بسم الله الرحمن الرحيم

Biostatistics

L V

16th –July 2023

PROF. DR. WAQAR AL-KUBAISY

Description statistics summarization

Presentation

- Graph
- Table

Numerical

- Measures of central tendencies
- Measures of Dispersion

the choice of the most appropriate measure depends crucially on the type of data involved

The interquartile range is not affected either by

- **Outlier**
- **skewness**

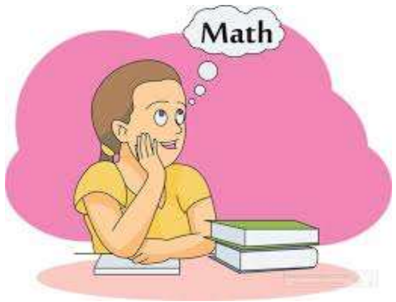
BUT

The limitation of iqr

it does not use all of the information in the data set since it ignores the bottom and top quarter of values.

So

- **I have to use the whole data values**
- **variation of each value from the other??**



An alternative approach use the idea of summarizing spread by measuring

✓ ***measure the variation of one observation from the other***

✓ **Standard deviation**

Standard deviation

75, 70, 75. 80, 85.
60, 65, 55, 70, 75, 75, ,70, 80,

variation of each value, from the other??

60, 65, 55, 70, 75, 75, ,70, 80, 40, 45, 53, 77, 75, 95, ,100, 88, 68, 95, 57, 78, 35, 95, ,78, 85, 67, 69, 35, 71, 79, 77

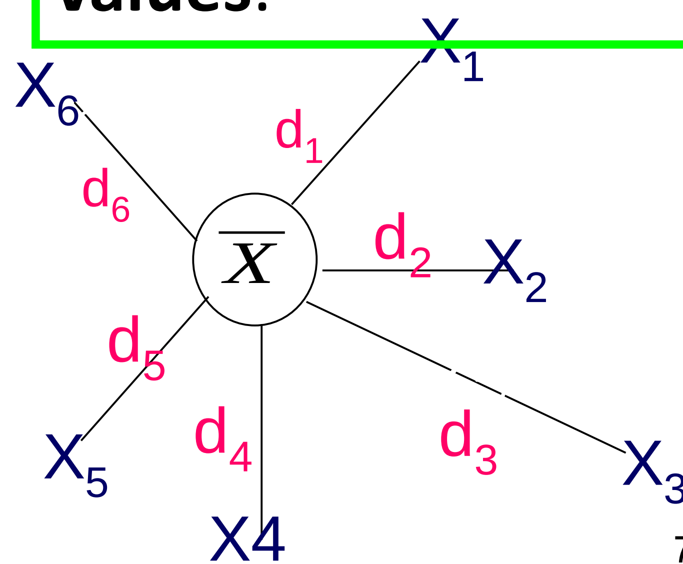
variation of each value from the other???

75, 70, 75. 80, 85. **Mean = ????**

60, 65, 55, 70, 75, 75, ,70, 80, Mean= ????

$$\overline{X} = \frac{\sum X}{N}$$

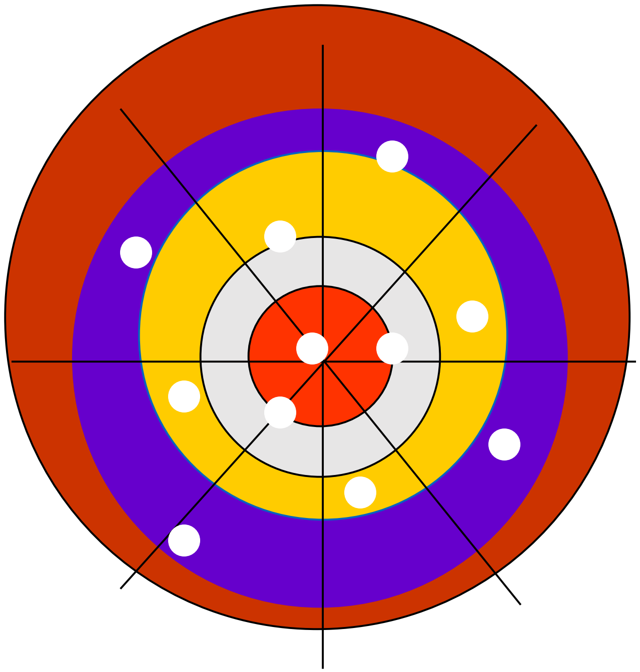
the mean (average) variation of all data values from the over all mean of all values.

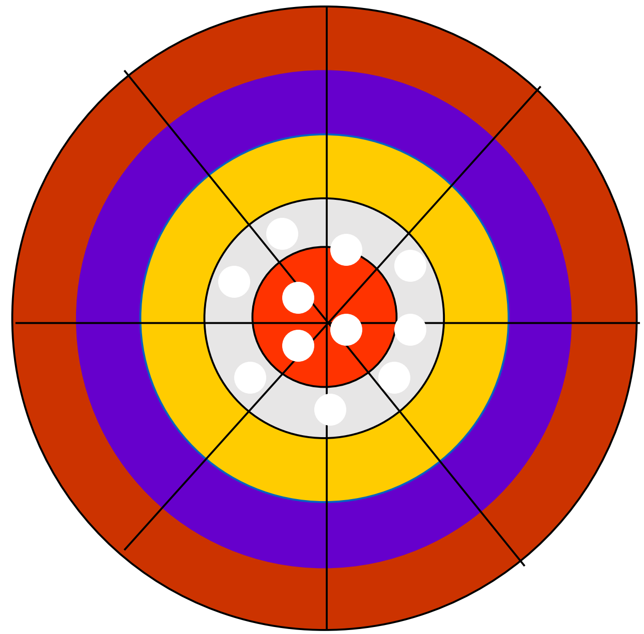


the mean (average) distance of all data values from the over all mean of all values.

Measures of Dispersion

SHOOTER A **SHOOTER B**





Both shooters are hitting around the "centre" but shooter B is more "accurate"

- **The smaller the mean distance is**
- ✓ **the narrower the spread of values**

The limitation of iqr it does not use all of the information in the data since it omits the top, and bottom quarter of values.

- **An alternative approach use the idea of summarizing spread by measuring**
- **the mean (average) distance of all data values, from the over all mean of all values.**

- **The smaller the mean distance is**
- ✓ **the narrower the spread of values must be and visa versa**

this is known as **standard deviation**

Measures of Dispersion
(Measures of Variation)
(Measures of Scattering)
measures of spread

Measures of Dispersion

1- **Range**

2- Variance

3- Stander Deviation

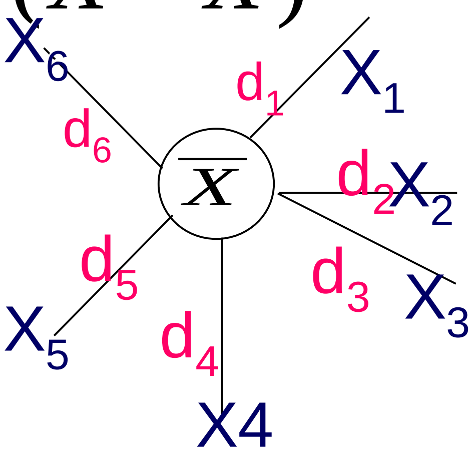
4- Coefficient of variance

Measures of Dispersion

student No.	score	$x - \overline{x}$
1st	**6**	**6 – 3** = +3
2nd	**2**	**2 – 3** = **-**1
3rd	**4**	**4 – 3** = +1
4th	**1**	**1 – 3** = **-**2
5th	**3**	**3 – 3** = 0
6th	**2**	**2 – 3** = **-**1

the **mean**(average) variation of **all** data **values** from the **over all mean** of all values.

$$(X - \overline{X})$$



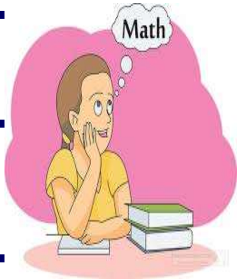
$$\boxed{\sum X = 18}$$

$$\boxed{\sum (X - \overline{X}) = zero}$$

$$\boxed{\overline{X} = 3}$$

????

student No.	Score	$X - \bar{X}$	$X - \bar{X})^2$
1st	6	$6 - 3 = +3$	9
2nd	2	$2 - 3 = -1$	1
3rd	4	$4 - 3 = +1$	1
4th	1	$1 - 3 = -2$	4
5th	3	$3 - 3 = 0$	0
6th	2	$2 - 3 = -1$	1



$$\bar{X} = 3$$

$$\sum X = 18$$

$$\sum (X - \bar{X}) = zero$$

$$\sum (X - \bar{X})^2 = 16$$

$$(X - \bar{X}) \Rightarrow (X - \bar{X})^2$$

Variance

$$S^2 = \frac{\sum (X - \bar{X})^2}{N - 1} =$$

$$S^2 = \frac{16}{5}$$

3.2 score^2

????

Variance S^2

It is the Average of **squared deviation** of **observation from the mean in a set of data** .

$$S^2 = \frac{\sum (X - \overline{X})^2}{N-1}$$

3.2 score ^{2}

????

The Disadvantage or drawback of variance

that its unit is squared Kg^2 , bacteria^2 ....., So Restore the squared unit into its **original form**

by

taking the square root of this (S^2) value, this is **known as Stander Deviation** (**S.D).**

$\sqrt{3.2}$ = ±1. 789 score

Standard Deviation ± S.D.

It is the square root of variance. S.D = √S²

$$S^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

$$\sqrt{\frac{\sum (X - \overline{X})^2}{N - 1}} = \pm S.D$$

± S.D (S) it is the square root of the **Average square deviation** of **observation from the mean in a set of data**

One advantage of SD is that unlike the iqr
it uses all the information in the data

Steps in calculating S.D

1.Determine the **mean** $\overline{X}$

2-Determine the deviation of each value from the mean $(X - \overline{X})$

3-.**Square each deviation of value from mean** $(X - \overline{X})^2$

4-Sum these square deviation of value from mean $\sum(X - \overline{X})^2$

(sum of square) . $\sum(X - \overline{X})^2$

5-**Divide** this **square deviation of value from mean by N-1**

$$\frac{\sum(X - \overline{X})^2}{N-1}$$

.

6-Take the square root of deviation of value from mean by N-1

$$\boxed{\sqrt{\frac{\sum(X - \overline{X})^2}{N-1}} = \pm S.D}$$

Short Cut Method

	score	Score ²
1	6	36
2	2	4
3	4	16
4	1	1
5	3	9
6	2	4
total	**18**	70

$$S^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

$$\sum (X - \overline{X})^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$\boxed{S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}}$$

$$\frac{70 \;\; - 18 X 18 / 6}{5} = \quad 70\text{-}54\text{=}16/5\text{=}\ 3.2$$

$\sqrt{3.2}$ = 1.789???????

Short Cut Method for S.D

1-Square **each absolute** individual value . X^2

2-Sum these **squared values** $(\sum X)^2$.

3-Sum the all absolute value of observation $X_1 + X_2 + X_3 + ... = \sum X$

4-Square this **sum** of absolute values

5-Divide this sum of absolute values by N $\frac{(\sum X)^2}{N}$

6-Subtract $\frac{(\sum X)^2}{N}$ from $\sum X^2$ $\longrightarrow$ $\sum X^2 - \frac{(\sum X)^2}{N}$ (**sum of square**)

7-Divided all this result by N-1 , $S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}$

8-Take the square root of this last result,

$$\boxed{S.D = \pm\sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}}}$$

Example

Short Cut Method

	Score	Freq.(No.of Students)	XF	X^2F
	6	2	$6\times2=12$	$6^2\times2=72$
	2	4	$2\times4=8$	$2^2\times4=16$
	4	3	$4\times3=12$	$4^2\times3=48$
	1	5	$1\times5=5$	$1^2\times5=5$
	3	2	$3\times2=6$	$3^2\times2=18$
	2	6	$2\times6=12$	$2^2\times6=24$
total		22	55	183

$$S^2 = \frac{\sum (X-\overline{X})^2}{N-1}$$

$$\sum (X-\overline{X})^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$\boxed{S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}}$$

$$S^2 = \frac{183 - \frac{55^2}{22}}{22-1} = \frac{183-137.5}{21} = 2.166$$

scor²

S.D = √2.166 = 1.472

??????

Disadvantage Limitation or Drawback of S.D

It is depend on the unit of measurement,

we can't compare between two or more data

to overcome this

Coefficient of Variation C.V

It is representing by measuring the variation in relation to the percentage of mean of that data

$$\boxed{C.V = \frac{S.D}{\overline{X}} \times 100}$$

$$C.V = \frac{1.47}{2.5} X100 = 58.8\%$$

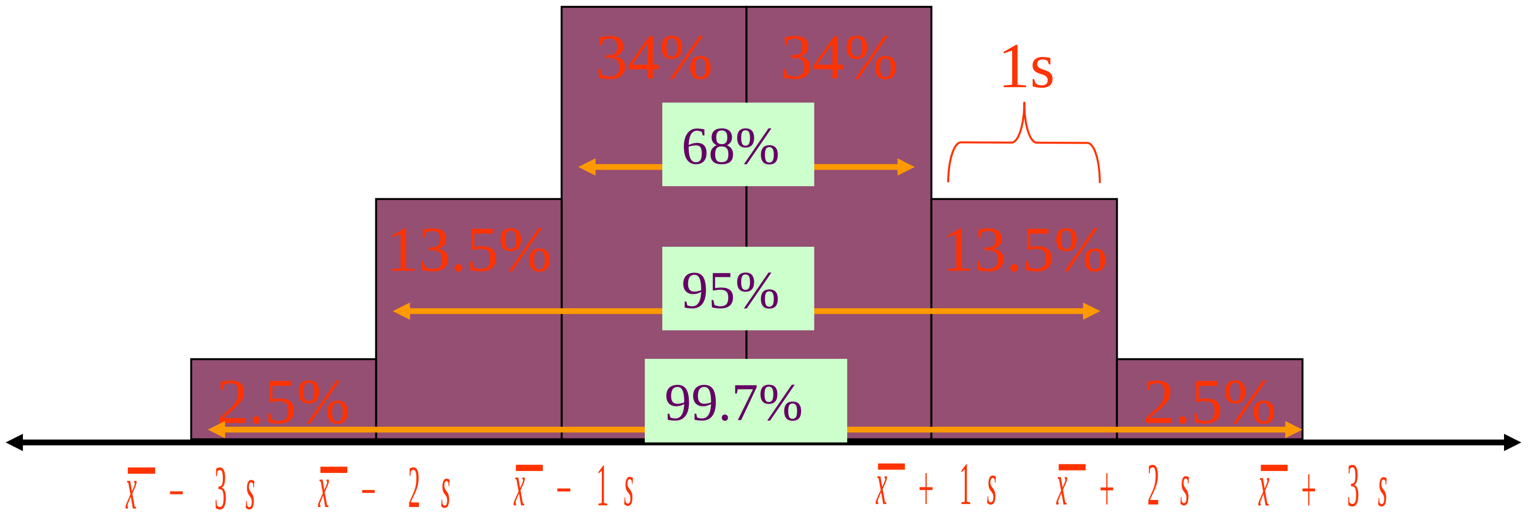
????

-C.V is used

to compare between two or more data

- with different units of measurement .
- data with large difference between their means .

Interpreting Standard Deviation



For bell-shaped shaped distributions, the following statements hold:

- Approximately 68% of the data fall between $\bar{x} - 1s$ and $\bar{x} + 1s$
- Approximately 95% of the data fall between $\bar{x} - 2s$ and $\bar{x} + 2s$
- Approximately 99.7% of the data fall between $\bar{x} - 3s$ and $\bar{x} + 3s$

For NORMAL distributions, the word 'approximately' may be removed from The above statements.


Thank you!

Q1

SD used with median

SD used with rang

SD used in nominal data

IQR used with the mean

Variance is the best measurement of dispersion

Q2 Measures of dispersion are

1

2

3

4

5

6

Thank You

1. Median is the value with a highest frequency
2. When the data is skewed , median is the appropriate measures of CT
3. Mean is appropriate measures of Ct in ordinal data

4. Mode used when we have Metric continuous data

5- mean is unique what ever the size of data is

Q1

Thirty (30) pregnant women attending Al- Karak antenatal clinic during 23-februry 2023 showing gain in weight as follows:

Weight gain (kg	NO.of women
4	3
7	5
10	10
12	8
16	4

1-Present this data graphically,

2- Compute the measures of Central tendency

3- Compute Measures of Dispersion