

# Necmettin Erbakan Üniversitesi

## Bilgisayar Mühendisliği

Dr. Öğr. Üyesi Murat KARAKOYUN  
([mkarakoyun@erbakan.edu.tr](mailto:mkarakoyun@erbakan.edu.tr))

Ders: Veri Önışlemeye Giriş  
[Konu: Veri Özetlenmesi ve Sunulması]

# Sunu İçeriği

---

- Giriş
- Histogram Analizi
- Yüzde, Yüzdelik ve Çeyrekler Kavramları
- Ortalama, Medyan ve Mod Kavramları
- Varyans ve Standart Sapma
- Verilerin Dağılımının Ölçülmesi
- Kutu Grafiği Analizi
- Örnekler

# Giriş

---

- ❑ Veri özetlemesi ve görsel sunum neden gereklidir?
- ❑ Ne zaman yapılmalıdır?

# Histogram Analizi

---

## Frekans:

- ❑ En basit tanımı ile bir değerin gözlenme sıklığı, tekrar sayısıdır.
- ❑ Frekans dağılımı ise ölçülen her kategoriye ve her kategori için tekrar sayısını listelemektedir.

# Histogram Analizi

## Frekans:

Kırmızı, Yeşil, Yeşil, Yeşil, Mavi, Mavi, Kırmızı, Mavi, Yeşil, Yeşil, Kırmızı, Kırmızı, Mavi, Yeşil, Kırmızı, Kırmızı.

RENK	FREKANS	KÜMÜLATİF FREKANS
Kırmızı	6	6
Yeşil	6	12
Mavi	4	16

# Histogram Analizi

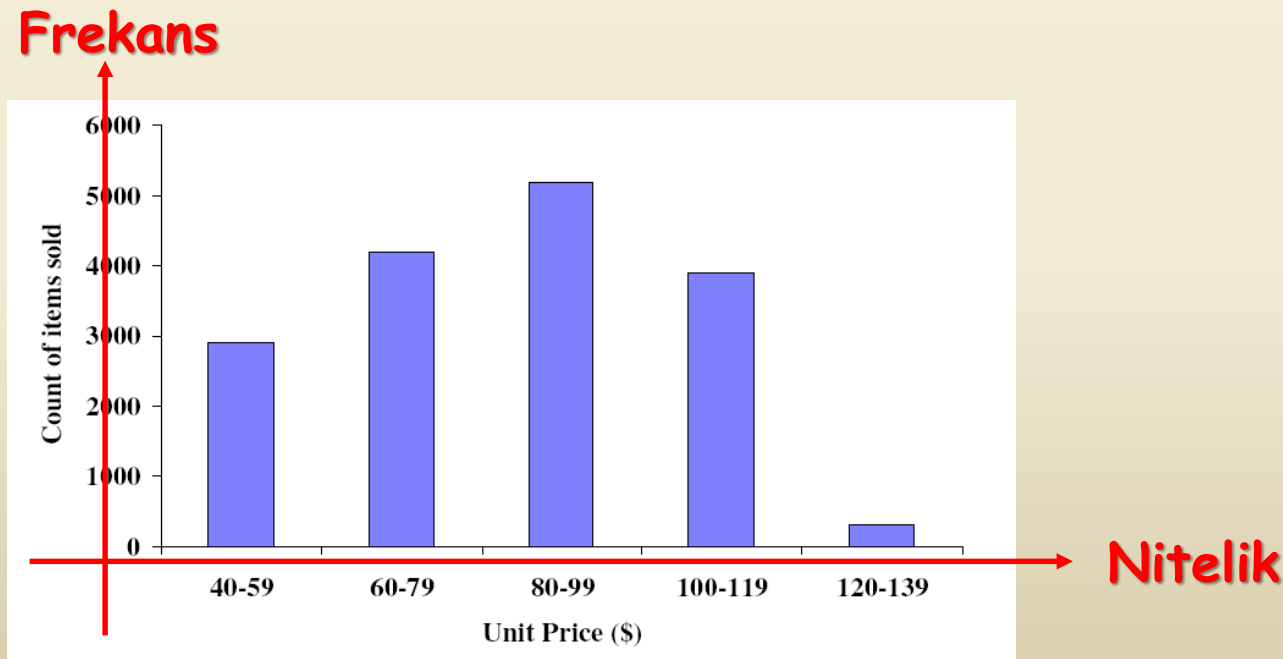
---

## Histogram:

- ❑ Yatay ekseninde ölçülen her kategoriye ve dikey eksenindeki her kategori için meydana gelme sayısını (frekans) listeleyen bir çubuk grafiktir.
- ❑ İki eksenin kesiştiği nokta her iki eksen için de en düşük değeri gösterir. Yatay eksen, her zaman 0'dan başlamak zorunda değildir.

# Histogram Analizi

Histogram, bar diagramı ya da çizgi grafiği çizilirken genellikle bir değişkenin düzeyleri ya da puanlar yatay ekseninde (X), bu düzey ya da puanlara karşılık gelen frekanslar dikey ekseninde (Y) gösterilir.



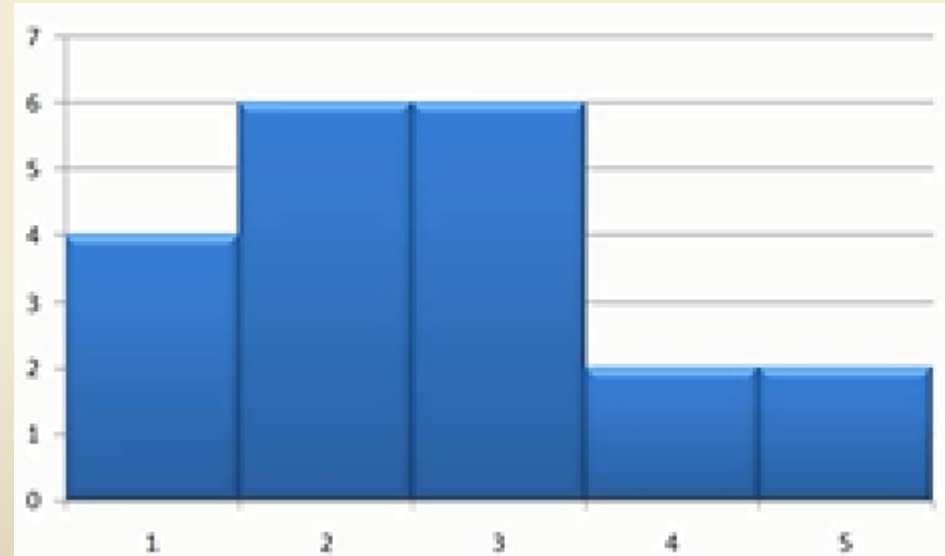
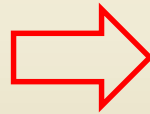
# Histogram Analizi

## Ayrık Histogram

3	1	3	2	2
1	2	3	2	2
5	5	2	3	3
4	4	1	1	3



DEĞER	FREKANS
1	4
2	6
3	6
4	2
5	2





# Histogram Analizi

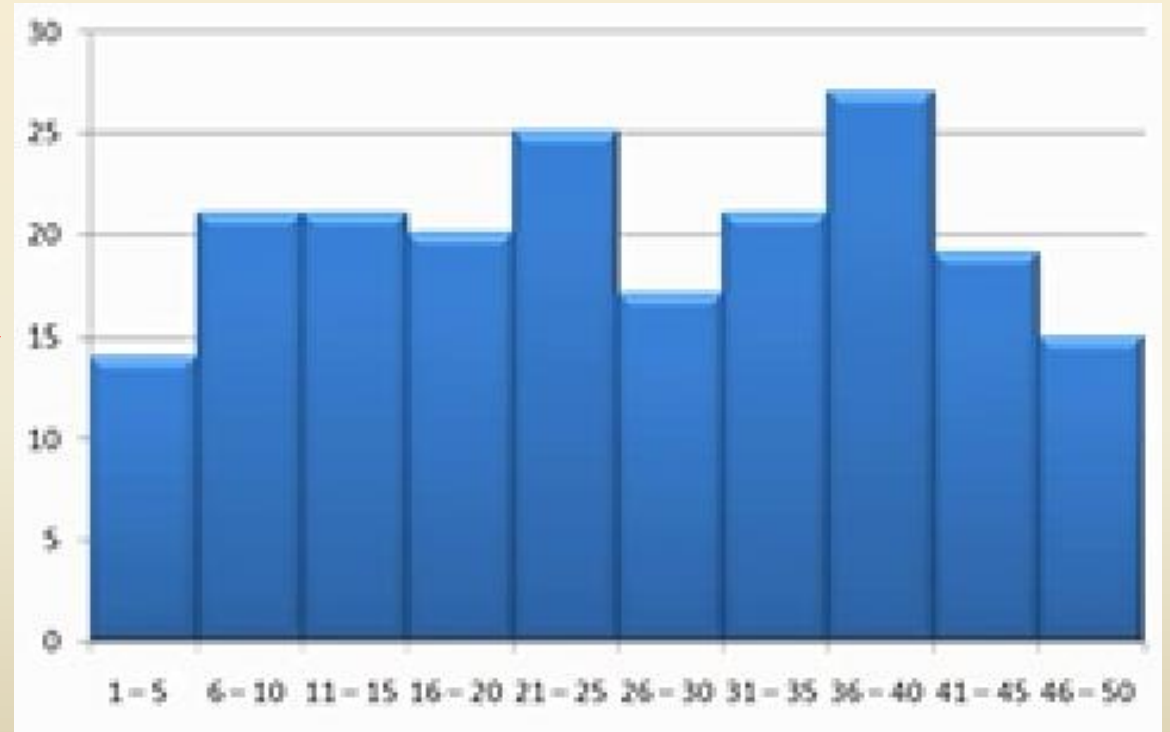
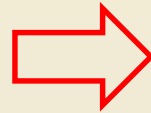
## Sürekli Histogram

38	29	42	24	46	15	23	13	16	42
36	37	35	19	11	28	8	10	25	25
43	44	8	22	27	2	15	28	4	42
38	35	14	8	10	7	1	2	38	39
22	12	35	16	35	34	12	34	23	3
34	10	3	48	47	29	31	20	2	26
46	13	2	3	38	5	46	45	6	29
19	13	30	28	22	40	23	40	14	20
28	22	10	49	44	40	48	32	31	40
17	29	7	35	45	37	50	50	28	29
17	22	9	28	45	51	38	36	25	4
7	34	37	13	16	21	13	34	18	38
42	48	33	40	18	6	14	42	17	39
33	29	16	44	31	34	36	1	47	37
32	46	49	6	5	42	38	39	13	14
26	9	37	10	45	25	22	23	38	21
13	20	22	32	17	33	8	11	37	20
25	23	36	45	22	12	42	29	22	42
19	12	36	20	5	47	44	22	11	45
9	19	33	9	14	23	7	21	7	19

# Histogram Analizi

## Sürekli Histogram

ARALIK	FREKANS
1-5	14
6-10	21
11-15	21
16-20	20
21-25	25
26-30	17
31-35	21
36-40	27
41-45	19
46-50	15



# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzde (Percentage):

Bir veri kümesinde ilgilenilen verinin bulunma oranıdır.

Veri kümesi: 1, 2, 3, 4, 5

# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzde (Percentage):

Bir veri kümesinde ilgilenilen verinin bulunma oranıdır.

Veri kümesi: 1, 2, 3, 4, 5

Veri kümesindeki çift sayıların yüzdesi?

Yüzde (%) = (Çift sayıların sayısı/toplam sayı)\*100

Yüzde (%) =  $(2/5)*100 = 40$

Sayıların % 40'ı çifttir.

# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzdelik (Percentile):

Kendisinin altında belirli bir veri yüzdesinin bulunduğu değerdir.

Veri kümesi: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzdelik (Percentile):

Kendisinin altında belirli bir veri yüzdesinin bulunduğu değerdir.

Veri kümesi: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, **10**, 11, 11, 12

«**10**» sayısının yüzdelik sıralaması (dilimi) nedir?

# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzdelik (Percentile):

$X$ 'in yüzdelik dilimi = ( $X$ 'ten küçük olanların sayısı / Toplam sayı) \* 100

Veri kümesi: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

10'un yüzdelik dilimi = (16 / 20) \* 100

10 sayısının yüzdelik dilimi: %80

# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzdelik (Percentile):

Verilen veri kümesinde %X'lik yüzdelik dilim sıralamasında hangi değer bulunur?

Veri kümesi: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

**%25**'lik yüzdelik dilim sıralamasında hangi veri bulunur?

$\text{Indeks}[\text{değer}] = (\% \text{ Değer} / 100) * (\text{toplam sayı} + 1)$

$\text{Indeks}[\text{değer}] = (25 / 100) * (20 + 1) = 5.25$



# Yüzde, Yüzdelik, Çeyrekler

---

## Yüzdelik (Percentile):

Veri kümesi: 2, 2, 3, 4, **5**, **5**, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

5.25 indeks değeri olamayacağından 5 ve 6 indekslerindeki değerlerin ortalaması alınır.

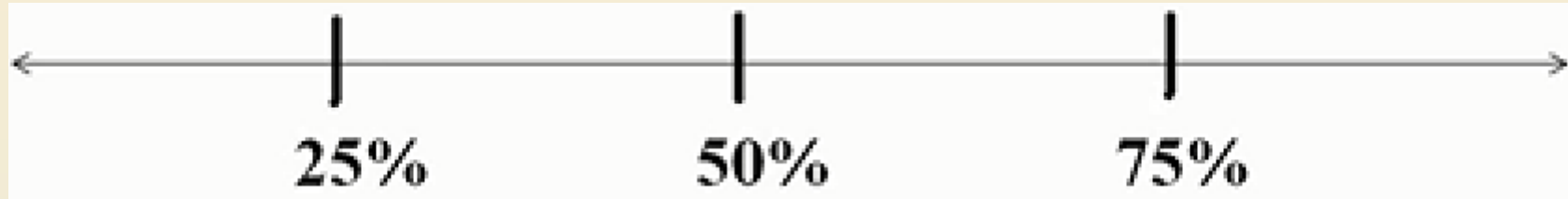
$$(5+5)/2 = 5$$

# Yüzde, Yüzdelik, Çeyrekler

---

## Çeyrekler (Quartiles):

Veri kümesindeki %25, %50 ve %75 dilim sıralamasında kalan değerlerin oluşturduğu kümedir.



İlk Çeyrek - 25%

İkinci Çeyrek - 50%

Üçüncü Çeyrek - 75%

# Ortalama, Medyan, Mod

---

## Ortalama (Mean):

- ❑ İstatistiksel bilim dalında ve veri analizinde kullanılan bir veri kümesinin orta konumunu, tek bir sayı ile ifade eden ölçüdür.
- ❑ Ortalamayı bulmak için tüm değerler toplanır, ardından veri kümesi sayısına bölünür.

**Veri kümesi: 1, 1, 2, 2, 2, 3, 3, 4, 5, 5**

$$\bar{X} = \frac{\sum x}{n} = \frac{1+1+2+2+2+3+3+4+5+5}{10} = \frac{28}{10} = 2.8$$

# Ortalama, Medyan, Mod

## Medyan (Median):

Ortanca olarak da isimlendirilen bu değer; bir veri kümesindeki sıralanmış verileri tam ortadan bölen sayıdır.

Veri kümesi: 1, 1, 2, 2, 2, **3**, 3, 4, 5, 5, 6

Medyan değeri =  $(n+1)/2$  ile hesaplanan sıradaki değerdir.

$$(n+1)/2 \rightarrow (11+1)/2 = \mathbf{6}$$

$$\text{Medyan} = \mathbf{3}$$

# Ortalama, Medyan, Mod

---

## Medyan (Median):

Hesaplanan sıra değeri her zaman bir tam sayı olmayabilir.

Veri kümesi: 1, 1, 2, 2, 2, 3, 3, 4, 5, 5

$$(n+1)/2 \rightarrow (10+1)/2 = 5.5$$

Bu durumda bulunan değerin sağında ve solunda kalan sıradaki iki değerin ortalaması ile **medyan** bulunur.

$$(2+3)/2 = 2.5$$

# Ortalama, Medyan, Mod

---

## Mod (Mode):

Bir veri kümesinde en sık görülen değerdir.

Başka bir deyişle frekansı en yüksek değerdir.

Veri kümesi: 1, 1, 2, 2, 2, 3, 3, 4, 5, 5

Bu veri kümesi için **mod** değeri 2'dir çünkü en sık görülen değer 2'dir.

# Ortalama, Medyan, Mod

---

Mod (Mode):

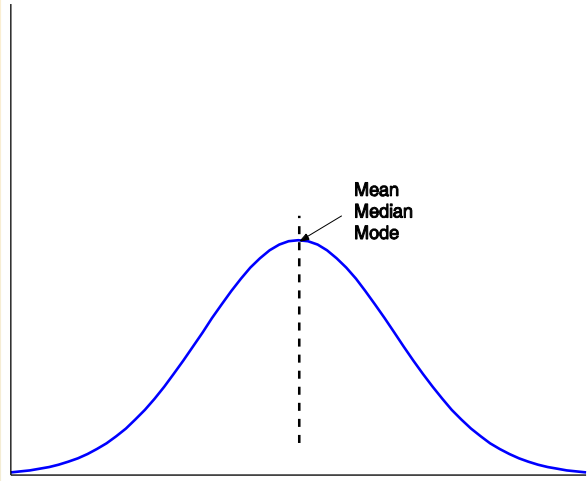
Veri kümesi: 1, 1, 2, 2, 2, 3, 3, 3, 4, 5, 5

Mod hangisidir?

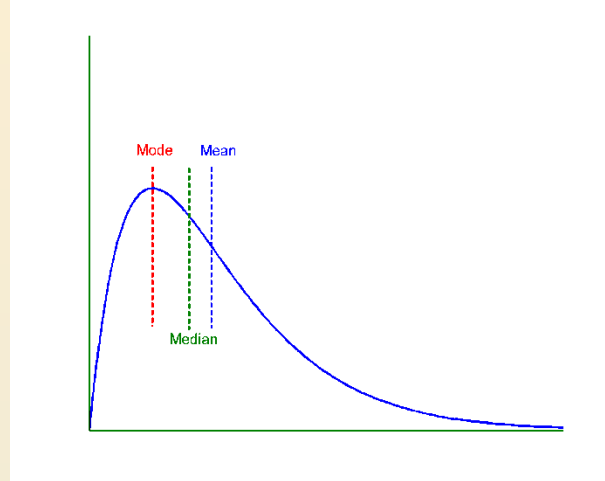
Bu veri kümesi **bi-modal** olarak geçmektedir ve mod değerleri 2 ve 3'tür.

Bir veri kümesinde birden fazla mod değeri varsa veri kümesi **multi-modal** olarak adlandırılmaktadır.

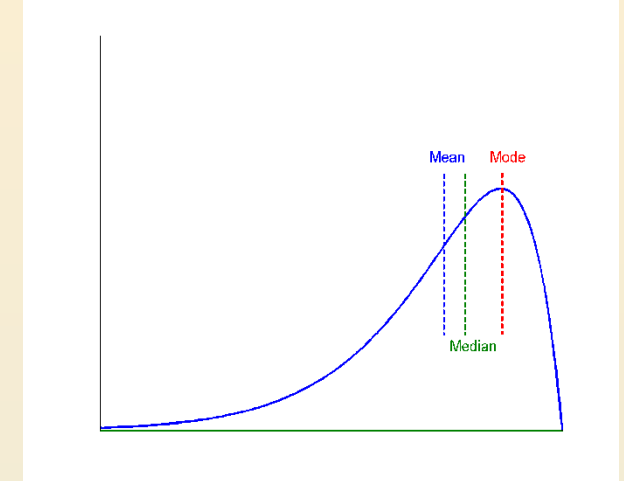
# Ortalama, Medyan, Mod



**Simetrik:** Verilerin dengeli dağılması durumudur. Dağılımın kütlesi merkeze doğrudur.



**Pozitif çarpıklık:** Bu halde sağdaki kuyruk daha uzundur. Dağılımın kütlesi grafiğin sol tarafında konsantre olmuştur. Bu türlü dağılım **sağdan (ya da sağa) çarpık** olarak anılır.



**Negatif çarpıklık:** Bu halde soldaki kuyruk daha uzundur ve dağılımın kütlesi grafiğin sağ tarafında konsantre olmuştur. Bu türlü dağılım **soldan çarpık** olarak anılır.



# Varyans ve Standart Sapma

---

## Varyans (Variance):

Verilerin (aritmetik) ortalamadan sapmalarının karelerinin toplamıdır.

## Standart Sapma (standart deviation):

Varyans değerinin karekök alınmış halidir.

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

**Varyans**

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

**Standart sapma**

# Verilerin Dağılımının Ölçülmesi

---

**Aralık:** Veri kümesindeki en büyük ve en küçük değerler arasındaki farktır.

Veri kümesi: **1**, 1, 2, 2, 2, 3, 3, 4, 5, 5, 7, 8, 8, 11, **12**

Bu veri kümesinin aralığı **11**'dir.

# Verilerin Dağılımının Ölçülmesi

---

## Çeyrekler:

- ❑  $Q_1$ : (%25'lik yüzdelerlik),
- ❑ **Medyan**: (%50'lik yüzdelerlik),
- ❑  $Q_3$ : (%75'lik yüzdelerlik).

# Verilerin Dağılımının Ölçülmesi

---

Çeyrekler Arası Aralık (Inter-Quartile Range, IQR):  $Q_3$  ile  $Q_1$  arasındaki farktır.

$$IQR = Q_3 - Q_1$$

# Verilerin Dağılımının Ölçülmesi

---

## Beş Sayı Özeti (Five Number Summary):

- ❑ Veri kümesindeki  $\min$ ,  $Q_1$ ,  $M$ ,  $Q_3$ ,  $\max$  değerlerinin ortaya koyduğu veri özetidir.
- ❑ Bu beş verinin görselleştirilmiş haline Kutu grafiği (Boxplot) denilmektedir.

# Verilerin Dağılımının Ölçülmesi

---

## Aykırı Değerler (Outlier):

Veri kümesinin ortalamasını ciddi anlamda etkileyen ve standart sapmanın yüksek çıkmasına sebep olan, verilerin genel yapısına uygun olmayan verilerdir.

Aykırı değerin tespiti:  $Q_3$  değerinden  $IQR \cdot 1.5$  miktardan daha büyük olan değerler veya  $Q_1$  değerinden  $IQR \cdot 1.5$  miktardan daha küçük değerlerdir.

# Verilerin Dağılımının Ölçülmesi

## Aykırı Değerler:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, 27

$$Q_1 = (25/100) * (19+1) = 5$$

$$Q_1 = 3$$

$$Q_3 = (75/100) * (19+1) = 15$$

$$Q_3 = 7$$

$$IQR = Q_3 - Q_1 = 4$$

# Verilerin Dağılımının Ölçülmesi

---

## Aykırı Değerler:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, 27

$$Q_1 = 3, Q_3 = 7, IQR = 4$$

Aykırı değerler için alt ve üst sınırları bulalım.

$$\text{Alt sınır: } Q_1 - 1.5 * IQR = 3 - 1.5 * 4 = -3$$

$$\text{Üst sınır: } Q_3 + 1.5 * IQR = 7 + 1.5 * 4 = 13$$

**Sonuç:** Veri kümesindeki **27** değeri aykırı bir değerdir.



# Verilerin Dağılımının Ölçülmesi

## Aykırı Değerler:

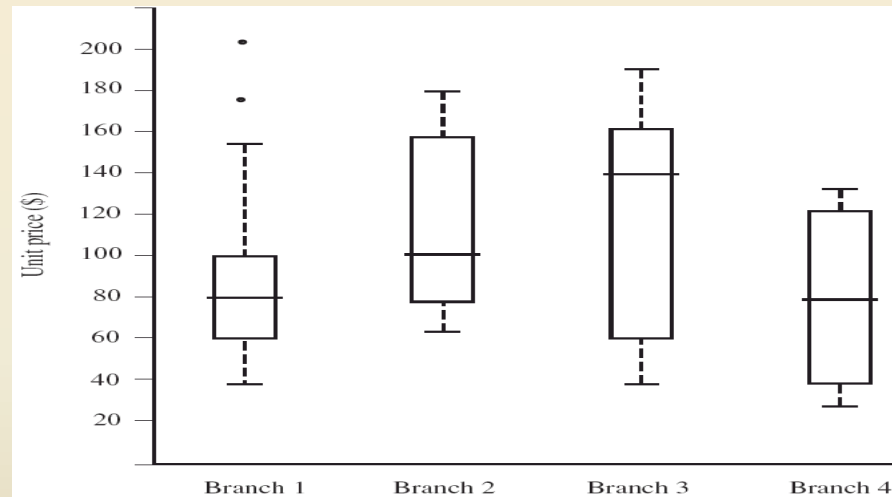
Veri kümesi: 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, **400**

	Aykırı değersiz	Aykırı değer ile
Ortalama	2.58	35.38
Medyan	2.5	2.5
Mod	2	2
Standart Sapma	1.08	114.74

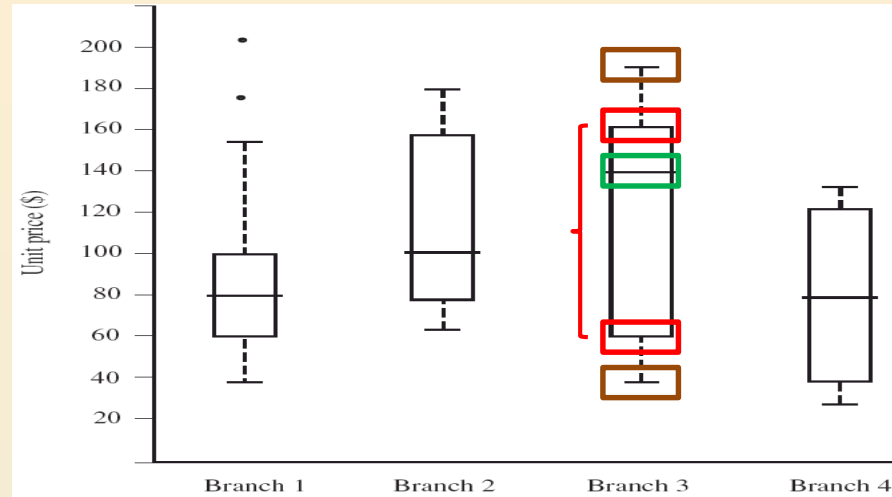
# Verilerin Dağılımının Ölçülmesi

## Kutu Grafiği Analizi:

- Beş sayı özetinin görselleştirilmesi ile elde edilen yapı **kutu grafiği (box plot)** olarak isimlendirilmektedir.



# Verilerin Dağılımının Ölçülmesi



Kutunun uçları birinci ( $Q_1$ ) ve üçüncü ( $Q_3$ ) çeyreklerdir.

Yani kutunun yüksekliği IQR'dır.

Medyan kutu içinde bir çizgi ile işaretlenmiştir.

Kutunun dışındaki iki çizgi Min ve Max değerlerini temsil eder ve Bıyık (whisker) olarak isimlendirilir.

# Verilerin Dağılımının Ölçülmesi

---

## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9, 9

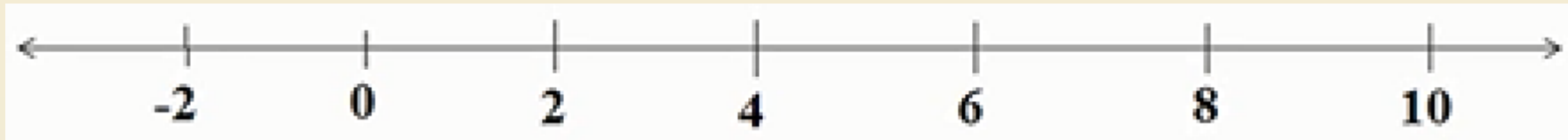
- |                              |   |
|------------------------------|---|
| 1. Minimum:                  | 1 |
| 2. Birinci Çeyrek ( $Q_1$ ): | 3 |
| 3. Medyan:                   | 5 |
| 4. Üçüncü Çeyrek ( $Q_3$ ):  | 7 |
| 5. Maksimum:                 | 9 |

# Verilerin Dağılımının Ölçülmesi

## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

Min= 1   Q1= 3   Median= 5   Q3= 7   Max= 9

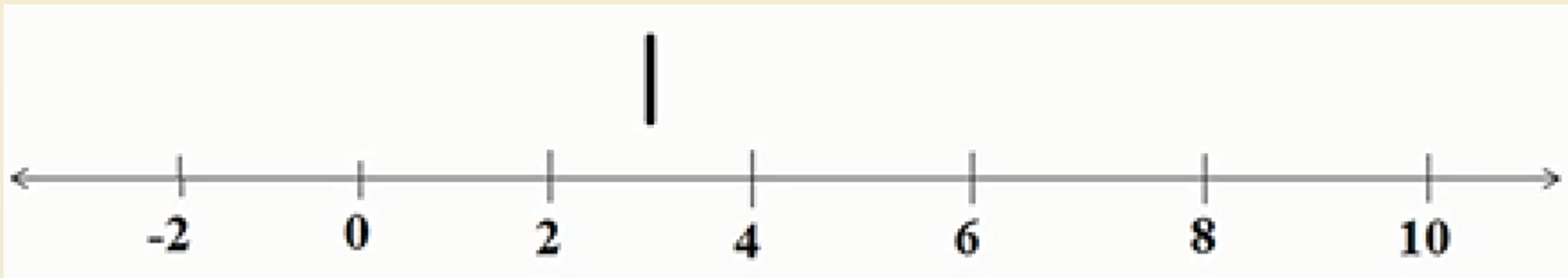


# Verilerin Dağılımının Ölçülmesi

## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

Min= 1   **Q1= 3**   Median= 5   Q3= 7   Max= 9



# Verilerin Dağılımının Ölçülmesi

## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

Min=1 **Q1=3** Median=5 **Q3=7** Max=9

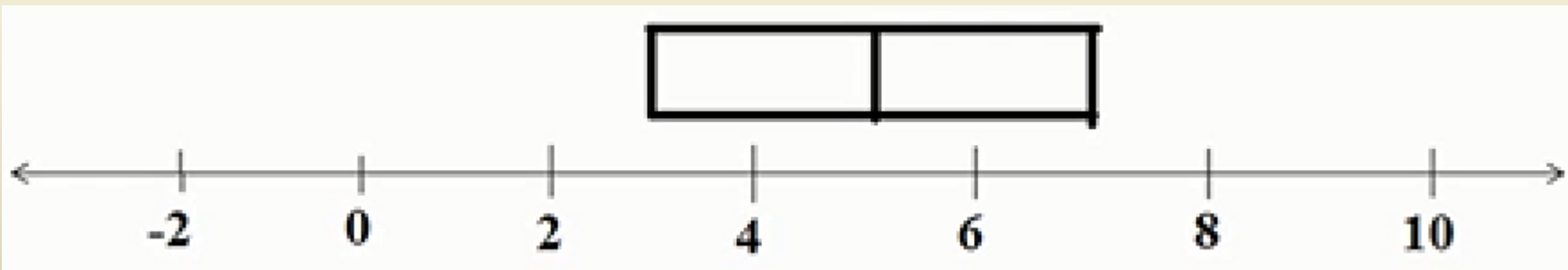
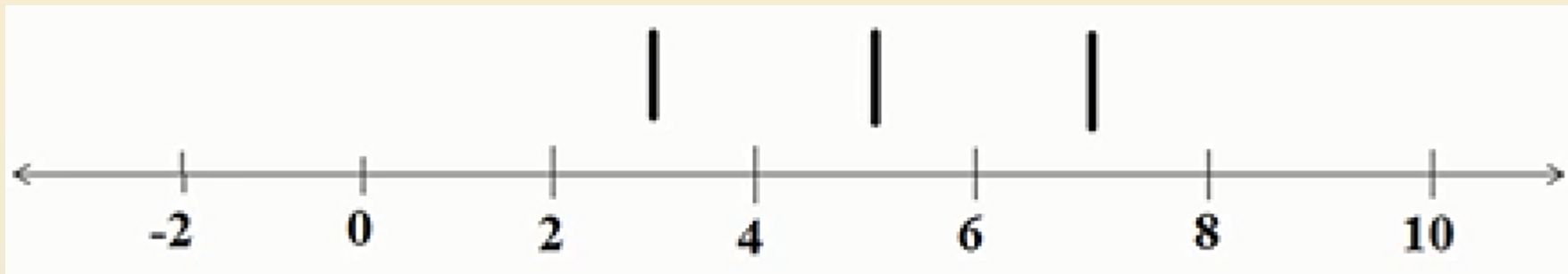


# Verilerin Dağılımının Ölçülmesi

## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

Min=1   **Q1=3**   **Median=5**   **Q3=7**   Max=9



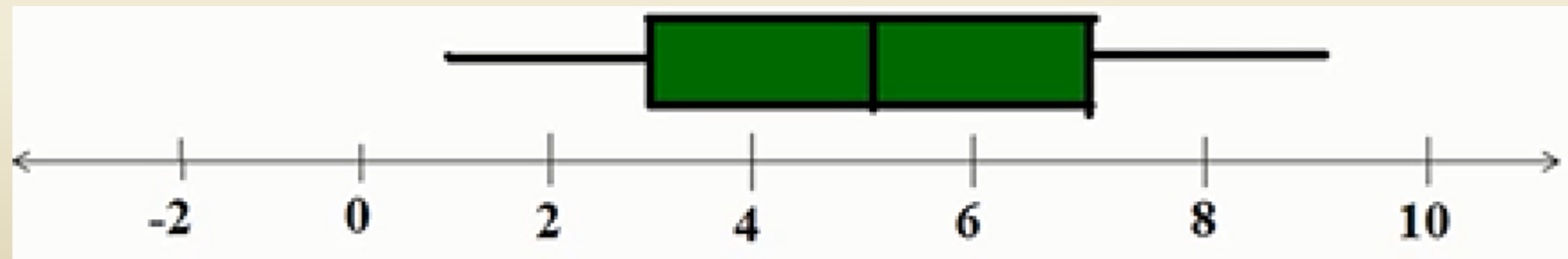
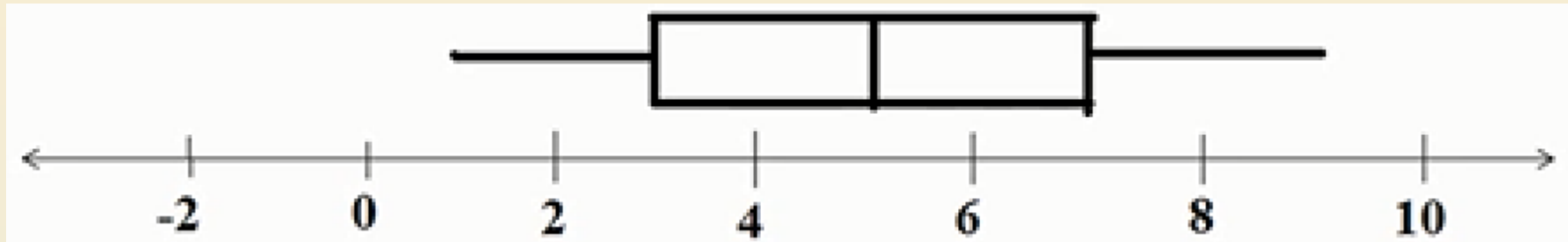


# Verilerin Dağılımının Ölçülmesi

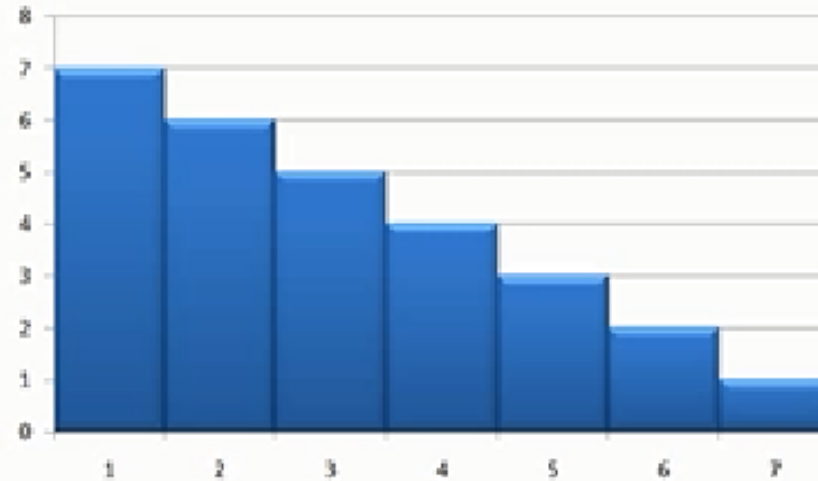
## Kutu Grafiği Çizimi:

Veri kümesi: 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 6, 7, 8, 8, 9

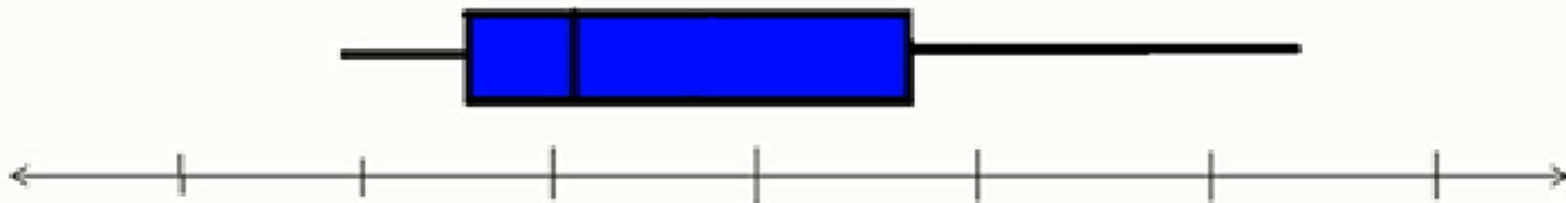
Min=1   Q1=3   Median=5   Q3=7   Max=9



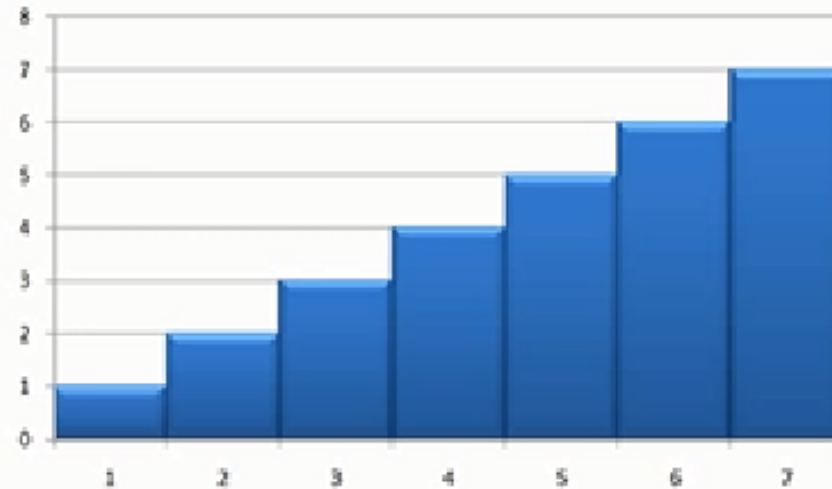
# Verilerin Dağılımının Ölçülmesi



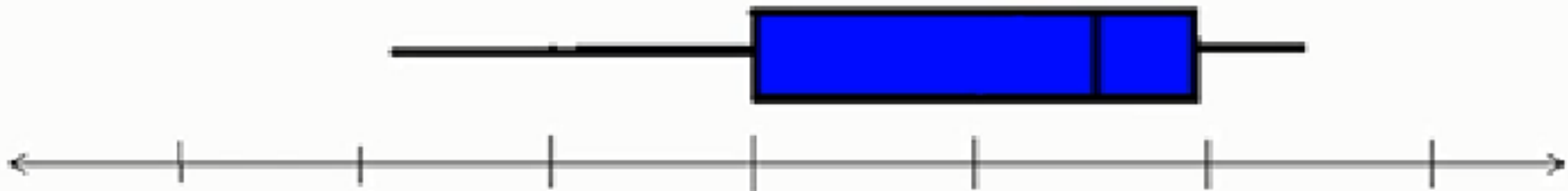
This distribution is **skewed to the right**. The **mean** is pulled to the left from the center.



# Verilerin Dağılımının Ölçülmesi



**This distribution is *skewed to the left*. The *mean* is pulled to the right from the center.**



# Verilerin Dağılımının Ölçülmesi

---

Örnek:

Veri kümesi: 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 7, 8, 8, 9, 10, 10, 11, 11

Veri kümesi için kutu grafiğini çiziniz!

# Verilerin Dağılımının Ölçülmesi

Veri kümesi: 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 6, 7, 8, 8, 9, 10, 10, 11, 11

1. Minimum: 2

2. Birinci Çeyrek ( $Q_1$ ): 4

$$Q1 (\%25) = (25/100) * (19+1) = 5. \text{ değer}$$

3. Medyan: 6

4. Üçüncü Çeyrek ( $Q_3$ ): 9

$$Q3 (\%75) = (75/100) * (19+1) = 15. \text{ değer}$$

5. Maksimum: 11

# Özet

---

Bu ders kapsamında:

- ❑ Verinin özetlenmesi ve görselleştirilmesi ile ilgili temel kavramlar incelenmiştir.
- ❑ Veri dağılımının görsel olarak yorumlanması yapılmıştır.
- ❑ Kutu grafiği analizi yapılmıştır.