

# **Olasılık ve İstatistik - 3. Hafta**

## **Betimsel İstatistik II**

Prof. Dr. Rüya ŞAMLI

# Dağılım Ölçüleri

- Varyans: Verilerin ortalamadan sapması
- Standart Sapma: Varyansın karekökü
- Değişim Katsayısı (CV):  $\sigma / \bar{x} \times 100\%$

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**
- Varyans, verilerin ortalamadan ne kadar uzaklaştığını ölçer. Daha büyük varyans → daha fazla dağılma; küçük varyans → veriler ortalama etrafında toplanmış demektir.

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**

## Örnek 1: Öğrenci sınav notları

- Sınıf A: [70, 72, 71, 69, 68] → Ortalama = 70, varyans düşük → öğrenciler birbirine yakın seviyede.
- Sınıf B: [40, 55, 70, 85, 100] → Ortalama = 70, varyans yüksek → bazıları çok düşük, bazıları çok yüksek not almış.

# Dağılım Ölçüleri

- **Varyans:** Verilerin ortalamadan sapması

## Örnek 2: Günlük sıcaklıklar

- Şehir X (ılıman iklim): [20, 21, 19, 22, 20] → Ortalama  $\approx 20$ , varyans küçük → hava stabil.
- Şehir Y (çöl iklimi): [10, 40, 15, 35, 5] → Ortalama  $\approx 21$ , ama varyans çok büyük → gündüz/gece farkı çok yüksek.

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**

## Örnek 3: Fabrika üretim hatları

- Hat 1: [100, 102, 98, 101, 99] → Ortalama  $\approx 100$ , varyans düşük → üretim istikrarlı.
- Hat 2: [50, 120, 80, 130, 70] → Ortalama  $\approx 90$ , varyans yüksek → üretim kalitesinde büyük dalgalanmalar.

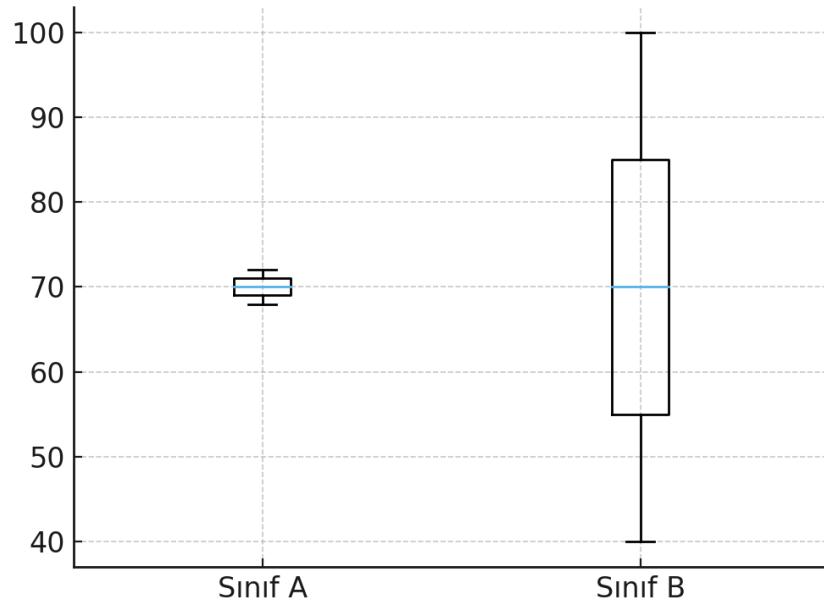
# Dağılım Ölçüleri

- **Varyans:** Verilerin ortalamadan sapması

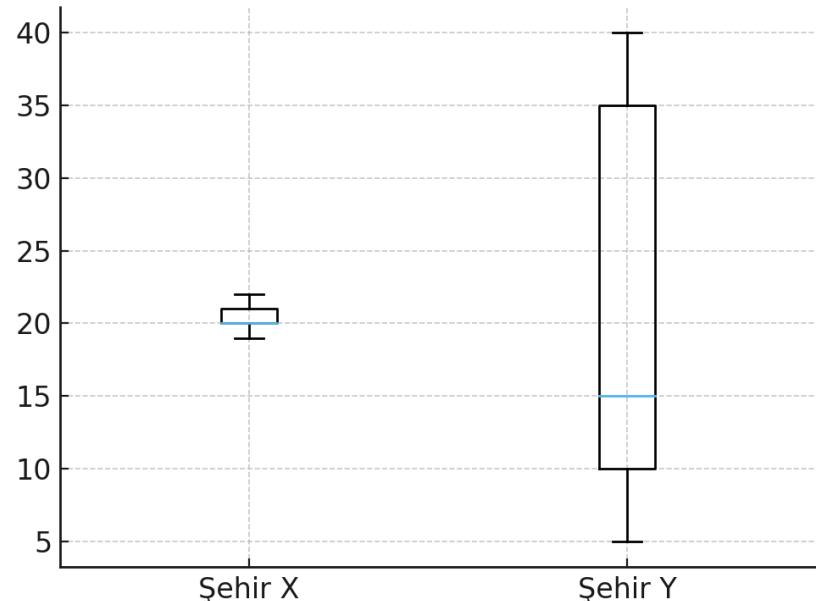
## Örnek 4: Sporcuların koşu süreleri

- Grup A: [12,1 – 12,3 – 12,2 – 12,4 – 12,3] saniye → Varyans düşük, performanslar benzer.
- Grup B: [10,5 – 14,0 – 11,0 – 15,5 – 9,5] saniye → Varyans yüksek, bazıları çok hızlı bazıları çok yavaş.

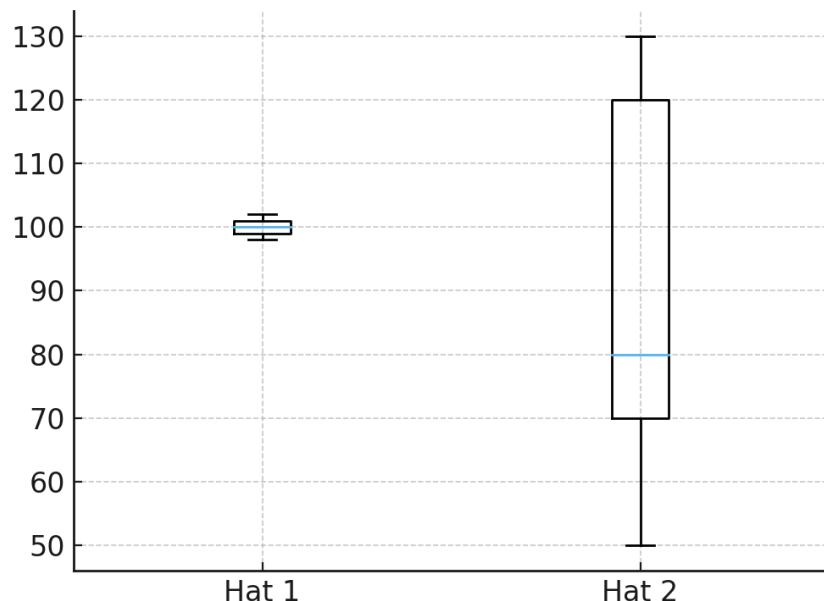
### Sınav Notları



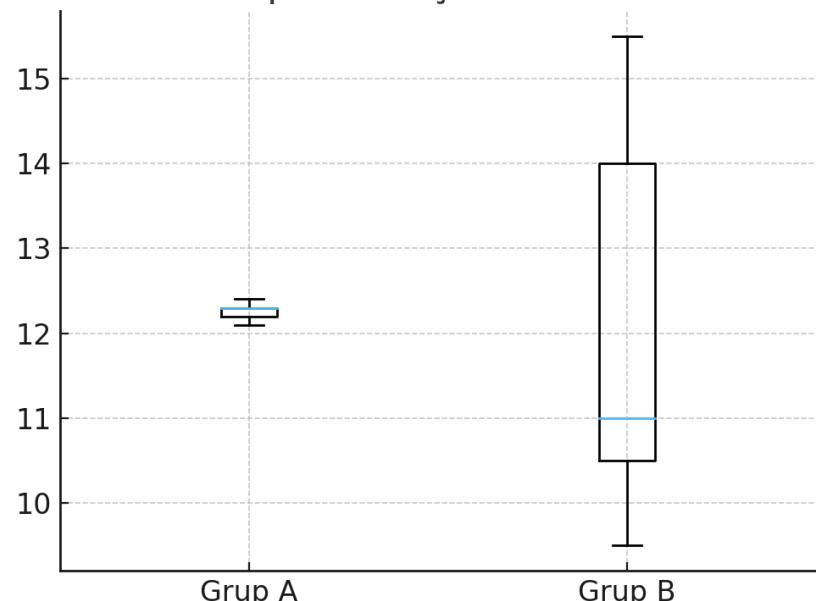
### Günlük Sıcaklıklar



### Fabrika Üretim Hatları



### Sporcu Koşu Süreleri



# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**
- **Sınav Notları:** Sınıf A öğrencilerinin notları birbirine yakın (küçük varyans), Sınıf B’de fark çok büyük (yüksek varyans).
- **Günlük Sıcaklıklar:** Şehir X sabit sıcaklıklara sahip, Şehir Y’de gece-gündüz farkı büyük.
- **Fabrika Üretim Hatları:** Hat 1 istikrarlı üretim yapıyor, Hat 2’de ciddi dalgalanmalar var.
- **Sporcu Koşu Süreleri:** Grup A benzer sürelerde koşarken, Grup B’de çok hızlı ve çok yavaş sporcular var.

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**

## Örnek 1: Küçük varyans

Veriler: [49, 50, 51]

- Ortalama = 50
- Her değer ortalamaya çok yakın.
- Varyans küçük → dağılım dar.

## Örnek 2: Büyük varyans

Veriler: [10, 50, 90]

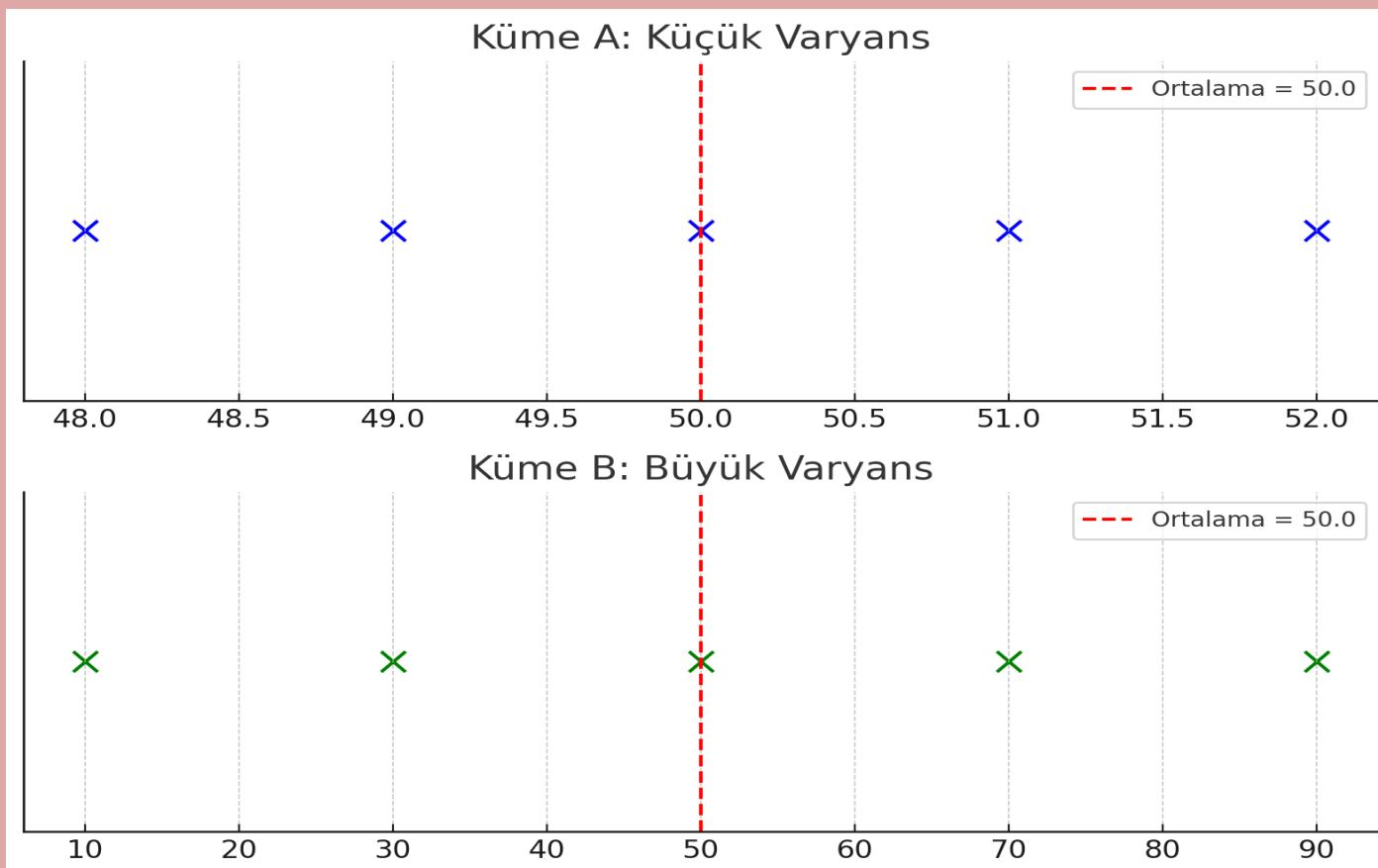
- Ortalama = 50
- Değerlerden ikisi ortalamadan çok uzak.
- Varyans büyük → dağılım geniş.

## Örnek 3: Aynı ortalama, farklı varyans

- Küme A: [48, 49, 50, 51, 52] → Ortalama = 50, varyans küçük.
- Küme B: [10, 30, 50, 70, 90] → Ortalama yine 50, ama varyans çok daha büyük.

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**
- **Küme A (küçük varyans):** Veriler ortalamanın (50) hemen yanında toplanmış.
- **Küme B (büyük varyans):** Ortalaması aynı (50), fakat değerler ortalamadan çok uzaklara dağılmış.



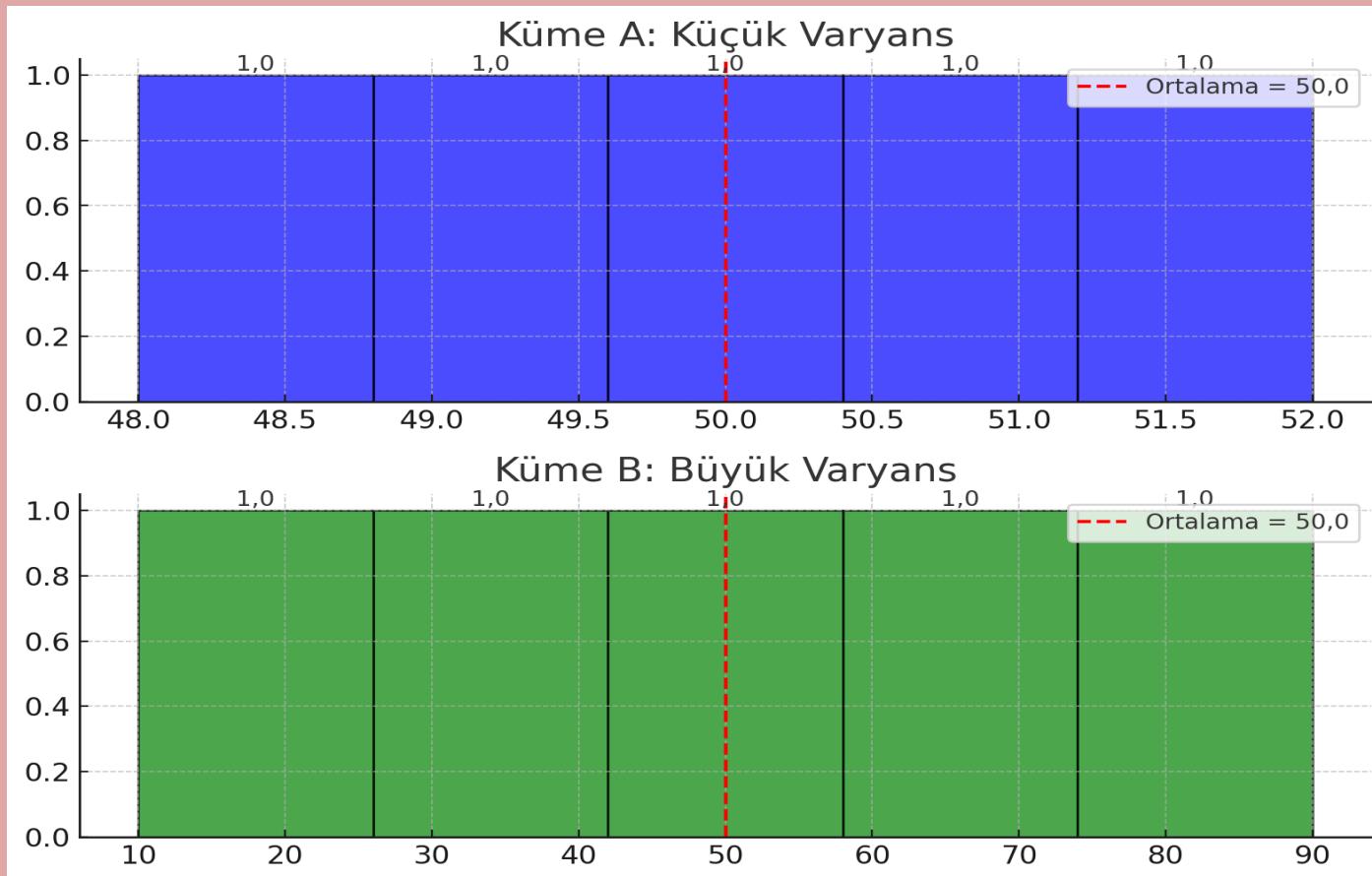
# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**
- **Küme A (48, 49, 50, 51, 52)**
  - Varyans = **2,0**
  - Standart Sapma = **1,41**
- **Küme B (10, 30, 50, 70, 90)**
  - Varyans = **800,0**
  - Standart Sapma = **28,28**

İki kümenin ortalaması aynı (50), fakat varyans/standart sapma çok farklı.  
Bu da Küme B'nin verilerinin çok daha geniş aralığa yayıldığını gösteriyor.

# Dağılım Ölçüleri

- **Varyans: Verilerin ortalamadan sapması**
- **Küme A:** Ortalamaya yakın dar dağılım
- **Küme B:** Ortalaması aynı ama çok daha geniş dağılım



# Dağılım Ölçüleri

- **Standart Sapma: Varyansın karekökü**

**Standart sapma**, varyansın kareköküdür ve verilerin ortalama etrafında ne kadar dağıldığını ölçer.

## Temel Özellikler

- **Küçük standart sapma:** Veriler ortalamaya yakın, dağılım dar.
- **Büyük standart sapma:** Veriler ortalamadan uzak, dağılım geniş.
- Birimi, verinin ölçü birimiyle aynıdır (örneğin notlar için "puan", sıcaklık için " $^{\circ}\text{C}$ ").
- Varyansa göre daha yorumlanabilir, çünkü değerlerle aynı ölçekten gelir.

# Dağılım Ölçüleri

- **Standart Sapma: Varyansın karekökü**

## Örnek

- Notlar: [50, 50, 50, 50, 50]
  - Ortalama = 50
  - Standart sapma = 0 → herkes aynı notu almış.
- Notlar: [0, 50, 100]
  - Ortalama = 50
  - Standart sapma  $\approx 40,8$  → notlar ortalamadan çok uzaklara dağılmış.

# Dağılım Ölçüleri

- **Standart Sapma: Varyansın karekökü**

## Günlük hayattan benzetme

- **Küçük standart sapma:** Bir fabrikanın ürettiği vidaların boyu hep aynı (standart üretim).
- **Büyük standart sapma:** Aynı fabrikanın ürettiği vidaların boyları birbirinden çok farklı (düzensiz üretim).

# Dağılım Ölçüleri

- **Değişim Katsayısı (CV):**  $\sigma / \bar{x} \times 100\%$
- CV, standart sapmanın ortalamaya oranıdır.
- Verilerin **göreli değişkenliğini** ölçer.
- Farklı ortalamalara sahip veri setlerini karşılaştırmada kullanılır.
- Boyutsuz bir ölçüdür (% cinsinden ifade edilir).
- **Küçük CV:** Veriler ortalamaya göre daha homojen.
- **Büyük CV:** Veriler ortalamaya göre daha değişken.

# Dağılım Ölçüleri

- Değişim Katsayısı (CV):  $\sigma / \bar{x} \times 100\%$

1. Sınıf A → Ortalama = 80,  $\sigma$  = 4

$$CV = \frac{4}{80} \times 100 = 5\%$$

→ Homojen başarı.

2. Sınıf B → Ortalama = 40,  $\sigma$  = 8

$$CV = \frac{8}{40} \times 100 = 20\%$$

→ Daha değişken başarı.

# Dağılım Ölçüleri

- **Değişim Katsayısı (CV):**  $\sigma / \bar{x} \times 100\%$

## Nerede Kullanılır?

- Finans → Yatırımin riskini ölçmek (düşük CV = daha istikrarlı yatırım).
- Üretim → Kalite kontrol (düşük CV = daha standart ürün).
- Bilimsel araştırmalar → Farklı büyüklükteki ölçümleri karşılaştırmak.

# Çarpıklık ve Basıklık

- Çarpıklık (Skewness): Dağılımin simetrisi
  - - Pozitif: Sağ kuyruğu uzun
  - - Negatif: Sol kuyruğu uzun
- Basıklık (Kurtosis): Sivrilik / yayvanlık ölçüsü

# Çarpıklık ve Basıklık

- **Çarpıklık (Skewness): Dağılımın simetrisi**

## Çarpıklık Türleri

### 1.Simetrik Dağılım (Çarpıklık = 0)

1. Ortalama, medyan ve mod birbirine eşittir.
2. Örnek: Normal dağılım (çan eğrisi).

### 2.Sağ Çarpık (Pozitif Çarpıklık, $> 0$ )

1. Kuyruk sağ tarafa doğru uzundur.
2. Ortalama  $>$  Medyan  $>$  Mod
3. Örnek: Gelir dağılımları (çoğu insan düşük-orta gelirde, az sayıda çok yüksek gelirli).

### 3.Sola Çarpık (Negatif Çarpıklık, $< 0$ )

1. Kuyruk sol tarafa doğru uzundur.
2. Ortalama  $<$  Medyan  $<$  Mod
3. Örnek: Emeklilik yaşı (çoğu insan belli bir yaştan, az sayıda daha erken).

# Çarpıklık ve Basıklık

- **Çarpıklık (Skewness): Dağılımın simetrisi**

## Çarpıklık Değerlerinin Yorumu

- $\approx 0 \rightarrow$  Simetrik dağılım.
- $0$  ile  $+1$  arası  $\rightarrow$  Hafif sağa çarpık.
- $> +1 \rightarrow$  Kuvvetli sağa çarpık.
- $0$  ile  $-1$  arası  $\rightarrow$  Hafif sola çarpık.
- $< -1 \rightarrow$  Kuvvetli sola çarpık.

# Çarpıklık ve Basıklık

- **Çarpıklık (Skewness): Dağılımın simetrisi**

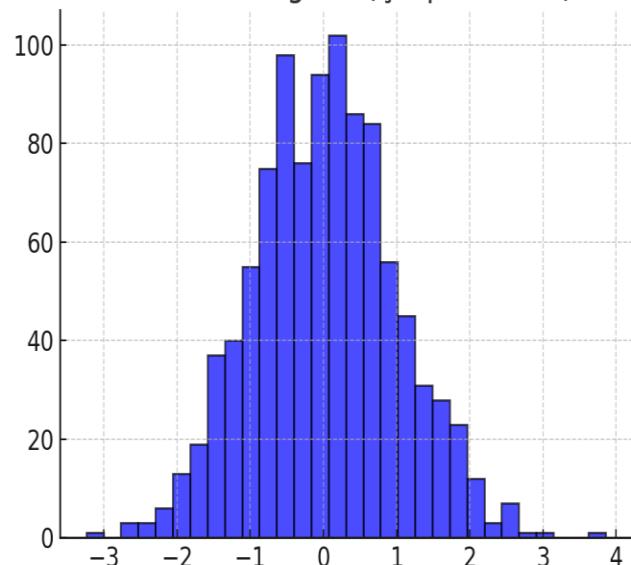
## Günlük Hayattan Örnekler

- **Pozitif çarpıklık:** İnsanların aylık maaşları.
- **Negatif çarpıklık:** Çok kolay bir sınavdan alınan notlar (çoğu yüksek, birkaç düşük).

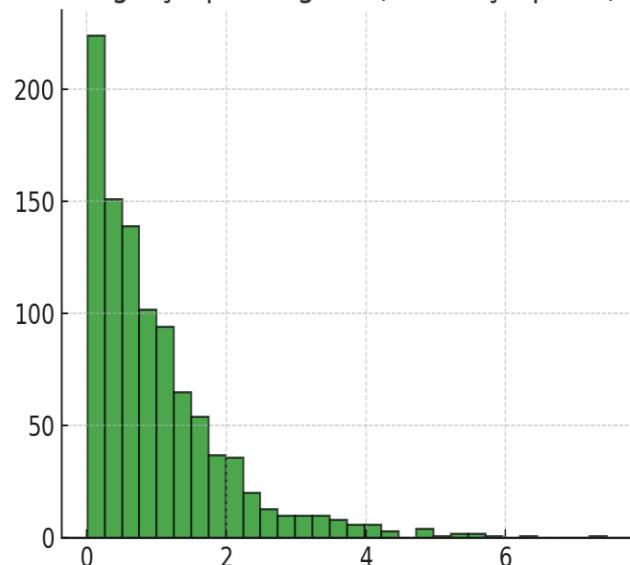
# Çarpıklık ve Basıklık

- Çarpıklık (Skewness): Dağılımın simetrisi

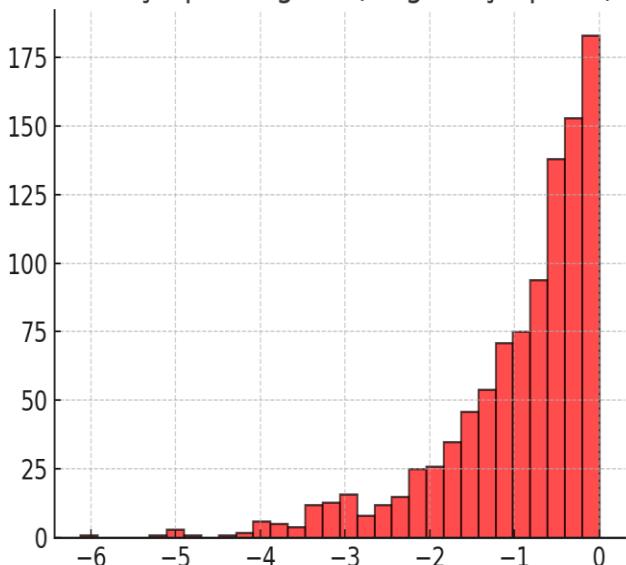
Simetrik Dağılım (Çarpıklık  $\approx 0$ )



Sağ Çarpık Dağılım (Pozitif Çarpıklık)



Sola Çarpık Dağılım (Negatif Çarpıklık)



# Çarpıklık ve Basıklık

- **Çarpıklık (Skewness): Dağılımın simetrisi**
- **Simetrik dağılım ( $\approx 0$ )** → Ortalama, medyan ve mod aynı yerde.
- **Sağ çarlık dağılım (pozitif çarpıklık)** → Kuyruk sağa doğru uzuyor, ortalama  $>$  medyan  $>$  mod.
- **Sola çarlık dağılım (negatif çarpıklık)** → Kuyruk sola doğru uzuyor, ortalama  $<$  medyan  $<$  mod.

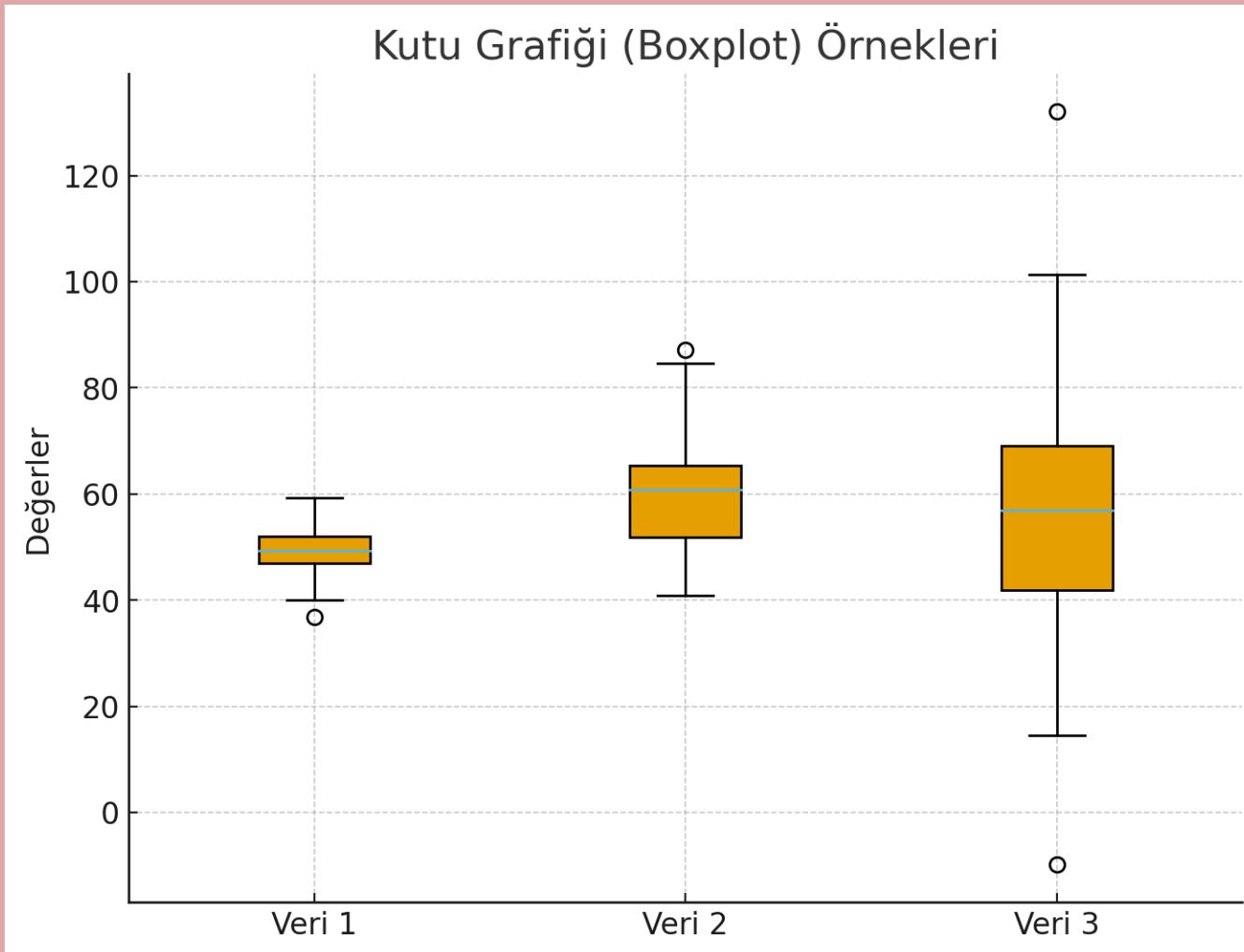
# Kutu Grafiği (Boxplot)

- Minimum, Q1, Medyan, Q3, Maksimum
- Aykırı değerleri gösterir
- Veri dağılımını özetlemede kullanılır

# Kutu Grafiği (Boxplot)

- **Kutu Grafiğinin Elemanları**
- **Medyan (Q2)**
  - Verilerin ortadaki değeri.
  - Kutu içindeki yatay çizgi ile gösterilir.
- **Alt Çeyrek (Q1)**
  - Verilerin %25'inin altında olduğu değer.
- **Üst Çeyrek (Q3)**
  - Verilerin %75'inin altında olduğu değer.
- **Kutu (IQR – Interquartile Range)**
  - Q1 ile Q3 arasındaki alan.
  - Veri setinin ortadaki %50'sini kapsar.
  - Kutu ne kadar küçükse, veriler o kadar yoğundur.
- **Bıyıklar (Whiskers)**
  - $Q1 - 1,5 \times IQR$  ile  $Q3 + 1,5 \times IQR$  aralığını gösterir.
  - Normal değerlerin yayıldığı alan.
- **Uç Değerler (Outliers)**
  - Bıyıkların dışına düşen noktalar.
  - Veri setinde istisnai gözlemleri belirtir.

# Kutu Grafiği (Boxplot)



# Kutu Grafiği (Boxplot)

- **Veri 1:** Ortalama etrafında dar dağılım → kutu küçük, bıyıklar kısa.
- **Veri 2:** Daha geniş dağılım → kutu ve bıyıklar daha uzun, birkaç uç değer var.
- **Veri 3:** Çok geniş dağılım → büyük kutu, uzun bıyıklar ve çok sayıda uç değer görülmüyor.

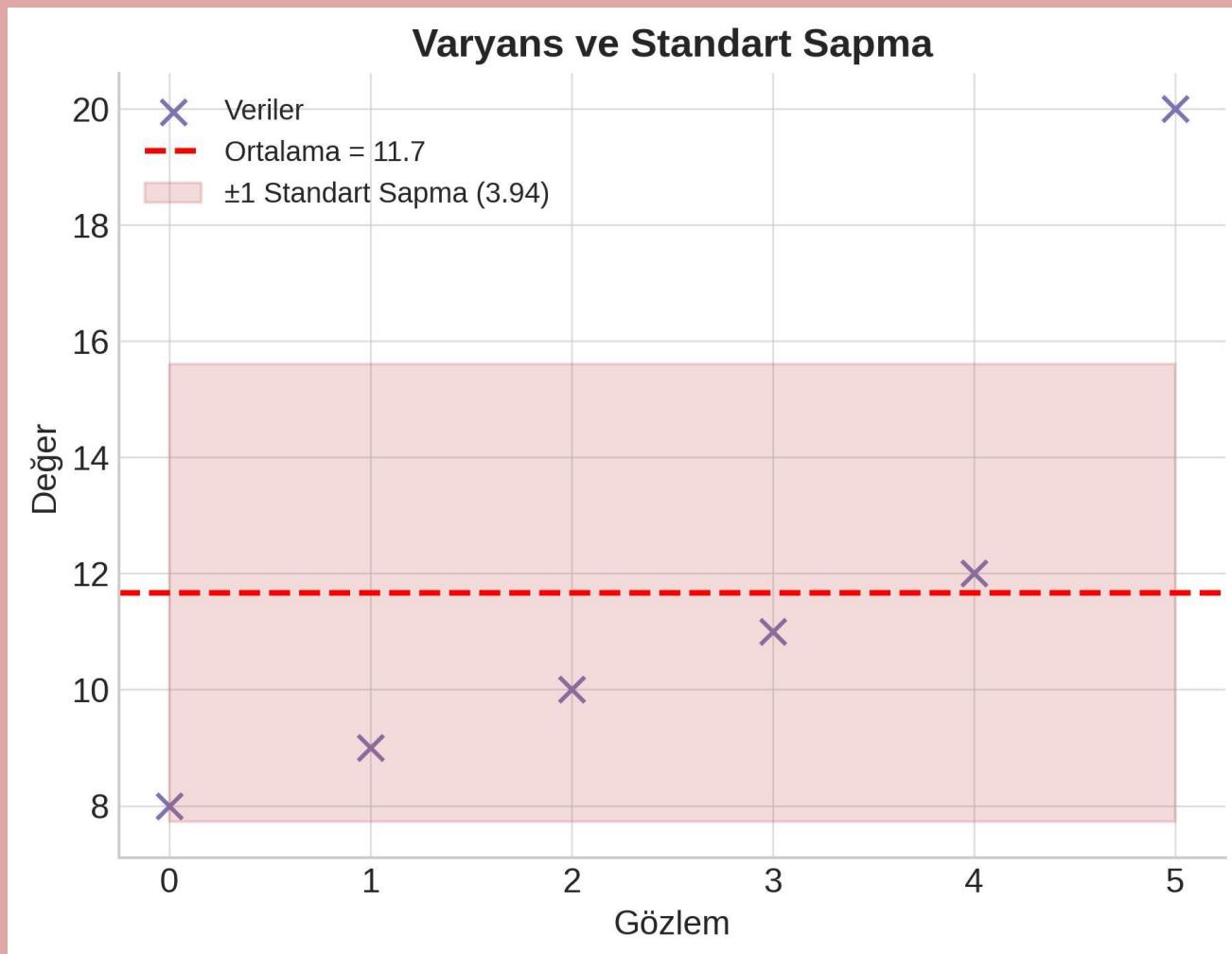
# Örnek Problemler

- 1. Günler: 8, 9, 10, 11, 12, 20
  - Ortalama, standart sapma, CV hesaplayın.
- 2. Veriler: 5, 6, 6, 7, 8, 10, 15, 20, 21
  - Kutu grafiğini çiziniz.

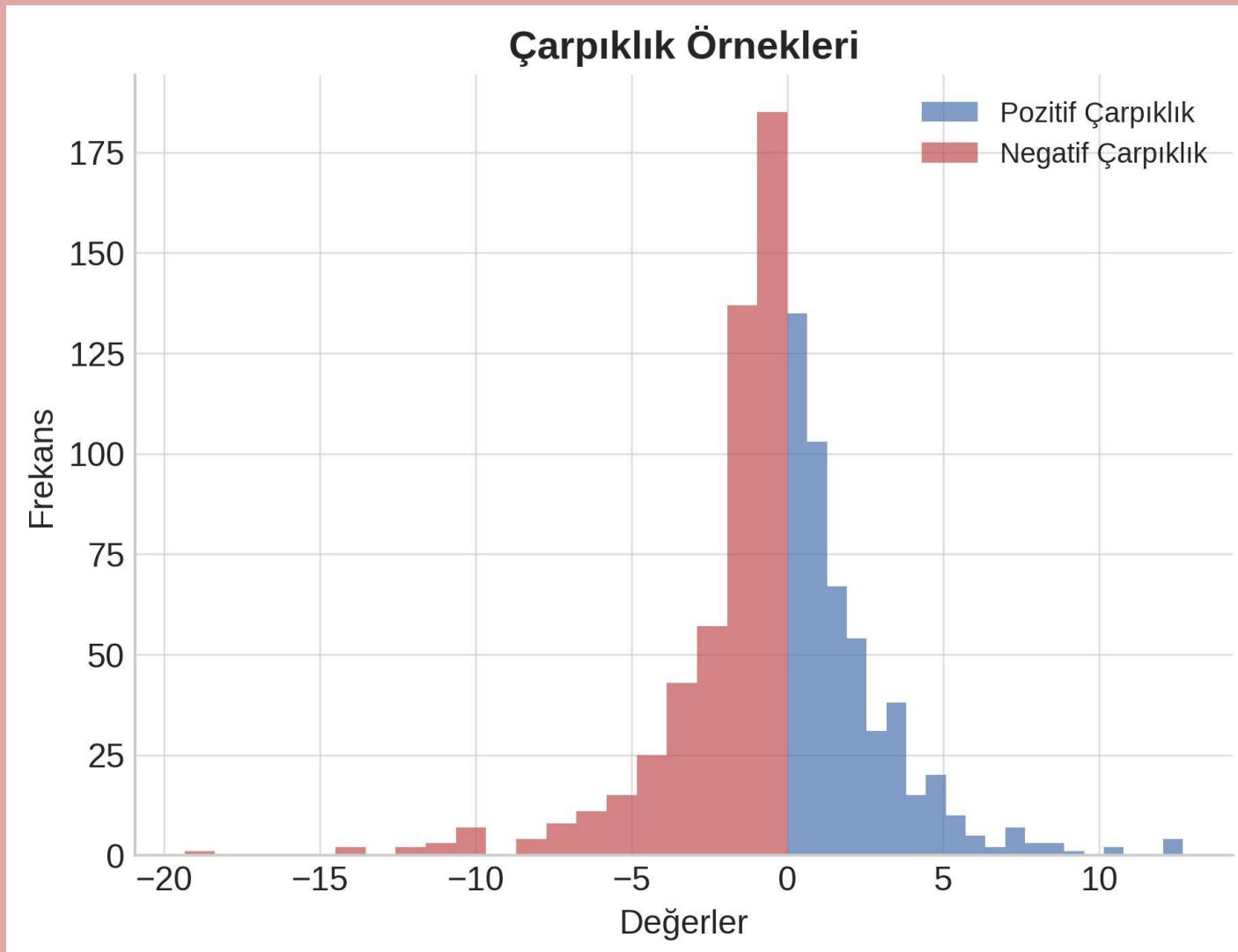
# Küçük Ödev

- 1. Küçük bir veri seti bulun (örn. yaş, maaş).
  - Ortalama, varyans, standart sapma hesaplayın
  - Çarpıklık ve basıklık yorumu yapın
  - Boxplot çizin
- 2. CV değerini hesaplayın ve karşılaştırın

# Varyans ve Standart Sapma



# Çarpıklık Örnekleri



# Boxplot Örneği

**Kutu Grafiği (Boxplot)**

