

# مروری بر روش های آمار توصیفی

دکتر یددا... محرابی

دانشکده بهداشت - دانشگاه علوم پزشکی شهید بهشتی

[ymehrabi@gmail.com](mailto:ymehrabi@gmail.com)

---

---

---

---

---

---

---

---

## Numerical Summaries

- **Center of the data**

- mean
- Median

- **Variation**

- range
- quartiles (interquartile range)
- variance
- standard deviation

2

---

---

---

---

---

---

---

---

## Mean or Average

- Traditional measure of center
- Sum the values and divide by the number of values

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

3

---

---

---

---

---

---

---

---

## Median ( $M$ )

- A *resistant measure* of the data's center
- At least half of the **ordered** values are less than or equal to the median value
- At least half of the **ordered** values are greater than or equal to the median value
- If  $n$  is odd, the median is the middle ordered value
- If  $n$  is even, the median is the average of the two middle ordered values

4

---

---

---

---

---

---

---

---

## Median ( $M$ )

Location of the median:  $L(M) = (n+1)/2$ ,  
where  $n$  = sample size.

Example: If 25 data values are recorded, the  
Median would be the  
 $(25+1)/2 = 13^{\text{th}}$  ordered value.

5

---

---

---

---

---

---

---

---

## Median

- Example 1 data: 2 4 6
- Example 2 data: 2 4 6 8
- Example 3 data: 6 2 4

6

---

---

---

---

---

---

---

---

## Comparing the Mean & Median

- The mean and median of data from a symmetric distribution should be close together. The actual (true) mean and median of a symmetric distribution are exactly the same.
- In a skewed distribution, the mean is farther out in the long tail than is the median [the mean is 'pulled' in the direction of the possible outlier(s)].

7

---

---

---

---

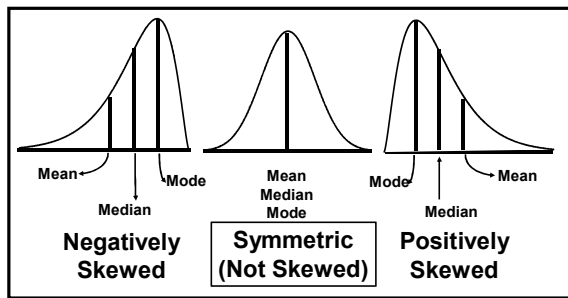
---

---

---

---

## Skewness



8

---

---

---

---

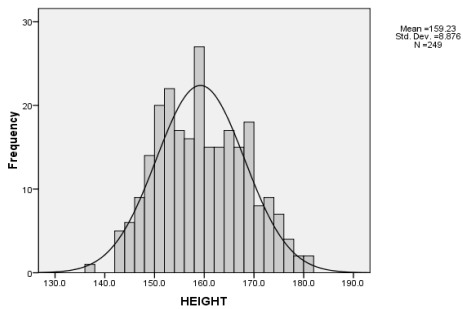
---

---

---

---

HEIGHT



9

---

---

---

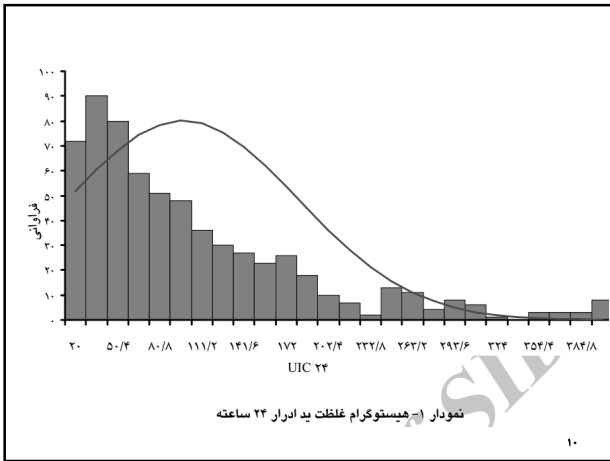
---

---

---

---

---




---



---



---



---



---



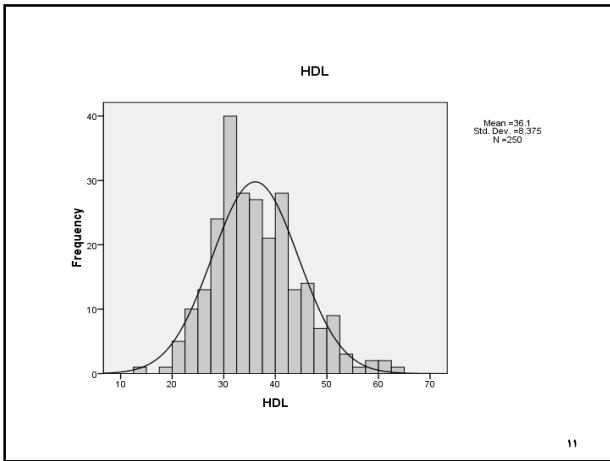
---



---



---




---



---



---



---



---



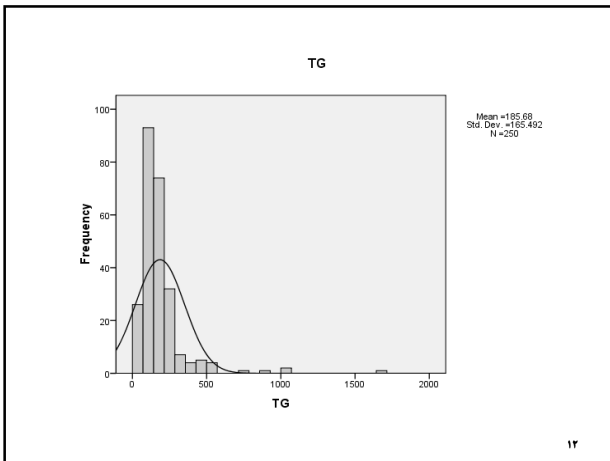
---



---



---




---



---



---



---



---



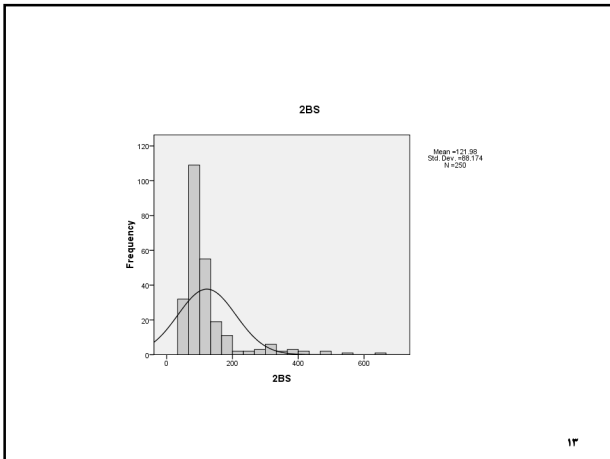
---



---



---




---

---

---

---

---

---

---

---

**مثال**

	Valid	WEIGHT	HEIGHT	HDL	TG	2hrBS
N	250	249	250	250	250	250
Mean		67.5	159.2	36.1	185.6	121.9
Median		67.2	158.5	35.0	148.5	97.0
Std. Deviation		12.40	8.8	8.3	165.5	88.2
Range		63.5	44.0	49	1645	619
Minimum		38.5	137.0	14	43	34
Maximum		102.0	181.0	63	1688	653

14

---

---

---

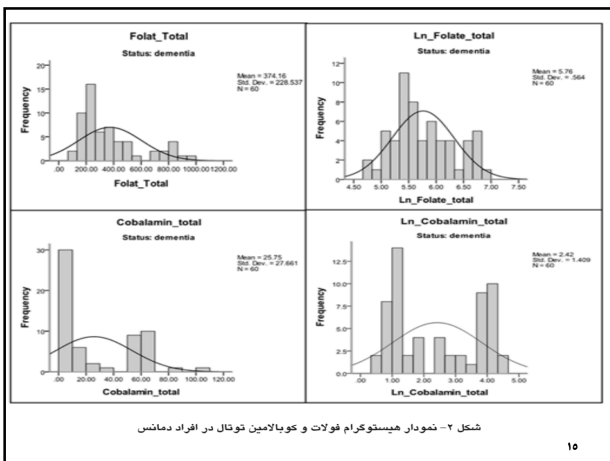
---

---

---

---

---




---

---

---

---

---

---

---

---

## Quantiles چندک ها

Centiles صدک ها

Tentiles دهک ها

Quintiles پنجک ها

Quartiles چارک ها

---

---

---

---

---

---

---

---

## Quartiles

- Three numbers which divide the ordered data into four equal sized groups.
- $Q_1$  has 25% of the data below it.
- $Q_2$  has 50% of the data below it. (Median)
- $Q_3$  has 75% of the data below it.

---

---

---

---

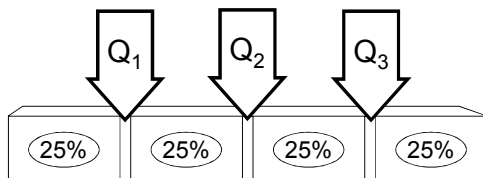
---

---

---

---

## Quartiles



---

---

---

---

---

---

---

---

## Cholesterol

100	124	148	170	185	215
101	125	150	170	185	220
106	127	150	172	186	260
106	128	152	175	187	
110	130	155	175	192	
110	130	157	180	194	
119	133	165	180	195	
120	135	165	180	203	
120	139	165	180	210	
123	140	170	185	212	

19

---

---

---

---

---

---

---

---

---

---

## Cholesterol Quartiles

- $Q_1 = 127.5$
- $Q_2 = 165$  (Median)
- $Q_3 = 185$

20

---

---

---

---

---

---

---

---

---

---

## Five-Number Summary

- minimum = 100
- $Q_1 = 127.5$
- $M = 165$
- $Q_3 = 185$
- maximum = 260

$$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \begin{array}{l} \textit{Interquartile} \\ \textit{Range (IQR)} \\ = Q_3 - Q_1 \\ = 57.5 \end{array}$$

IQR gives spread of middle 50% of the data

21

---

---

---

---

---

---

---

---

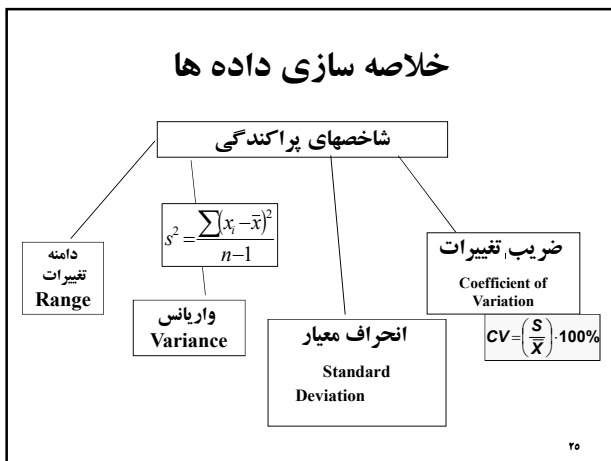
---

---





## خلاصه سازی داده ها




---

---

---

---

---

---

---

---

## انحراف معیار Standard Deviation

- مهم ترین شاخص پراکندگی
- پراکندگی حول میانگین را نشان می دهد

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

انحراف معیار نمونه

۲۱

---

---

---

---

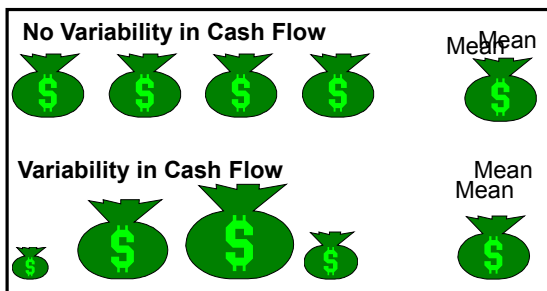
---

---

---

---

## Variability



۲۲

---

---

---

---

---

---

---

---

**Variance:** a measure of how data points differ from the mean

• Data Set 1: 3, 5, 7, 10, 10

Data Set 2: 7, 7, 7, 7, 7

What is the mean and median of the above data set?

Data Set 1: mean = ? median = ?

Data Set 2: mean = ? median = ?

But we know that the two data sets are not identical!  
The **variance** shows how they are different.

We want to find a way to represent these two data set numerically.

18

---

---

---

---

---

---

---

---

---

---

### How to Calculate sd?

X	X-Mean	(X-Mean) <sup>2</sup>
3	3-7=-4	16
5	5-7=-2	4
7	7-7=0	0
10	10-7=3	9
10	10-7=3	9
35		38

$$sd = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}} = \sqrt{\frac{38}{5-1}} = \sqrt{9.5} = 3.1$$

19

---

---

---

---

---

---

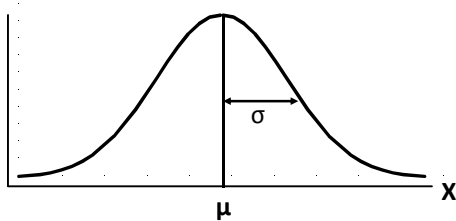
---

---

---

---

### The Population Mean and Standard Deviation



30

---

---

---

---

---

---

---

---

---

---

**Example: Age of Patients**

Male	Female
28	27
22	27
21	28
26	6
18	27

Find the mean, median, range?

<b>mean</b>	<b>23</b>	<b>23</b>
<b>median</b>	<b>22</b>	<b>27</b>
<b>range</b>	<b>10</b>	<b>22</b>

What can be said about this data?

**Due to the outlier, the median is more typical of overall performance.**

Which diver was more consistent?

31

---

---

---

---

---

---

---

---

---

---

---

---

	Age $X$	$x - \bar{x}$	$(x - \bar{x})^2$
1	28		
2	22		
3	21		
4	26		
5	18		

31

---

---

---

---

---

---

---

---

---

---

---

---

**Example**

Metabolic rates of 7 men (cal./24hr.) :

1792 1666 1362 1614 1460 1867 1439

33

---

---

---

---

---

---

---

---

---

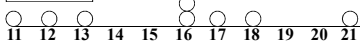
---

---

---

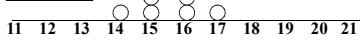
### مقایسه انحراف معیار داده ها

Data A



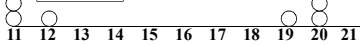
Mean = 15.5

Data B



Mean = 15.5

Data C



Mean = 15.5

۳۴

---

---

---

---

---

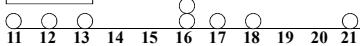
---

---

---

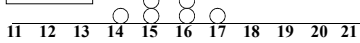
### مقایسه انحراف معیار داده ها

Data A



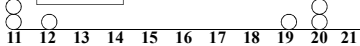
Mean = 15.5  
sd = 3.338

Data B



Mean = 15.5  
sd = .9258

Data C



Mean = 15.5  
sd = 4.57

۳۵

---

---

---

---

---

---

---

---

SYSTOLIC BP						
Age Groups	N	Mean	Std. Deviation	Minimum	Maximum	Range
<35	54	109.54	11.421	90	140	50
35-44	60	112.08	17.303	85	180	95
45-54	52	125.87	20.430	100	185	85
55-64	54	136.30	22.554	100	180	80
65+	30	134.33	23.034	100	210	110
Total	250	122.30	21.848	85	210	125

۳۶

---

---

---

---

---

---

---

---

Coefficient of Variation ضریب تغییرات

- پراکندگی نسبت به میانگین را نشان می دهد
- مقایسه پراکندگی دو یا چند گروه
- مقایسه پراکندگی دو متغیر با واحد اندازه گیری متفاوت

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

۳۷

---

---

---

---

---

---

---

---

Coefficient of Variation ضریب تغییرات

$$\bar{X} = 3000 \quad Sd = 300$$

وزن نوزادان

$$\bar{X} = 40Kg \quad Sd = 2$$

وزن افراد ۱۲ ساله

۳۸

---

---

---

---

---

---

---

---

Coefficient of Variation ضریب تغییرات

$$\bar{X} = 3000 \quad Sd = 300$$

$$\bar{X} = 3000 \text{ gr} \quad Sd = 300 \text{ gr} \quad CV = \frac{300}{3000} \times 100 = 10\%$$

وزن نوزادان

$$\bar{X} = 40Kg \quad Sd = 2$$

وزن افراد ۱۲ ساله

$$\bar{X} = 40kg \quad Sd = 2kg \quad CV = \frac{2}{40} \times 100 = 5\%$$

۳۹

---

---

---

---

---

---

---

---

ضریب تغییرات داده های زیر را محاسبه کنید. کدام یک پراکندگی بیشتری دارد؟

SYSTOLIC BP			
Age Groups	N	Mean	Std. Deviation
<35	54	109.54	11.421
35-44	60	112.08	17.303
45-54	52	125.87	20.430
55-64	54	136.30	22.554
≥65	30	134.33	23.034
Total	250	122.30	21.848

---



---



---



---



---



---



---



---