بسم الله الرحمن الرحيم

Biostatistics

L I

3rd -7-2022

PROF. DR. WAQAR AL-KUBAISY

: Biostatistics

- The terms/concepts:
  - Variable
- Distinguish between
  - Nominal
  - Ordinal
  - Discrete
  - continuous variable
  - Distinguish between quantitative and quantitative data
  - Frequency distribution
  - Relative frequency
  - Cumulative frequency

✓Transform data set into information in the form of

- Tables,
- Graphs

Biostatistics consist of

1-Collection of data .

2-Presentation of data

3-.Estimation of data

Statistics

Is a field of study that concern with

The Collection ,Organization and Summarization of data.

And

Drawing of inference about a body of data when only part of data are observed

Biostatistics

When data being analyzed are derived from biological sciences and Medical observation .

Biostatistics Is a field of study that concern with the **Collection** Organization and Summarization of data. **Drawing o**f inference about a body of data when only part of data are observed

Biostatistics breaks into two main distinct components or two distinct subcategories:

I- Descriptive Biostatistics.

II- Inferential Biostatistics.

I Descriptive Statistics

- It is a series of procedures designed to clarify the data , so that its principal characteristics and main features. } are revealed
- for the purpose of conclusion at a late stage.

This one serve as devices for

- *organizing and summarizing data and*
- *bringing into a focus their essential characteristics*
- Reduce the information to a manageable size

This includ

Descriptive statistics

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**

Biostatistics consist of

1-Collection of data .

2-Presentation of data

3-.Estimation of data

Data

- Data are the values you get when you measure a variable example
  20 years old, (age)
  55 males. (Sex)
  170 cm height
- The values of the observations for the variable is known as data.

- Data are the raw material of statistics
- Data carry little or no meaning when considered alone
- needs further steps to become valuable (information)

- Data consist one or more variable

Variable (Y)

It is the characteristics ,that observed in:

persons, places or things .

This characteristic is not the same when observed in different possessors

It is any aspect of an individual that is measured Like;

B.P. cholesterol age, sex ,Blood sugar ??

Variable

is some thing whose value can vary

example

age ,sex, weight height??

An important thing is the type of the variable concerned

Type of variable

There are two major types of variable

Each of these can be subdivided into two subtypes

1. Categorical variable
(Qualitative Variable)

1.Categorical variable

a- Nominal

b- ordinal

2 Metric variable
(Quantitative Variable)

2 Metric variable

a-Continuous

b-Discrete

1.Categorical variable
a- Nominal
b- ordinal

1. Categorical variable

a- Nominal

Example

Blood group of 100 persons Just categorize the blood group into

A, B, AB, & O

then counting the No. of individuals (frequency) in each group

(1) Data do not have any unit

(2) ordering of the categories is completely subjective,

AB, A, B,& O

O, AB, A & B

1.Categorical variable
a- Nominal
b- ordinal

b- Ordinal ترتيبي

example

grading of tumor I II III IV V

the order category in a meaningful

The difference between any adjacent two grades is not necessarily be the same (equal)

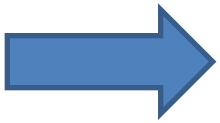
Therefore

- **1-the data are *not properly measures***

but

assessed in some way

- ***2-these data are not real numbers* and**



Continou.. ordinal

as it is not real data

❖3--we cannot apply any arithmetic's roles

no adding,
no subtracting.
no multiply or
no divide

the ordinal vales

❖4- Data do not have any unit

❖5- ordering of the categories is not subjective
the order category in a meaningful way

Ordinal

- * order category in a meaningful way

- *difference between any adjacent two grades is not necessarily be equal
- Have no *interval property*
- **not properly measures*
- **not real numbers*
- cannot apply any arithmetic's roles
- Data do not have any unit

2 Metric variable
a-Continuous
b-Discrete

2 Metric variable

Continuous variable

Example Height ,Weight

- 1-usually comes from measuring
  Can be properly measured
  so
- 2- they are a real numbers
  so
- 3- we can apply all mathematics' operations
- 4- All have units of measurement attached to them
- 5-The difference between any pairs of adjacent values are
- exactly the same (equal) this is
  known as

the *interval property*

50-60-70-80

continuous Variable cont.

- Can be properly measured
- may assume any value along a continuum .
- The value of a C.V. is not limited to the set of integers

Height :158,160,

- 157.9 , 16o.6 160.8
- dose not possess a gap or interruptions.

ex.

B.P. Hb Blood sugar .

??????????

2 Metric variable
a-Continuous
b-Discrete

2 Discrete variable

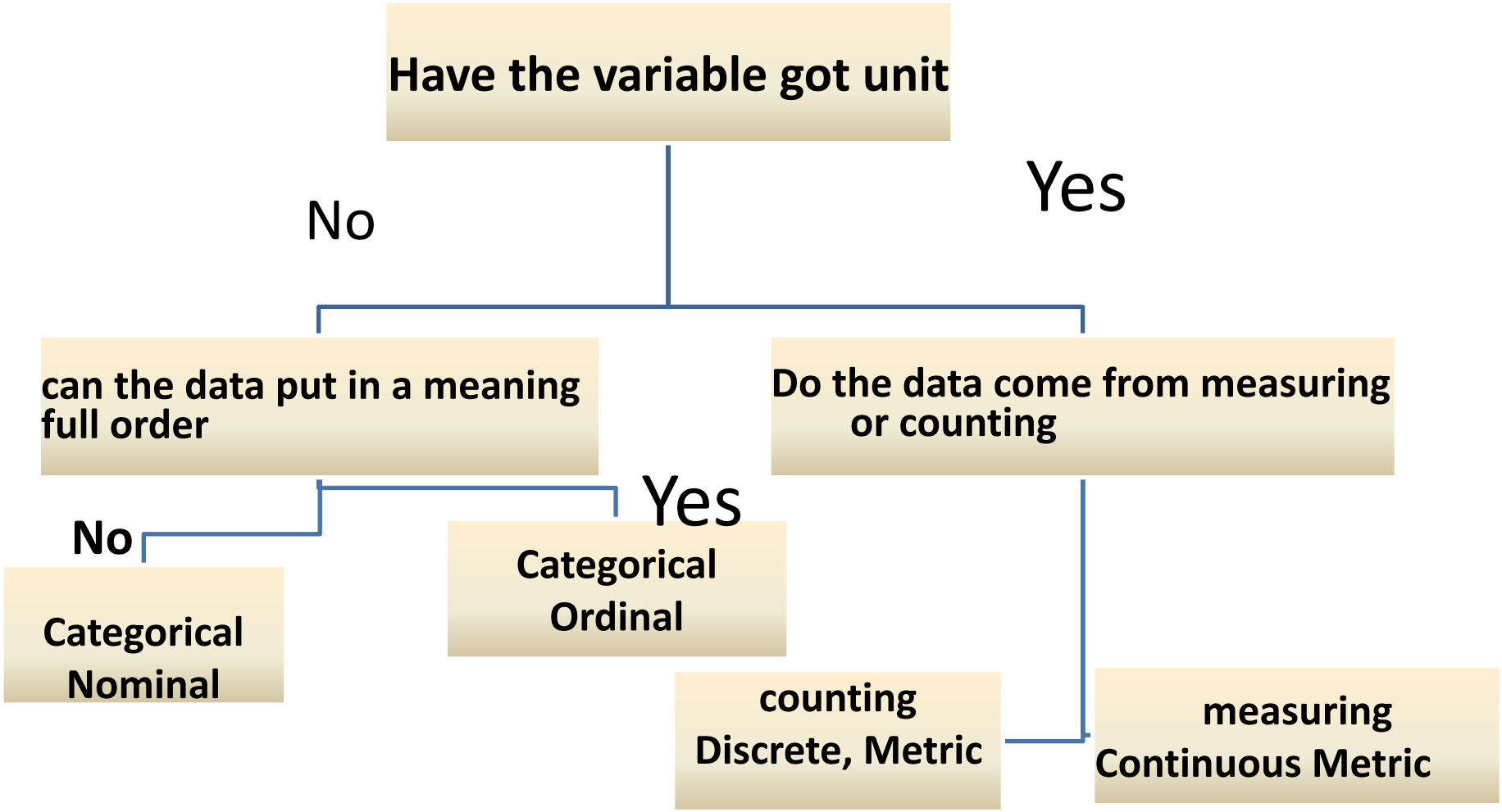
usually comes from counting such as

No. of death.
No. of students
No. of patients
??????

all are
discrete metric variable

- It is real numbers So……..???
- It can be counted
- It have a unit of measurements
- It is integer, measurement or values are integers

- They have the same interval and ratio properties as the continues variables



An important thing is the type of the variable concerned.

Quantitative Variable

The one that can be measured by the usual sense .

Qualitative Variable

The one which are not capable of being measured by the usual sense .

Biostatistics consist of

1-Collection of data .

2-Presentation of data .

3- Estimation of data

age of 50 patients

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

???????

Descriptive Statistics

organizing and summarizing data *and* ***bringing into a focus their essential characteristics***

Descriptive statis.

reduce the information to a manageable size

Presentation of Data

Data that collected from any source, are inadequate for planning .

Data need to be transformed into information

- **by reducing them,**
- **by summarization and**
- **Arrange it in a simple and useful way to**
- **bring out the *important point clearly & concise***

This mean that

display the important feature of the sample .

This is Descriptive Statistics

Descriptive Statistics

This one serve as devices for
organizing and summarizing data
and
bringing into a focus their essential characteristics
Descriptive statistics .
reduce the information to a manageable size

This include

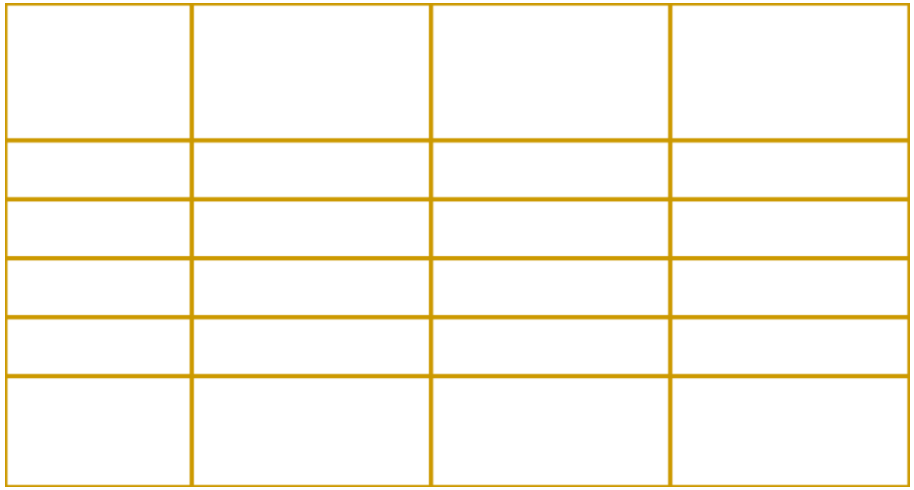
- **table**
- **graph, chart** or
- **Numerical Description**

An important thing is the type of the variable concerned.

Table

It is first step in data presentation .

Is the simplest and often most useful summary of data



An important thing is the type of the variable concerned.

Nominal Simple Frequency Table

example

Blood group of 95 children with leukemia shows as (22)A, (25)B, (18)AB, & O(30)

count the No. of observation in each category, these count are called Frequencies .

- **frequency**
- **Relative frequency**
- **percentage frequency**

} **of Bl. group distribution for of 95 children with leukemia**

An important thing is the type of the variable concerned.

Simple Frequency Table

Blood group	Frequency N=95	Relative Frequency	Percentage %
A	22	???	??
B	25		
AB	18		
O	30		
Total	95		

Nominal Simple Frequency Table continuo..

Relative frequency

Frequency each category divided by the total frequency.
No. of children of each category (Bg) divided by the total no of children.

Percentage frequency

Frequency of each category divided by the total frequency X 100

Relative or percentage frequency are often more useful than the actual number of individuals in each category.
???????

Simple Frequency Table

Blood group	Frequency N=95	Relative Frequency	Percentage %
A	22	22/95=0.231	23.15
B	25	0.26315	26.315
AB	18	0.18947	18.947
O	30	0.3157	31.5789
Total	95	?????	????

Relative or percentage frequency are often more useful than the actual number of individuals in each category. Why ????????

Type of feeding

Infants 600 600

Breast 478

Bottle 65

Mixed 57

Simple Frequency Table

Type of feeding	No. of cases (F)	R.F.	percentage %
Breast	478	0.79	79. 7
Bottle	65	0.108	10.8
Mixed	57	0.095	9.8
Total	600	????	100

b- ordinal Variable

120 individuals were asked about their level of satisfaction toward the health care given by Hospital X. The response as follows

29 very satisfied, 39 satisfied, 20 neutral

18 unsatisfied, 14 highly unsatisfied

level of satisfaction	Frequency N=120	Relative Frequency	Percentage %
very satisfied	29	0.24166??	24.166
satisfied	39	0.325 ??	32.5
neutral	20	0.1666	16.66
unsatisfied	18	0.15	15
highly unsatisfied	14	0.11666	11.66
Total	120	????	99.929

Ex.

The mathematic marks of 26 secondary school students at Amman in 2022

15.2	31.3	14.9	16.3	19.3	18.2	20.2	12.8	14.7
29.4	21.1	20.4	13.6	22.4	14.0	14.3	22.8	26.7
18.9	13.7	17.7	27.2	19.3	16.1	13.5	11.2	

? ? ? ?

Metric variable

Continuous variable

? ? ? ?

Continuous Metric variable

The most useful way for presenting data of CMV to produce grouped frequency distribution

❖ **grouping data first These group of data we call it class interval**

✓ **Each group of data (class interval)consist of values within certain range**

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
	26	

Continuous Metric variable

to produce grouped frequency distribution table

- Grouping data into groups of equal width
- then construct frequency distribution table for grouped data
- Counting the frequency of observation within the groups(class interval)
- Each group of data contain No. of observation

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
	26	

Use sturges rule :

$$K \quad 1 \quad 3.322 \ (\log N)$$

K= No. of class intervals. **N= sample size .**

Width of class intervals :

$$W \qquad \frac{R}{K}$$

W= width .

R= Range = highest – lowest .

K= No. of class intervals .

Continuous Metric variable

Frequency distribution of mathematic marks of 26 secondary school students at Amman in 2022

mathematic marks	Frequency	Cumulative frequency
10.0- 14.9	9	9
15.0- 19.9	8	17
20.0- 24.9	5	22
25.0- 29.9	3	25
30.0- 34.9	1	26
Total	26	

frequency distribution table

? ? ? ?

Example

The following data representing age (years) of 50 patients with diabetes Mellitus collected from Al Karak Hospital during march 2022

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

? ? ? ?
????????/

An important thing is the type of the variable concerned.

Age(year) of 50 patients with diabetes Mellitus attending Al Karak Hospital during march 2015

AGE year	Freq	Commul. frequency	Relative frequency	% R.F.	Cumul. R.F.	%cum Freq.
20-29	1		??	??		
30-39	2		???	??		
40-49	2		??			
50-59	3	?				
60-69	12					
70-79	14	?				
80-89	12					
90-99	4					
total	50	?	?	?	?	?

Relative Frequency (proportion)

Dividing the No. of values (observation, frequency) in a particular class interval by the total No. of values (observation frequency) in whole data

$$\frac{1}{50} \quad \frac{2}{50} \quad \frac{3}{50} \quad \frac{12}{50} \quad \frac{14}{50} \quad \frac{12}{50} \quad \frac{4}{50}$$

Percentage of Frequency

Dividing frequency of each class interval by the total No. of observation and then multiply by 100 .

$$\frac{1}{50}\,100 \quad \frac{2}{50}\,100 \quad \frac{3}{50}\,100 \quad \frac{12}{50}\,100 \quad \frac{14}{50}\,100 \quad \frac{12}{50}\,100 \quad \frac{4}{50}\,100$$

❑ Cumulative Freq. Dist.

- That is to convert the frequencies distribution into less than and more than .

❖ This is done by simply

- Adding two or more classes frequency
- Starting either at the top or at the bottom of the distribution .

1+2+ +2+3+12+14+12+4= 50

❑ Cumulative Relative and Percentage Dist.

- add two or more Relative frequencies together .

0.02+0.04+0.04+0.06+0.24+0.28+0.24+0.08=1

- Add the % instead of the frequencies, starting either at the top or at the bottom .

AGE year	frequency	Commutative frequency	Relative frequency	% R.F.	Cumulativ R.F.	%cum Freq.
20-29	1	1	0.02	2	0.02	2
30-39	2	3	0.04	4	0.06	6
40-49	2	5	0.04	4	0.1	10
50-59	3	8	0.06	6	0.16	16
60-69	12	20	0.24	24	0.4	40
70-79	14	34	0.28	28	0.68	68
80-89	12	46	0.24	24	0.92	92
90-99	4	50	0.08	8	1.00	100
total	50	---	1	100	---	---

Points should be keep in mind

1-No. of class intervals (5-15) .

2-Classes interval should notoverlapping .

3-All classes interval should have the same width across all data (constant width) .

4-There should be no gaps between class interval .

5-Every observation will be uniquely classifiable into one and only one class interval .

❑Class Marks

- ❖ It is the midpoint of the class interval .
- ❖ *It could be obtain by adding the lower and upper limits of a class interval and divided by two*


Thank You

Presentation of Data
table
graph, chart or
Numerical Description

Graphical Techniques

some times table presentation will give some difficulties to the reader especially to non numerical readers

- **Picture speaks lauder than thousand words .**
- **Graph have powerful impact on the imagination of population .**
- **Relationships, Trends** and **Contrasts** are often more **readily appreciated from diagram than table ..**

An important thing is the type of the variable concerned.