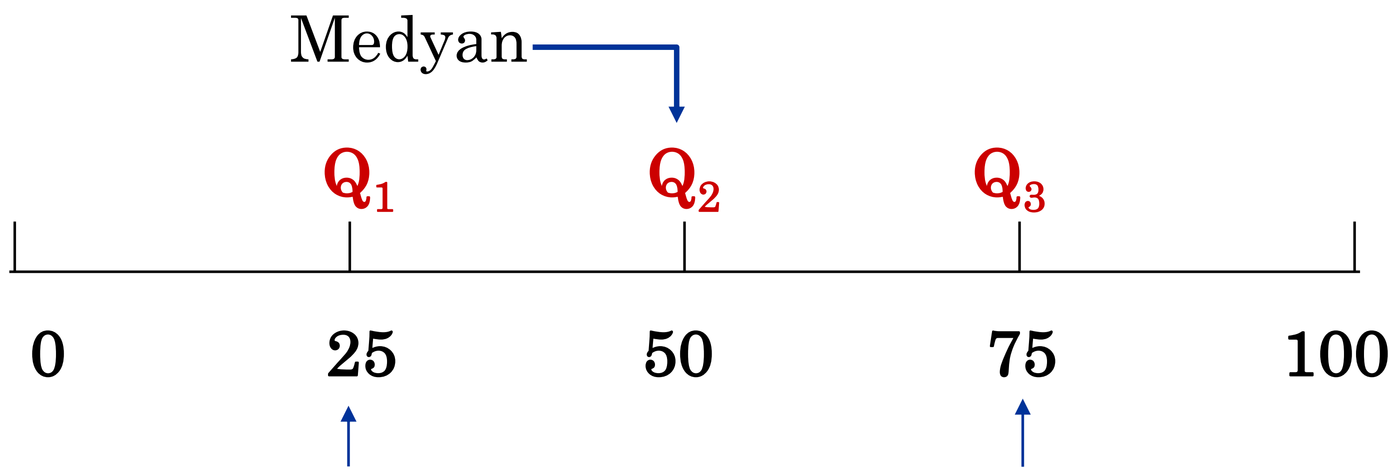
Yaşların standart sapması 10.2'dir.

§ 2.5

Konum Ölçümleri

Çeyreklikler

Üç çeyrek Q1, Q2 ve Q3, sıralı verileri yaklaşık dört eşit parçaya böler.



Q1, Q2 altındaki verilerin ortancasıdır.

Q3, Q2 üzerindeki verilerin ortancasıdır.

Çeyreklik Bulma

Örnek:

15 öğrencinin sınav puanları aşağıda listelenmiştir. Skorların birinci, ikinci ve üçüncü çeyreğini bulun.

28 43 48 51 43 30 55 44 48 33 45 37 37 42 38

Verileri sıralayın.

$$\overbrace{28\quad 30\quad 33\quad \underset{\substack{\uparrow \\ Q_1}}{37}\quad 37\quad 38\quad 42}^{\text{Alt yarı}}\quad \underset{\substack{\uparrow \\ Q_2}}{43}\quad \overbrace{43\quad 44\quad 45\quad \underset{\substack{\uparrow \\ Q_3}}{48}\quad 48\quad 51\quad 55}^{\text{Üst yarı}}$$

Öğrencilerin yaklaşık dörtte biri 37 veya daha az puan alır; yaklaşık yarısı 43 ya da daha az; ve yaklaşık dörtte üçü 48 veya daha az puan alır.

Çeyrekler Arası Genişlik

Bir veri kümesinin çeyrekler arası genişlik (IQR), üçüncü ve ilk çeyrekler arasındaki farktır.

$$\text{Çeyrekler arası genişlik(IQR)} = Q_3 - Q_1.$$

Örnek:

15 sınav puanı için quartiles aşağıda listelenmiştir. Çeyrekler arası genişliği bulun.

$Q_1 = 37$ $Q_2 = 43$ $Q_3 = 48$

$$\begin{aligned}(\text{IQR}) &= Q_3 - Q_1 \\ &= 48 - 37 \\ &= 11\end{aligned}$$

Veri setinin orta bölümündeki sınav puanları en fazla 11 puan arasında değişmektedir.

Kutu Grafiği

Box-and-whisker plot bir veri setinin önemli özelliklerini vurgulayan bir keşif veri analizi aracıdır.

Beş tane değer, grafiği çizmek için kullanılır:

- minimum veri
- Q_1
- Q_2 (medyan)
- Q_3
- maximum veri

Örnek:

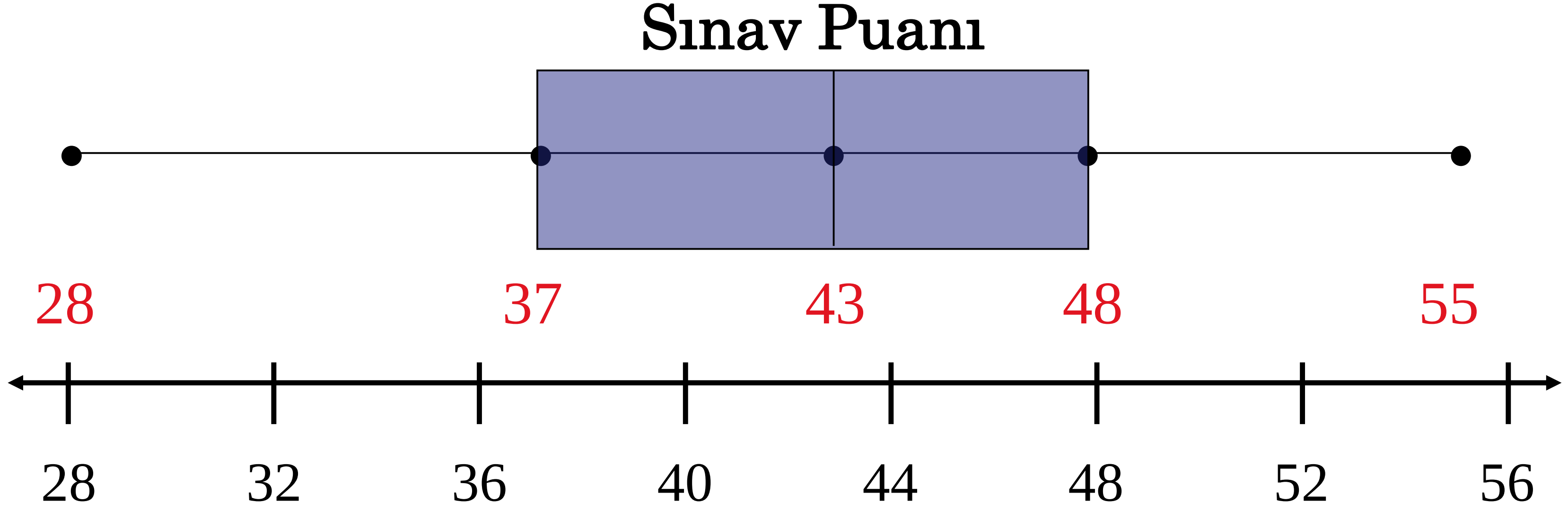
Box-and-whisker plot çizmek için 15 sınav puanından verileri kullanın.

28 30 33 37 37 38 42 43 43 44 45 48 48 51 55

Kutu Grafiği

5 tane değer

• minimum veri	28
• Q_1	37
• Q_2 (median)	43
• Q_3	48
• maximum veri	55



Yüzdelikler ve Ondalıklar

Fraktiller sıralı bir veri setini bölümleyen veya bölen sayılardır.

Yüzdelikler sıralı verileri 100 bölüme ayırır. 99 tane yüzdelik var : P_1, P_2, P_3…P_{99}.

Ondalıklar sıralı verileri 10 bölüme ayırır. 9 tane ondalık var : D_1, D_2, D_3…D_9.

80. yüzdelik sınav puanı (P_{80}), test puanının diğer tüm test puanlarının% 80'inden büyük olduğunu ve puanların % 20'sinden küçük veya ona eşit olduğunu gösterir.

Standart Puan

Standart puan veya z puanı, bir x veri değerinin μ ortalamadan düştüğü standart sapma sayısını gösterir.

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

Örnek:
Union College'daki tüm istatistik finalleri için test puanlarının ortalaması 78 ve standart sapma 7'dir. Aşağıdaki değerler için z puanını bulun.

a.) 85 test puanı,

b.) 70 test puanı,

c.) 78 test puanı.

Standart Puan

Örneğin devamı:

a.) $\mu = 78, \quad \sigma = 7, \quad x = 85$

$$z = \frac{x-\mu}{\sigma} = \frac{85-78}{7} = 1.0$$

Bu puan ortalamadan 1 standart sapma daha yüksektir

b.) $\mu = 78, \quad \sigma = 7, \quad x = 70$

$$z = \frac{x-\mu}{\sigma} = \frac{70-78}{7} = -1.14$$

Bu puan ortalamadan 1,14 standart sapma daha düşüktür.

c.) $\mu = 78, \quad \sigma = 7, \quad x = 78$

$$z = \frac{x-\mu}{\sigma} = \frac{78-78}{7} = 0$$

Bu puan ortalama ile aynıdır.

Göreli Z-Puanı

Örnek:
John sınıf ortalaması 73.2 ve standart sapması 4.5 olan bir testten 75 aldı. Samantha sınıf ortalaması 65 ve standart sapması 3.9 olan bir testten 68.6 aldı. Hangi öğrenci daha iyi test puanına sahiptir?

John'un *z*-puanı

$$z = \frac{x - \mu}{\sigma} = \frac{75 - 73.2}{4.5}$$
$$= 0.4$$

Samantha'nın *z*-puanı

$$z = \frac{x - \mu}{\sigma} = \frac{68.6 - 65}{3.9}$$
$$= 0.92$$

John'un puanı ortalamadan 0,4 standart sapma daha yüksek, Samantha'nın puanı ortalamadan 0,92 standart sapma daha yüksek bulundu. Bu nedenle Samantha'nın test puanı John'dan daha iyidir.