İstatistik 1
BÖLÜM 3
VERİ SETLERİNİN ÖZETLENMESİNDE KULLANILAN SAYISAL YÖNTEMLER

2017-2018 Ankara Üniversitesi, SBF Onur Özsoy

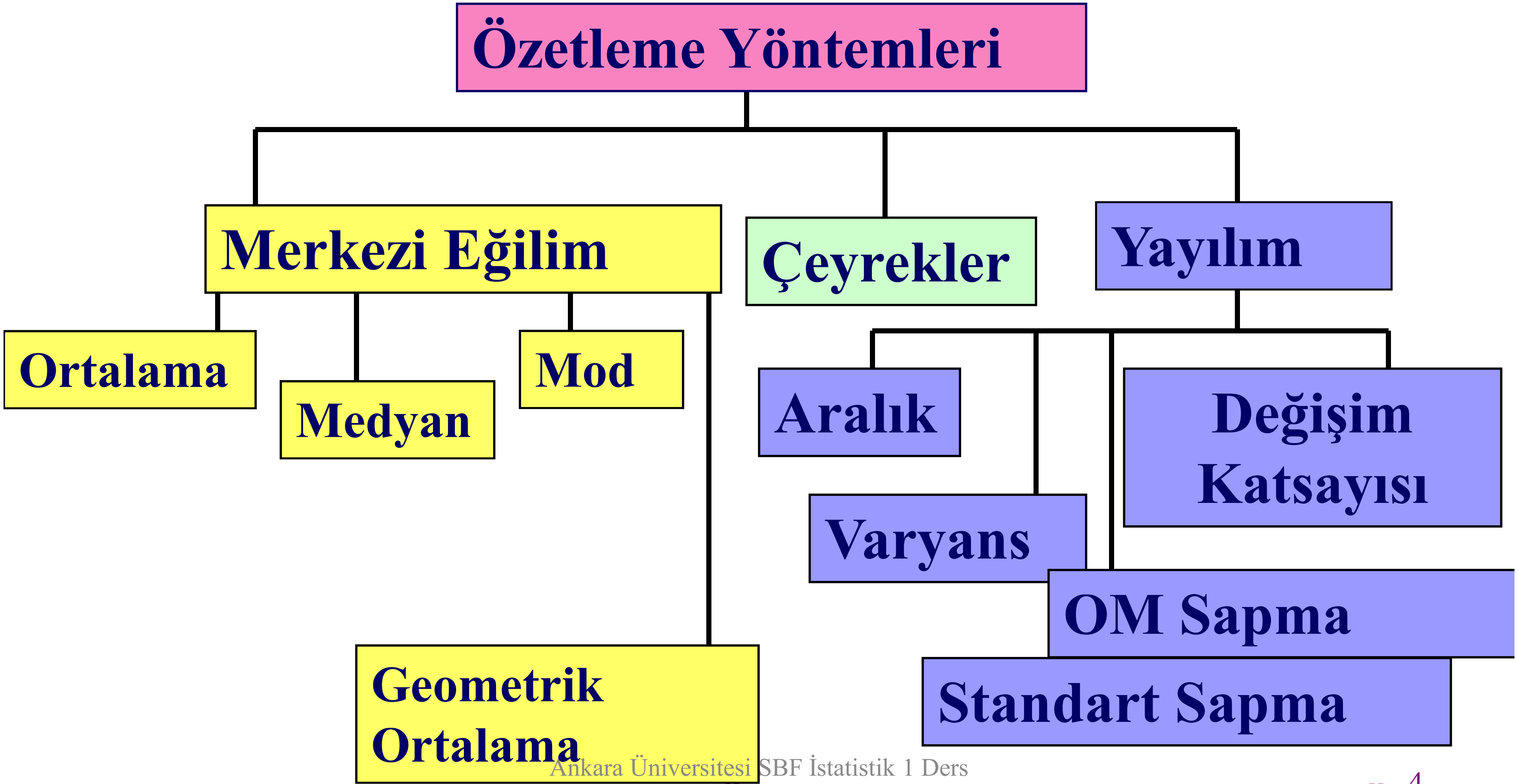
İşlenecek Konular

- Merkezi Eğilim Ölçüleri
  - Ortalama, medyan, mod, geometrik ortalama ve orta-aralık
- çeyrekler
- Merkezi Yayılım Ölçüleri
  - Açıklık, çeyrekler arası açıklık, varyans, standart sapma, ortalama mutlak sapma, ve değişim katsayısı
- Dağılımların Şekilleri
  - simetrik, çarpık, box-ve-whisker grafikleri

İşlenecek Konular

- Korelasyon Katsayısı
- Sayısal yöntemlerin kullanımı ile istatistiksel özetleme yaparken dikkat edilmesi gereken önemli noktalar

Özetleme Yöntemleri



Merkezi Eğilim Ölçüleri

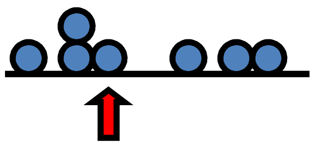
Merkezi Eğilim

- **Ortalama**

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

- **Medyan**



- **Mod**
- **Geometrik Ortalama**

$$\bar{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

Ortalama (Aritmetik)

- Ortalama
  - Örnek Ortalaması

Örnek gözlem sayısı

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

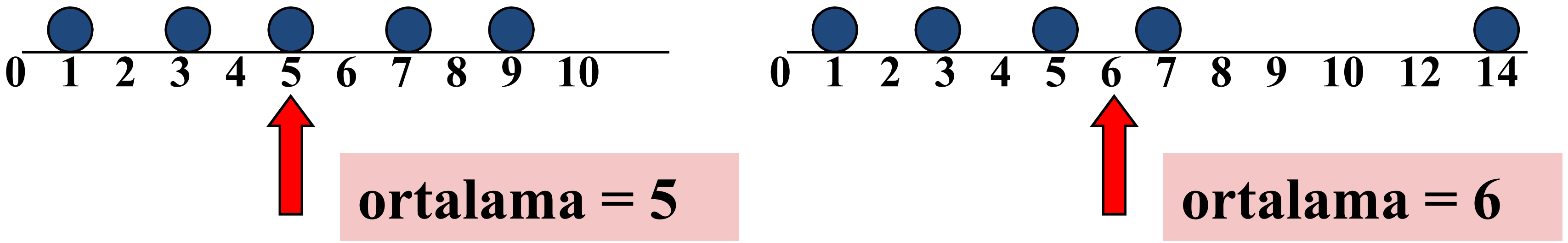
  - Populasyon Ortalam

Populasyon gözlem sayısı

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

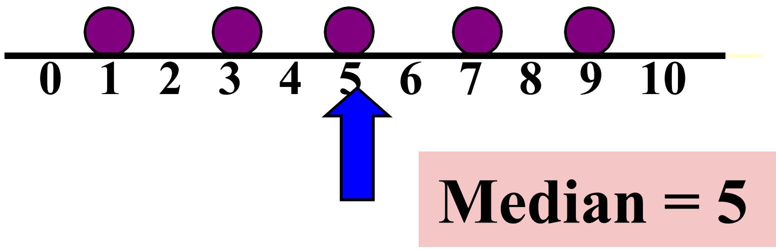
Ortalama

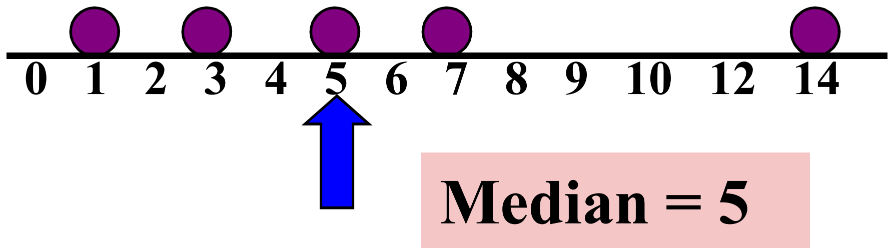
- En yaygın kullanılan merkezi eğilim ölçüsüdür.
- Veri setinde aşırı uçlar varsa bu ölçü etkilenir.



Medyan

- Aşırı uç değerler etkilemez.
- Daha sağlam bir ölçüdür.

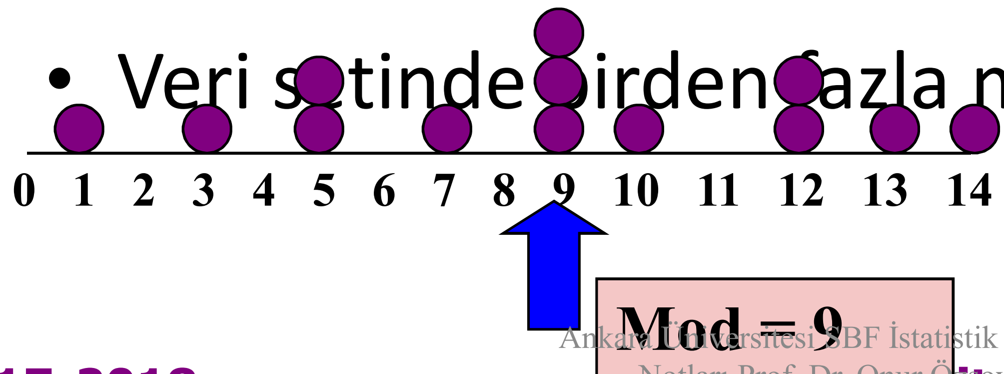


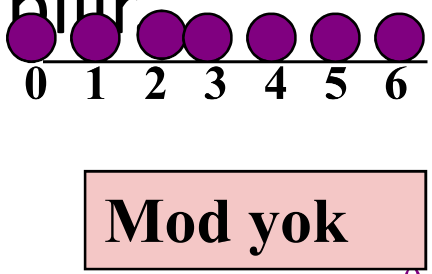


- Gözlem değerleri küçükten büyüğe doğru sıralandığında Medyan ortada kalan gözlemdir.
  - N yada n tek sayı ise ortada kalan gözlem
  - N yada n çift sayı ise ortadaki iki gözlemin ortalaması

Mod

- Başka bir merkezi eğilim ölçüsüdür.
- En sık gözlemlenen değerdir.
- Aşırı uçlardan etkilenmez.
- Sayılabilen yada sayılamayan veriler için kullanılabilir.
- Veri setinde bazen mod olmayabilir.
- Veri setinde birden fazla mod olabilir.





Geometrik Ortalama

- Bir değişkende zaman içinde meydana gelen değişim oranını hesaplamada kullanılır.

- Zaman içinde bir yatırımın getirisini hesaplar

$$\bar{X}_G = \left(X_1 \times X_2 \times \cdots \times X_n\right)^{1/n}$$

  – Yatırımın zaman içinde durumunun ne olduğunu hesaplar.

$$\bar{R}_G = \left[\left(1+R_1\right)\times\left(1+R_2\right)\times\cdots\times\left(1+R_n\right)\right]^{1/n} - 1$$

Örnek

$100,000 bir yatırımın değeri 1. sene sonunda $50,000 düşmüş ve ikinci yılın sonunda tekrar $100,000 yükselmiştir. Ortalama getiri ve geometrik ortalama nedir?

$$X_1 = \$100,000 \qquad X_2 = \$50,000 \qquad X_3 = \$100,000$$

Ortalama getiri

$$\bar{X} = \frac{(-50\%) + (100\%)}{2} = 25\%$$

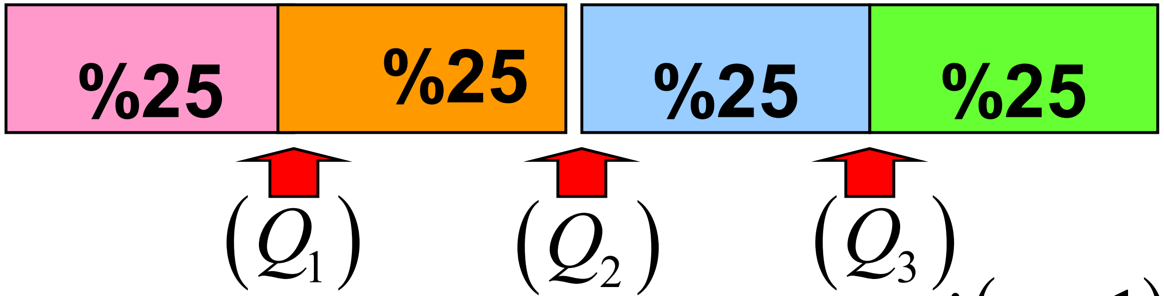
Geometrik getiri

$$\bar{R}_G = \left[\left(1 + (-50\%)\right) \times \left(1 + (100\%)\right)\right]^{1/2} - 1$$

$$= \left[(0.50) \times (2)\right]^{1/2} - 1 = 1^{1/2} - 1 = 0\%$$

Çeyrekler

- Veri setini dörde böler



- İ inci çeyreğin pozisyonu $(Q_i) = \dfrac{i(n+1)}{4}$

Veri seti: küçük...büyük: 11 12 13 16 16 17 18 21 22

Q_1 **inci çeyreğin pozisyonu** $= \dfrac{1(9+1)}{4} = 2.5 \qquad Q_2 = \dfrac{(12+13)}{2} = 12.5$

- Q_1 ve Q_3 merkezi eğilim ölçüsü değil
- Q_2 = Medyan, merkezi eğilim ölçüsüdür.

Çeyrekler

Yüzdeliğin yeri $$Y_y = (n+1)\frac{y}{100}$$

Y_{89} 89. yüzdeliği ifade etmektedir.

Çeyrekler

1. Çeyrek → veri setinde yer alan gözlemlerin yaklaşık %25'ini kapsar
2. Çeyrek → veri setinde yer alan gözlemlerin yaklaşık %50'sini kapsar
3. Çeyrek → veri setinde yer alan gözlemlerin yaklaşık %75'ini kapsar
4. Çeyrek → veri setinde yer alan gözlemlerin yaklaşık %100'ünü kapsar

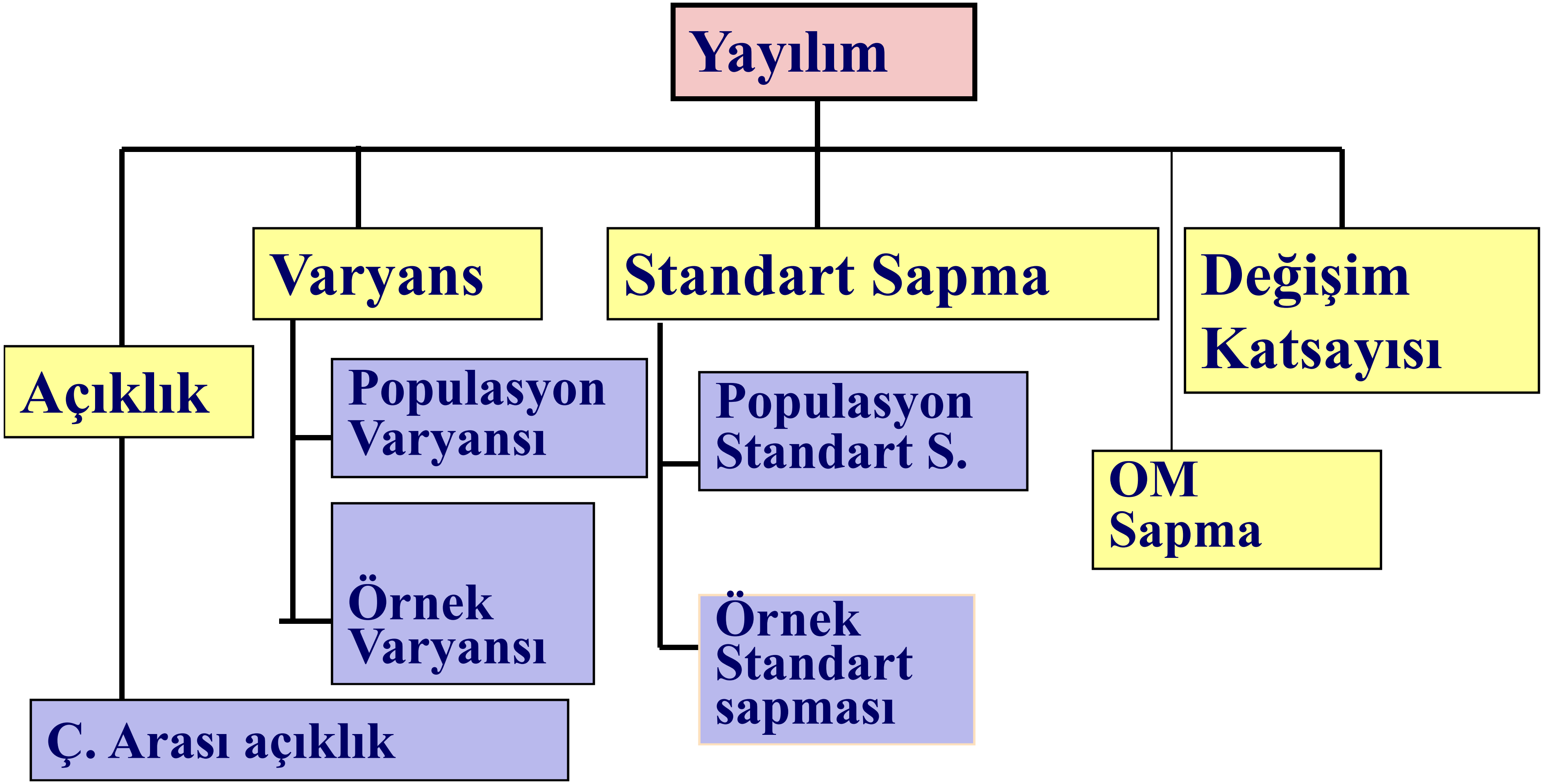
Çeyrekler

1. Çeyrek : $\dfrac{n+1}{4}$ üncü gözlem,

2. Çeyrek: $\dfrac{n+1}{2}$ inci gözlem (bu değer aynı zamanda medyan olarak da bilinmektedir)

3. Çeyrek: 3(n+1)/4'üncü gözlemdir (bu aynı zamanda 75. yüzdelik olarak da bilinmektedir.

Yayılım Ölçüleri



Açıklık

- Bir yayılım ölçüsüdür.
- En büyük gözlem ile en küçük gözlem arasındaki fark:

$$\textbf{Açıklık} = X_{\text{en büyük}} - X_{\text{en küçük}}$$

- Datanın dağılımını göz ardı eder

Açıklık = 12 - 7 = 5

Çeyrekler arası Açıklık

- Yayılım ölçüsüdür.
- Aynı zamanda orta yayılım olarak da bilinir
  - %50'lik yayılım
- 1. Ve 3. Çeyrekler arasındaki farktır.

Data büyükten küçüğe: 11 12 13 16 16 17 17 18 21

Çeyrekler arası açıklık $= Q_3 - Q_1 = 17.5 - 12.5 = 5$

- Aşırı uçlardan etkilenmez

Varyans

- En önemli yayılım ölçüsü
- Ortalamadan sapmaları gösterir-ölçümler
  - Örnek varyansı:

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

  - Populasyon varyansı: $\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \mu\right)^2}{N}$

Standart Sapma

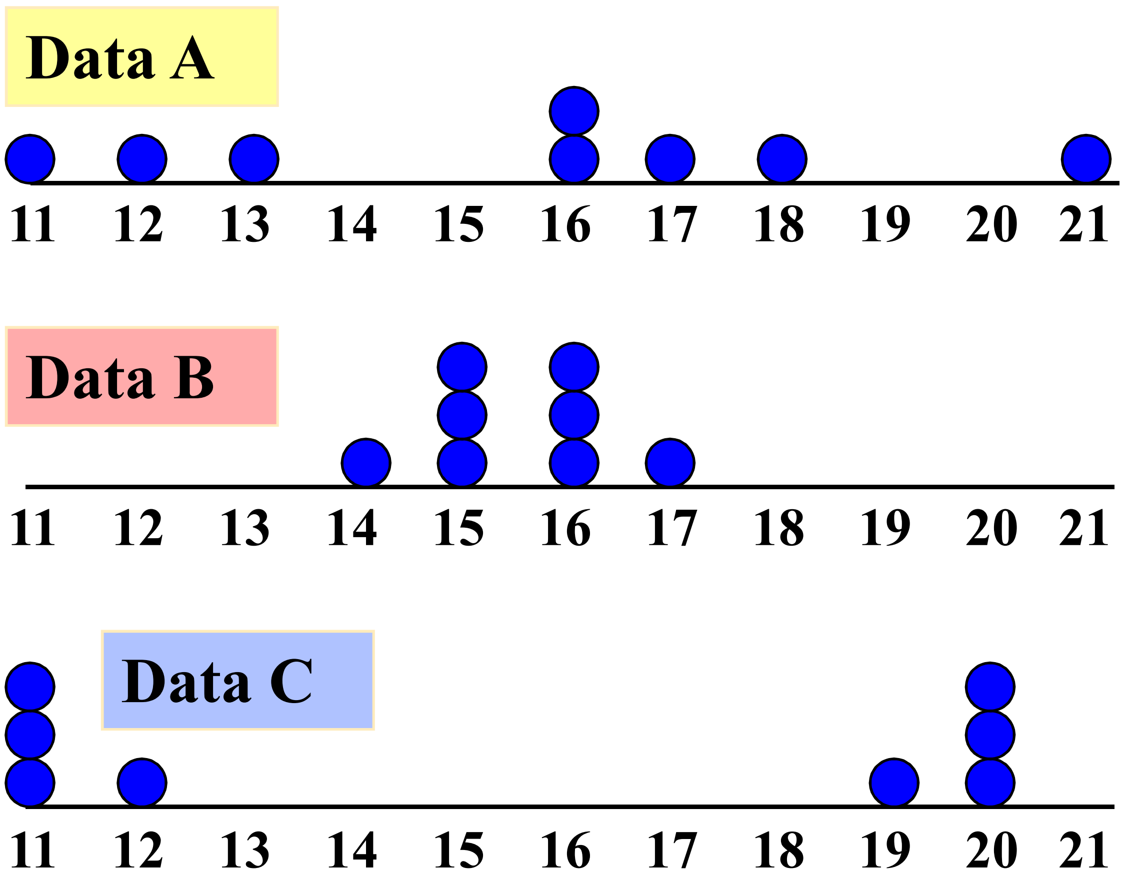
- En önemli yayılım ölçüsü
- Ortalamadan sapmaları gösterir-ölçümler
- Orijinal veri seti ile aynı birim cinsinden
  - Örnek standart sapması:

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

  - Populasyon Standart sapması:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} \left(X_i - \mu\right)^2}{N}}$$

Standart Sapmaların Karşılaştırılması



Data A: Ortalama = 15.5, s = 3.338

Data B: Ortalama = 15.5, s = .9258

Data C: Ortalama = 15.5, s = 4.57

Değişim Katsayısı

- Nisbi yayılımı gösterir-ölçümler
- Her zaman (%) formundadır
- Ortalamaya oranla yayılımı gösterir
- Farklı iki veya daha fazla sayıda veri setinin karşılaştırılmasına yarar
-

$$CV = \left(\frac{S}{\bar{X}} \right) 100\%$$

Değişim Katsayılarının Karşılaştırılması

- Hisse S. A:
  - Geçen yılki ortalama fiyat = \$50
  - Standart sapma = \$5
- Hisse S. B:
  - Geçen yılki ortalama fiyat = \$100
  - Standart sapma = \$5
- Değişim Katsayısı:
  - Hisse S. A: $$CV = \left(\frac{S}{\overline{\overline{X}}}\right)100\% = \left(\frac{\$5}{\$50}\right)100\% = 10\%$$
  - Hisse S. B: $$CV = \left(\frac{S}{\overline{\overline{X}}}\right)100\% = \left(\frac{\$5}{\$100}\right)100\% = 5\%$$

Ortalama Mutlak Sapma (OMS)

- OMS ortalamadan sapmaların mutlak değerlerinin ortalaması
- Yayılım ölçüsü
- Populasyon için

$$\frac{\sum_{i=1}^{N} |x_i - \mu|}{N}$$

- Örnek için

$$\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$