Shahid Beheshti University of Medical Sciences
School of Public Health and Safety
Department of Epidemiology

A Review on
Descriptive Statistical Methods

Dr Y Mehrabi

ymehrabi@gmail.com

Why Statistics?

- As a clinician why do we need to get familiar with statistics?

Variables

Type of variables:

- Quantitative
  - Continuous
  - Discrete
- Qualitative
  - Ordinal
  - Nominal

Role of variables:

- Dependent
- Independent
- Background
- Confounding

Numerical Summaries

- **Center of the data**
  - Mean
  - Median
  - Mode
- **Variation**
  - Range
  - Quartiles (interquartile range)
  - Variance
  - Standard Deviation

Median

- Half of the values are less than or equal to the median value
- Half of the 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

Median

- Example 1: 3 4 12 7 6
- Example 2: 2 4 8 11

- A ***resistant measure*** of the data center
- Example 4: 2 4 6 8 10
- Example 5: 2 4 6 8 100
- Example 6: 2 4 6 800 10000

Example: Age of Patients

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

Find the mean, median, range?

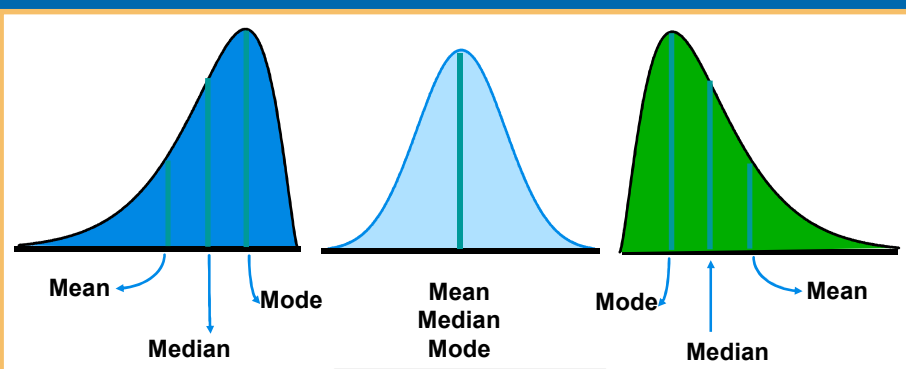
mean	**23**	**23**
median	**22**	**27**
range	**10**	**22**

What can be said about this data?

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

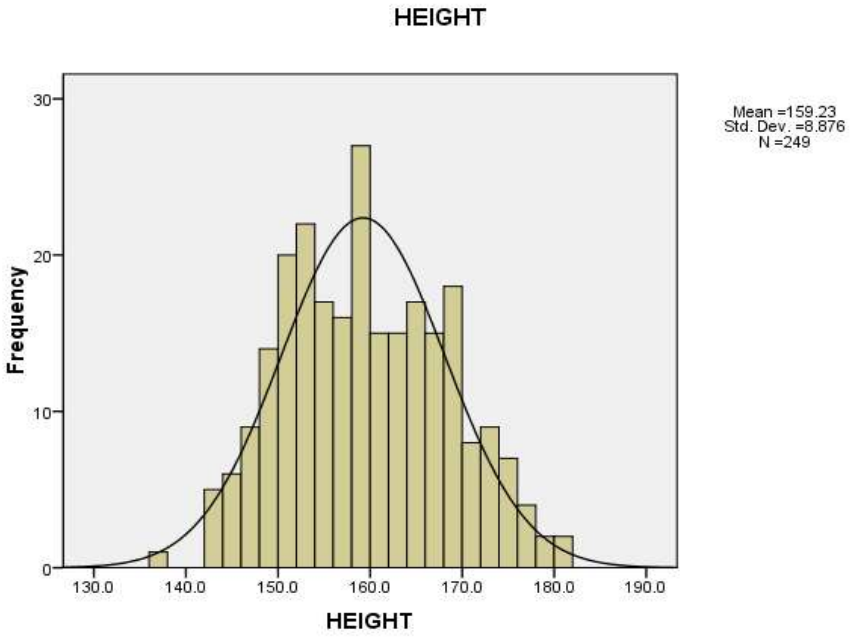
Which measure is more consistent?

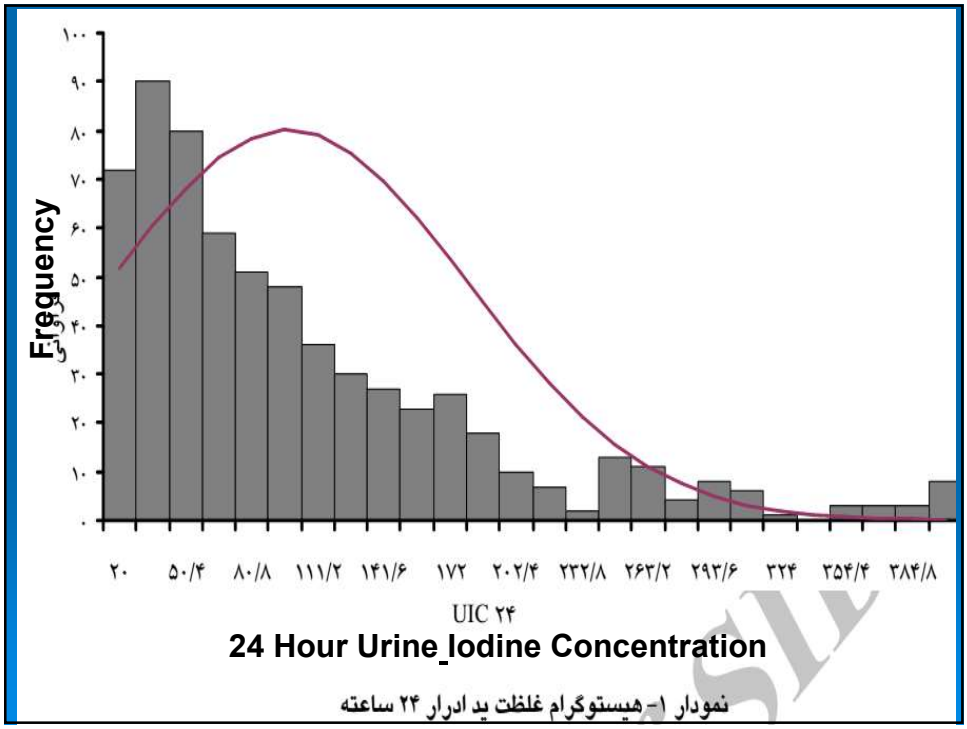
Skewness



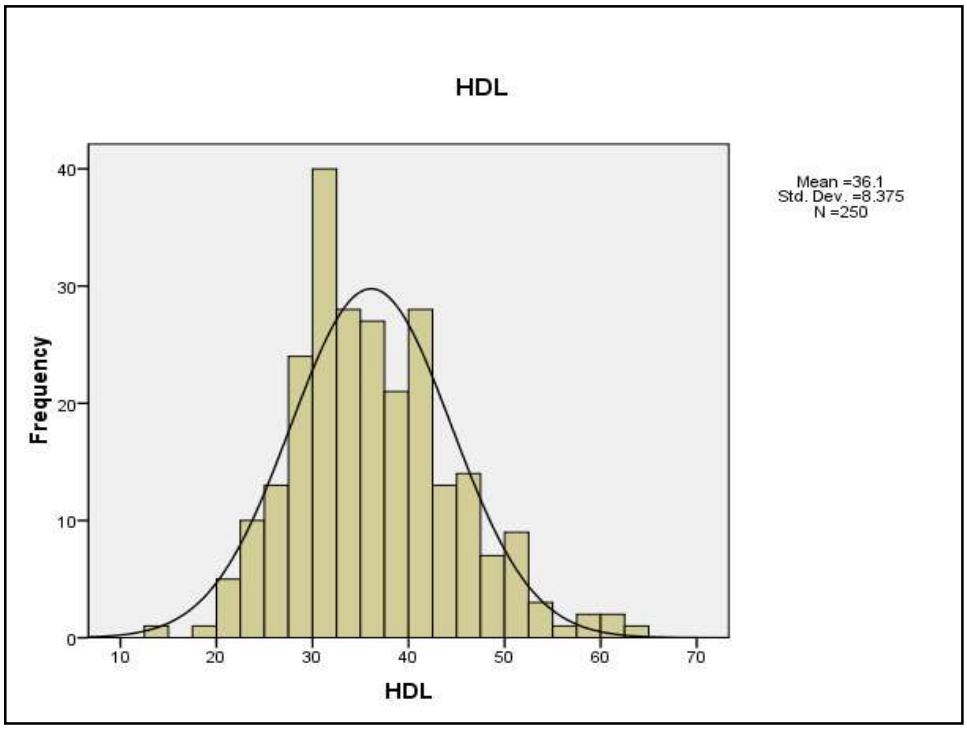
Comparing the Mean & Median

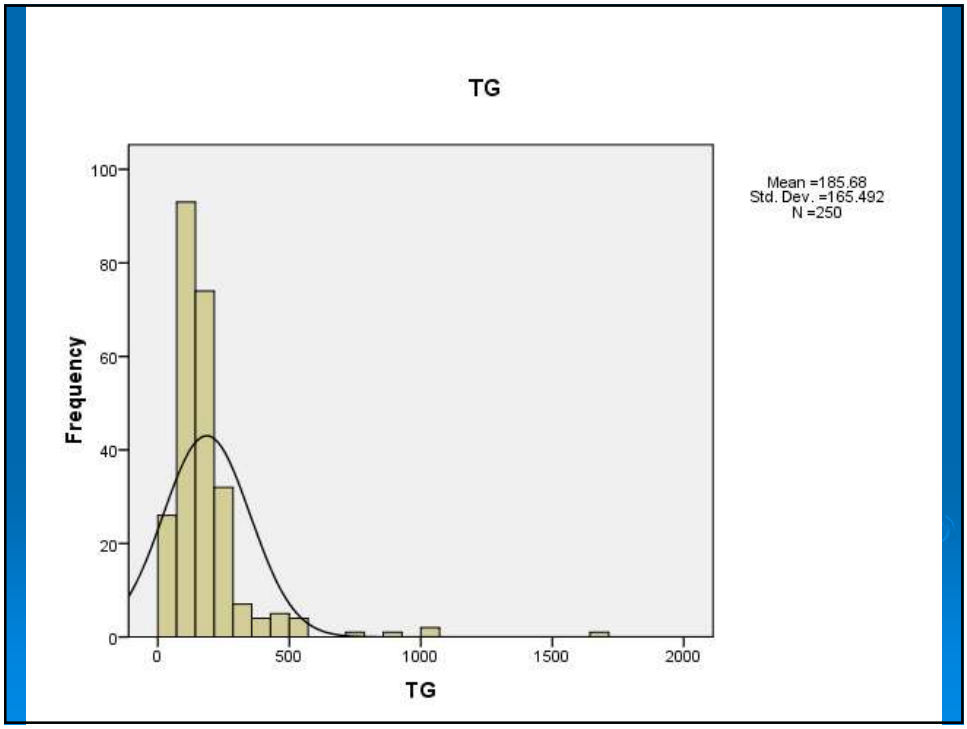
- The mean and median of data from a symmetric distribution should be close together. The actual (true) mean and median of a symmetric distribution are exactly the same.
- In a skewed distribution, the mean is farther out in the long tail than is the median [the mean is 'pulled' in the direction of the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

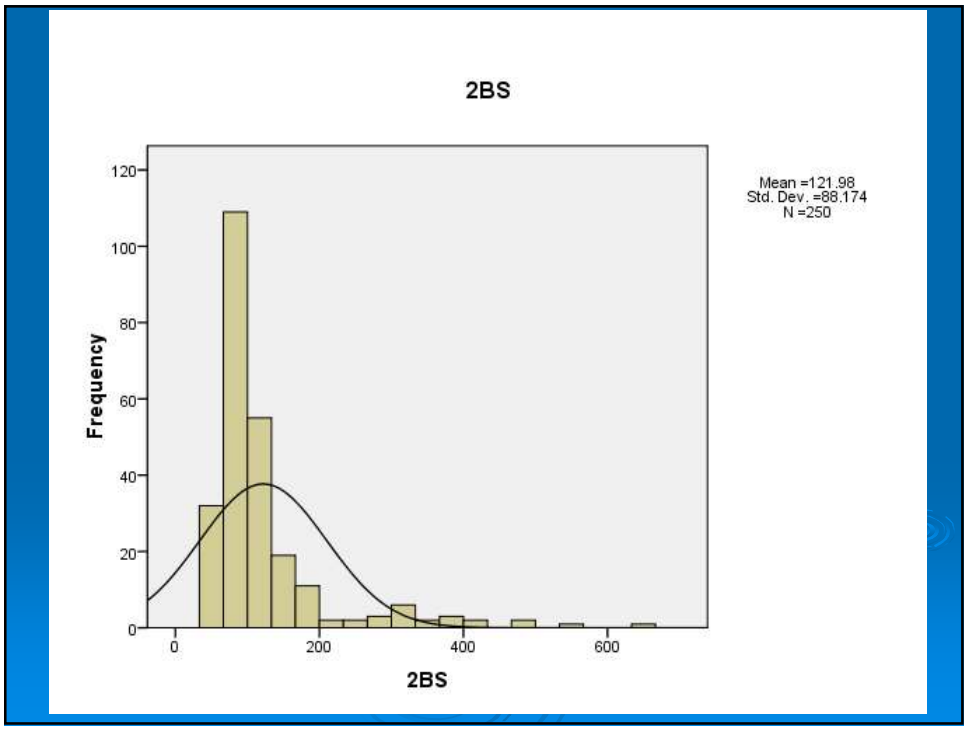




نمودار ۱- هیستوگرام غلظت ید ادرار ۲۴ ساعته

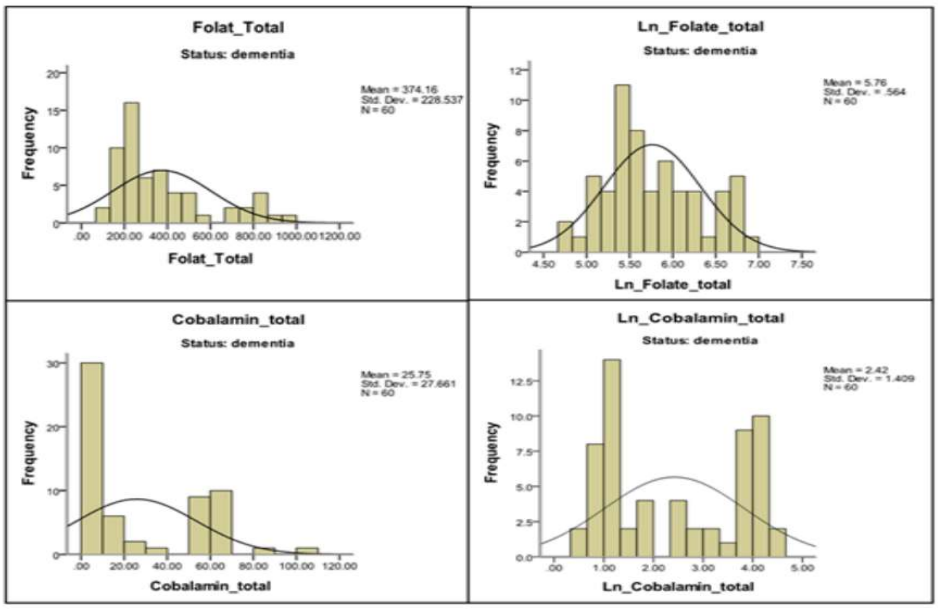



TG
Mean =185.68
Std. Dev. =165.492
N =250
Frequency
100
80
60
40
20
0
0
500
1000
1500
2000
TG


2BS
Mean =121.98
Std. Dev. =88.174
N =250
Frequency
120
100
80
60
40
20
0
0
200
400
600
2BS

Example

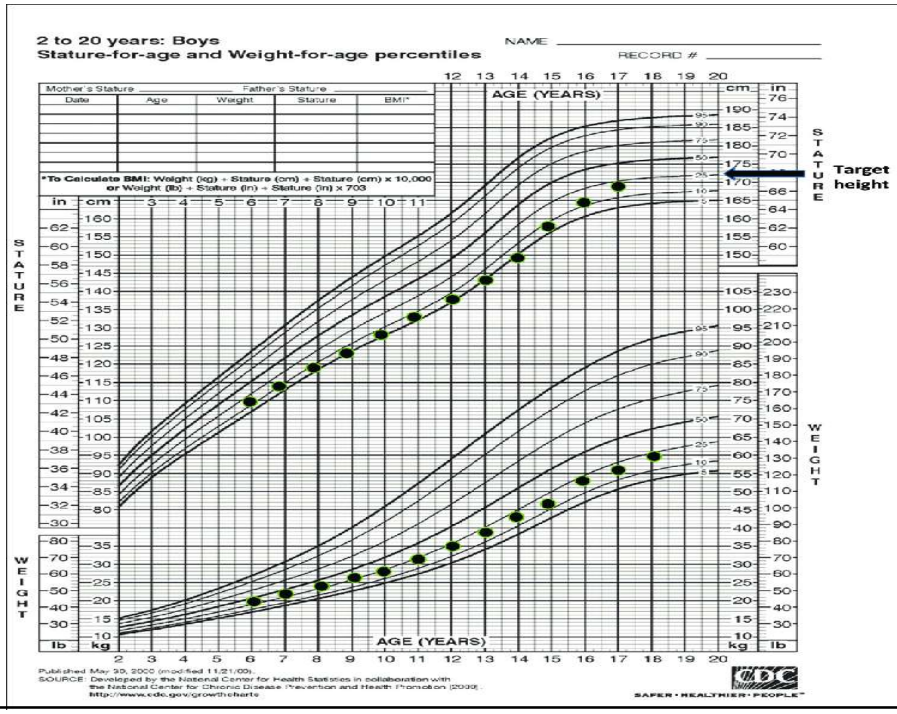
	WEIGHT	HEIGHT	HDL	TG	2hrBS
N Valid	250	249	250	250	250
Mean	**67.5**	159.2	36.1	185.6	121.9
Median	**67.2**	158.5	35.0	148.5	97.0
Std. Deviation	12.40	8.8	8.3	165.5	88.2
Range	63.5	44.0	49	1645	619
Minimum	38.5	137.0	14	43	34
Maximum	102.0	181.0	63	1688	653



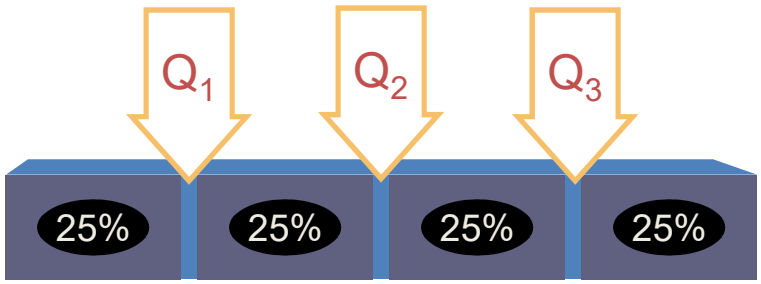
شکل ۲- نمودار هیستوگرام فولات و کوبالامین توتال در افراد دمانس

Quantiles چندک‌ها

- Centiles صدک ها
- Tentiles دهک ها
- Quintiles پنجک ها
- Quartiles چارک ها
- Thertiles سهک ها



Quartiles



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.

Cholesterol

L(M)=(53+1)/2=27

$L(Q_1)=(26+1)/2=13.5$

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	

Cholesterol Quartiles

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

Five-Number Summary

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

Interquartile Range (IQR) = $Q_3 - Q_1$
= 57.5

IQR gives spread of middle 50% of the data

Median (IQR)

TABLE 3. MEDIAN (INTERQUARTILE RANGE) 24-HOUR URINARY IODINE CONCENTRATION, SALT INTAKE, SALT PER CAPITA, AND IODINE CONTENT OF SALT IN MEN AND WOMEN IN DIFFERENT REGIONS OF TEHRAN

	n	*24-Hour UIC (μg/L)*[a]	*Salt intake (g/day)*[a]	*Salt per capita (g/day)*[a]	*Iodine content of salt (ppm)*[a]
All population					
Men	242	72.9 (37.0–136.1)[b]	7.7 (5.5–10.2)[b]	8.4 (5.7–12.0)[b]	21.2 (3.2–31.7)[b]
Women	397	68.6 (33.0–128.0)	7.5 (5.6–9.8)	8.5 (5.7–12.5)	20.1 (3.2–21.7)
Total	639	70.0 (34.0–131.2)	7.6 (5.5–9.8)	8.4 (5.7–12.2)	21.2 (3.2–31.7)
South					
Men	62	49.0 (34.0–87.5)	7.6 (5.2–9.3)	9.1 (6.6–13.9)	3.2 (1.8–27.5)
Women	85	48.0 (30.0–79.7)	7.8 (5.9–10.2)	9.1 (6.5–12.2)	3.2 (1.1–24.3)
Total	147	49.0 (30.0–86.5)[c,d]	7.7 (5.6–9.8)	9.1 (6.6–12.8)[c,d]	3.2 (1.6–25.9)[c,d]
West					
Men	58	84.0 (31.0–183.0)	8.9 (5.9–11.1)	7.6 (5.9–11.2)	22.4 (10.6–29.6)
Women	105	74.9 (45.2–141.6)	8.3 (6.3–10.3)	7.5 (5.2–11.1)	22.7 (10.6–29.6)
Total	163	76.0 (34.0–153.0)	8.5 (6.1–10.6)[d]	7.6 (5.6–11.1)	21.9 (10.6–29.6)[d]
East					
Men	61	78.0 (34.0–112.0)	7.3 (5.4–9.3)	9.2 (6.3–13.5)	15.3 (1.1–25.4)
Women	99	63.0 (29.0–121.5)	7.2 (5.5–9.6)	9.5 (7.5–13.2)	7.4 (1.1–25.4)
Total	160	69.0 (31.5–116.5)[d]	7.3 (5.4–9.5)	9.5 (6.7–13.3)[c,d]	8.5 (1.1–25.4)[c,d]
North					
Men	61	97.2 (49.5–187.2)	7.6 (5.6–11.1)	6.3 (4.7–10.8)	32.8 (25.4–39.1)
Women	108	105.0 (46.2–165.2)	6.8 (5.0–9.2)	7.0 (5.0–13.1)	31.7 (21.7–38.1)
Total	169	102.2 (47.5–166.8)	7.1 (5.4–9.8)	6.7 (4.9–12.2)	31.7 (23.3–38.1)

Characteristics and dietary intakes of study participants by quintile (Q) categories of dietary patt

	Healthy pattern score			
	Q1 (lowest) ($n = 97$)	Q3 ($n = 97$)	Q5 (highest) ($n = 97$)	P^2
Age (y)	49 ± 6[3]	50 ± 7	48 ± 6	0.18
BMI (kg/m^2)	30.4 ± 3.4	27.8 ± 3.9	25.7 ± 3.8	< 0.01
WHR	0.91 ± 0.08	0.89 ± 0.08	0.85 ± 0.05	< 0.01
Physical activity (MET·h/wk)	10.3 ± 9.1	14.7 ± 11.2	17.3 ± 10.8	< 0.01
Family history of diabetes (%)	9	11	11	< 0.05
Family history of stroke (%)	1	1	2	0.74
Current daily smoker (%)	1	1	0	0.81
Obese (%)[4]	47	35	20	< 0.01
Current estrogen use (%)	24	26	26	0.09
Metabolic syndrome (%)[5]	37	27	20	< 0.01

Breast Milk Iodine Concentration Rather than Maternal Urinar Iodine Is a Reliable Indicator for Monitori Iodine Status of Breastfed Neonates

Nazeri *et al.* (2018)

Table 2 Maternal and neonatal urinary iodine and thyrotropin concentrations and iodine content of salt by BMIC

Variables	Breast milk iodine concentration (μg/L)		*P*
	≥ 100	< 100	
Maternal UIC (μg/L)			
Median (IQR)	70 (42–144)	37 (25–100)	0.047
≥ 100, *n* (%)	88 (68.2)	4 (25.0)	0.580
< 100, *n* (%)	41 (31.8)	12 (75.0)	
Neonatal UIC (μg/L)			
Median (IQR)	230 (114–310)	76 (41–140)	< 0.001
≥ 100, *n* (%)	101 (78.3)	5 (31.3)	< 0.001
< 100, *n* (%)	28 (21.7)	11 (68.8)	
Neonatal TSH (mIU/L)			
Median (IQR)	1.00 (0.50–1.70)	1.05 (0.52–2.70)	0.590
> 2, *n* (%)	13 (10.2)	3 (18.8)	0.355
> 3, *n* (%)	6 (4.7)	2 (12.5)	
Iodine content of salt (ppm)			
Median (IQR)	25 (18–32)	20 (10–27)	0.066
≥ 20, *n* (%)	93 (73.2)	9 (43.8)	0.157
< 20, *n* (%)	34 (26.8)	7 (56.3)	

Intergroup comparisons were performed using the chi-square test or Mann-Whitney test as appropriate

BMIC breast milk iodine concentration, *UIC* urinary iodine concentration, *TSH* thyroid-stimulating hormone, *IQR* interquartile range

Quantiles

	WEIGHT	HEIGHT	SYSTOLIC BP	TG	CHOLESTEROL
Median	67.250	158.500	120.00	148.50	192.50
Percentiles 5	47.000	145.750	95.00	63.00	126.10
10	51.550	148.000	100.00	71.00	146.10
20	56.000	151.000	105.00	94.00	157.00
25	58.000	152.000	110.00	102.00	163.00
30	60.000	153.000	110.00	110.30	168.00
40	64.000	156.000	110.00	129.40	179.40
50	67.250	158.500	120.00	148.50	192.50
60	70.000	161.000	120.00	169.00	200.60
70	74.000	164.000	130.00	198.00	219.70
75	76.125	166.000	136.25	209.00	226.00
80	78.000	168.000	140.00	223.00	231.00
90	83.900	171.000	150.00	302.10	251.90
95	90.000	174.250	170.00	434.50	271.00

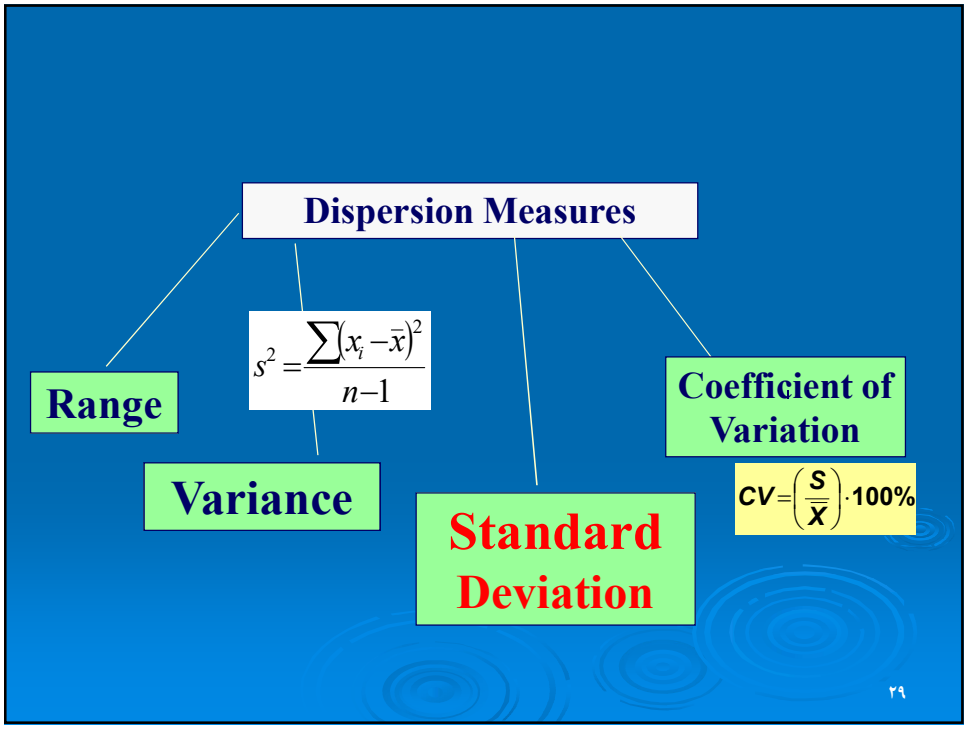
Shahid Beheshti University of Medical Sciences
School of Public Health and Safety
Department of Epidemiology

A Review on

Descriptive Statistical Methods

Dispersion Measures

Dr Y Mehrabi
ymehrabi@gmail.com



Standard Deviation

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

Variability

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.

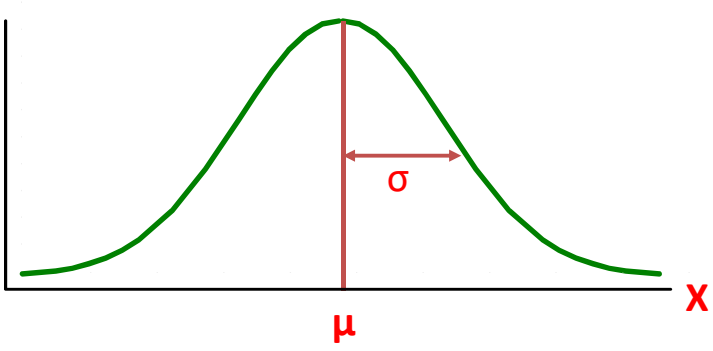
We want to find a way to represent these two data set numerically.

How to Calculate sd?

	X	X-Mean	$(X\text{-Mean})^2$
	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

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

The Population Mean and Standard Deviation



Example: Age of Patients

	Male	Female
	28	27
	22	27
	21	28
	26	6
	18	27
mean	**23**	**23**

Age X	$X - \overline{X}$	$(X - \overline{X})^2$
28		
22		
21		
26		
18		

Example: Age of Patients

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

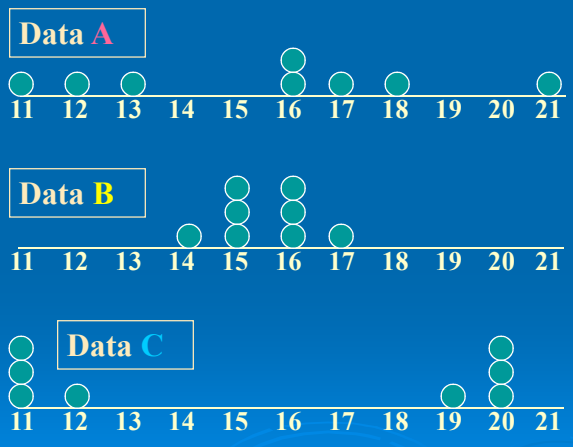
Example: Age of Patients

	Male	Female
	28	27
	22	27
	21	28
	26	6
	18	27
mean	23	23
	$sd = 4$	$sd = 9.5$

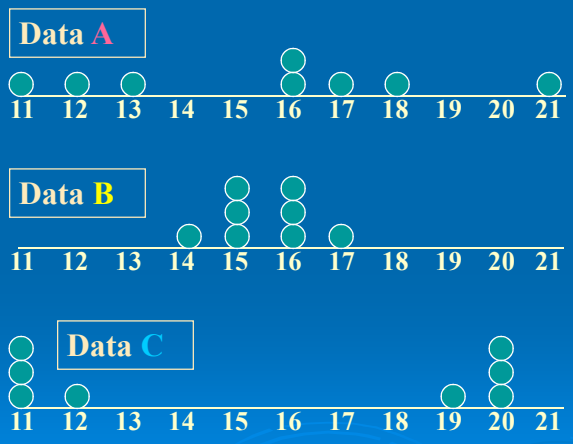
Age X	$X - \overline{X}$	$(X - \overline{X})^2$
28	5	25
22	-1	1
21	-2	4
26	3	9
18	-5	25
جمع	0	64

$$sd = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)}} = \sqrt{\frac{64}{5-1}} = \sqrt{16} = 4$$

Comparison of Data SD



Comparison of Data SD



Mean = 15.5
sd = 3.338

Mean = 15.5
sd = .9258

Mean = 15.5
sd = 4.57

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

Coefficient of Variation

- Variation relative to mean
- Comparison of two or more variables variation with different means
- Comparison of two or more variables variation with different units

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

Coefficient of Variation

Birth Weight $\overline{X} = 3000 \quad Sd = 300$

12 years old children weight $\overline{X} = 40Kg \quad Sd = 2$

Coefficient of Variation

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

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

12 years old weight $\bar{X} = 40Kg \quad Sd = 2$

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

Bone mineral density measurements

- The Hologic system DXA (Discovery QDR, USA) was applied to measure BMD (g/cm2), bone mineral content (BMC) in grams, bone area (BA) in square centimeters. The coefficient of variation for the femoral neck, the lumbar spine, and the total body (based on the measurements in ten children) was 2.4%, 0.51%, and 1% respectively.

Please compute the coefficient of variation for the systolic blood pressure of each age group:

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	