بسم الله الرحمن الرحيم

Biostatistics

L III

11 July 2023

PROF. DR. WAQAR AL-KUBAISY

Numerical Presentation

Numerical Description

Measures of Central Tendency

Measures of Dispersion

This include:

Presentation of data by

~~1. Graph and or~~

~~2. Tables~~

3. **Calculation or numerical summaries, such as Frequency, Average, Mean, Median, Mode Percentages**

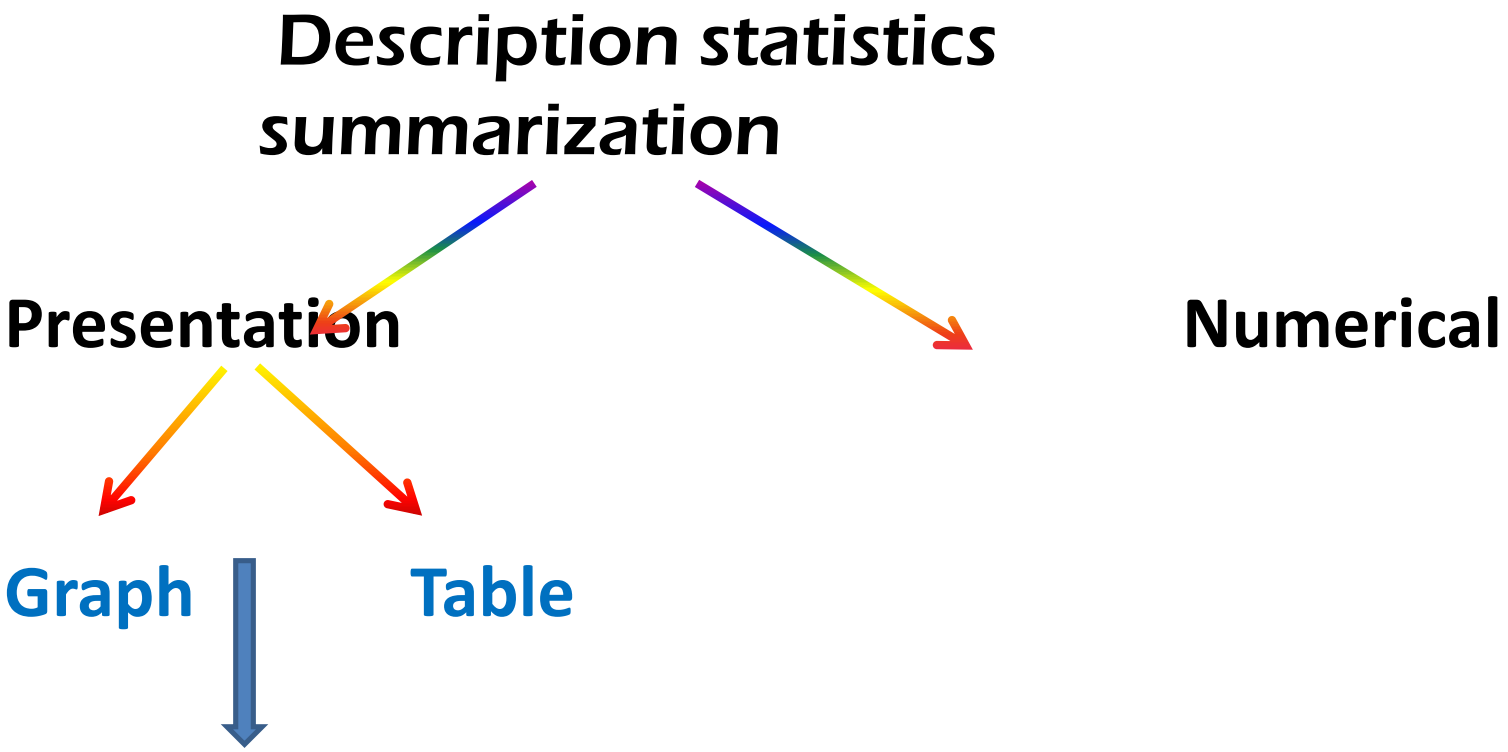
Descriptive statistics

Biostatistics consist of

1-Collection of data .

2-Presentation of data

3-.Estimation of data



-this approach might not be enough,

-comparisons between one set of data & another

-summarize data by one more step further .

-presenting a set of data by a

- single Numerical value

Numerical Presentation
Numerical Description

1-Measures of Central Tendency
2-Measures of Dispersion

Example

The following data representing age (years) of 50 patients with diabetes Mellitus collected from Al Karak Hospital during march 2023

68, 62, 62, 66, 68, 65, 64, 71,77, 74, 20, 33, 38. 42, 47. 50,55, 56, 60 72, 80 74, 75, 74, 77,80, 81, 89, 86, 85, 83,72, 70, 71, 79, 76, 77, 80, 90, 97, 94, 90,65, .60, 67, 63 88, 84, 84, 87

7

An important thing is the type of the variable concerned.

The central value as representative value in a set of data

1-Measures of central tendencies (Location) .

A value around which the data has a tendency to congregate (come together)or cluster

2-Measures of Dispersion, scatter around average

A value which measures the degree to which the data are or are not, spread out

-single Numerical value. ??

Are we using largest value ?

Are we using lowest value ?

As a single Number representation

The central value as
representative value in a set of data,

The central value as representative value in a set of data

1-Measures of central tendencies (Location) .

A value around which the data has a tendency to congregate (come together)or cluster

2-Measures of Dispersion, scatter around average

A value which measures the degree to which the data are or are not, spread out

Measures of Central Tendency

A value around which the data has a tendency to congregate or cluster

1- Mean

2- Median

3- Mode

4- weighted mean

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

Mode (Mo)

- **Most frequently occurring value in a set of observation**

5 1, 3, 2, 6 , 7, 10 5 ?????

Or

- **the value of observation which has the highest frequency in a set of observation .**

1 5 1, 3, 1, 2, 6 , 7, 10 5 ?????

- **Mode is the only measure of central tendency that can be used for qualitative data ???**
- **is not practically useful with the metric continuous data where no two value may be the same,**
- **If the observation all having different value**

5 1, 3, 2, 6 , 7, 10 ?????

So

the observation all having different value

there is no Mode 5 1 3 2 6 .

We might have one Mode 5 , 1 2, 3, 1, 6 uni modal

We might have more than one Mode

5, 1, 3, 5 7, 3, 6 , 2 Two Mode Bimodal

5, 1, 3, 5, 7, 3, 6, 2, 1 Three Mode Tri modal

5, 1, 3, 5, 7, 3, 6, 2, 1 , 3 ???

3

uni modal

Characteristics of Mode

Advantages and Disadvantages

1-Requires no calculation just counting

2- It may not exist (No Mode)

3-It is not necessarily be unique

there may be one mode unimodal

more than one mode in a set of data

Bimodal, Tri modal ....

- It is the only measure of central tendency that can be used for qualitative data

4 -Mode is not practically useful with the metric continuous data

Median (Md)

Measures of C T
1- Mean
2- Median
3- Mode

It is the middle value in ordered data
(from the lowest to the highest values).
-Divided the observations into two halves .

So

- 1/2 of observation their values less than the value of median
- 1/2 of observation their values More than the value of median

- Median is located the center of data by count and disregards the size .
- Median is thus a measure of centrals

Steps in calculating the median

1- Arrange the value.

From the lowest to the highest value .

Exam. marks

50 10 90 20 40 → 10 20 40 50 90

2- Find the Median position by this formula

$$\frac{n+1}{2} = \frac{5+1}{2} = 3rd$$

Calculate the value of the third observation = 40 marks .

Odd No. we have just one median position .

Even No. we have two median position or two median values

Median value =Average of the two values

Even No 50 10 90 20 40 95

10 20 40 50 90 95

$$\frac{n+1}{2} = \frac{6+1}{2} = \frac{7}{2} = 3.5$$

Median located (position)

between the 3rd and 4th .

Median value =Average of the two (3rd and 4th) values

$$Md = \frac{40+50}{2} = 45$$

Characteristics

10 20

20 40 50 90 95

10 20 40 50 90 95 99 100……..

10 20 40 50 70 85 90 99 100

1 20 40 50 70 85 90 99 100

10 20 40 50 70 85 90 99 1000.

two extremes

15 20 30 35 95 99 100

skewness

1 5 10 **35** 400 900 1000

Characteristics of the Median

It is always existed .

- It is always unique, there is one and only one Md .

- It is not affected by two extremes, not sensitive by two extremities .

- Not affected by skewness in the distribution or

- Not affected by presence of outliers

- It is discard a lot of information

  because it ignores most of the values apart from those in the center of distribution

-

Measures of C T
1- Mean
2- Median
3- Mode

Mean $\overline{X}$

Arithmetic Mean

- more commonly known as average
- -it is an arithmetic average of a set of observation obtained by

- Adding the values of all observation together .
- Dividing the sum by No. of observation in sample .
- *It represent the center of data according to the size of the values .*

Example :

following are the scores of five students

40 50 90 10 20

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

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

Σ = sigma = summation .

X = value of observation

N = No. of observation

$\overline{X}$ = is the sum of value of all observation divided by the total No. of observation

Characteristics of the Mean

Advantages and disadvantages

- Relatively easy to handle
- It is always exist
- It is always unique,

  there is one and only one Mean
- It takes into account every item in a set of data
- It uses all of the information in the data set.

- affected by skewness in the in the data set

- affected by presence of outliers

- it can not be used with the ordinal data ???

- **It is affected by the two extremes by**
  - **a very small or**
  - **a very large value .**
- **It is sensitive to the extremes**

1 2 3 4 5 mean = 3

1 2 3 4 50 mean = 12

1 2 3 4 500 mean = 102

- **this may produce a mean that is not very representative of the general mass of data**

another disadvantage ,

- **it can not be used with the ordinal data ???**

(ordinal data are not real numbers, so they cannot be added or divided)

Weighted mean

It is the average measure of a No. of means, when we take into consideration the frequencies of each mean .

It is used when some values of observation more important in some sense than others .

$$W.mean = \frac{W_1 \overline{X}_1 + W_2 \overline{X}_2 + W_3 \overline{X}_3 + \ldots\ldots\ldots\ldots + W_k \overline{X}_k}{W_1 + W_2 + W_3 + \ldots\ldots\ldots + W_k}$$

Group	$\overline{X}$ Hb	No. of person
I	13	5
II	14	10
III	13.5	15

$$W.mean = \frac{5\times13+10\times14+15\times13.5}{5+10+15} = \frac{407.5}{30} = 13.5\ gm/100\,ml$$

$$\frac{65+140+202.5}{5+10+15} = \frac{407.5}{30} = 13.58$$

Central Tendency In Grouped Data

Age (year)	F	M.P.	(M.P.)F	Cum. F	%
20-29	2	24.5	24.5 2 = 49	2	4
30-39	8	34.5	34.5 8 = 276	10	16
40-49	5	44.5	44.5 5 = 222.5	15	10
50-59	14	54.5	54.5 14 = 763	29	28
60-69	15	64.5	64.5 15 = 967.5	44	30
70-79	6	74.5	74.5 6 = 447	50	12
total	50	---		---	100

$$\sum(\text{M.P.})\text{F} = 2725$$

2725/50 =54.5 years

Choosing the most appropriate measure (Mean, Median or mode)

How do you chose the most appropriate measure of location in a given set of data ??

The main thing is to remember is that

mean can not be use with the ordinal data **(because they are not real numbers**

the median can be use for both ordinal & metric data.

and

the Median can be use for both ordinal & metric data.

when the later (metric data)

is skewed

Or

when there is outlier

the median is

more representative of data than the mean

????????

	Mode	Median	Mean
Nominal	**Yes**	**No**	**No**
Ordinal	**Yes**	**Yes**	**No**
Metric discrete	**Yes**	**Yes if distribution is markedly skewed**	**yes**
Metric continuous	**No**	**Yes** if distribution is markedly skewed	**yes**


Thank you ...


Thank
you


?
Any questions?
© Presentation-Process.com

1-Measures of central tendencies (Location) .
A value around which the data has a tendency to congregate (come together)or cluster
2-Measures of Dispersion, scatter around average
A value which measures
the degree to which the data are or are not , spread out

The central value as

1-Measures of central tendencies (Location)

75, 75, 75, 75, 75, 75, Mean = ????

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

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

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

2-Measures of Dispersion,

The central value as
1-Measures of central tendencies
2-Measures of Dispersion,

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