Biostatistics

Statistics :

is concerned with scientific methods for collection of data ,presentation of collected data and analysis of such data then making a valid conclusion and reasonable decision on the basis of such analysis

Biostatistics :

a application of the above mentioned methods in the field of biological sciences and medicine.

Statistical Methods

A) Methods Of Collection Of Data

B) Methods Of Presentation Of Data

C) Methods Of Analysis Of Data

A) Methods Of Collection Of Data

Sources of data	1. - Data collection through a comprehensive survey 2. - Data collection through sample survey 3. - Data collection through population census 4. - Data collection through hospital records 5. - Data collection through health office records 6. - Data collected through published vital statistics

TYPES OF DATA

A) Constant

B) Variable

1- Quantitative

- **A- Continuous (fraction, decimals)**
- **B- Discrete (integer)**

2- Qualitative

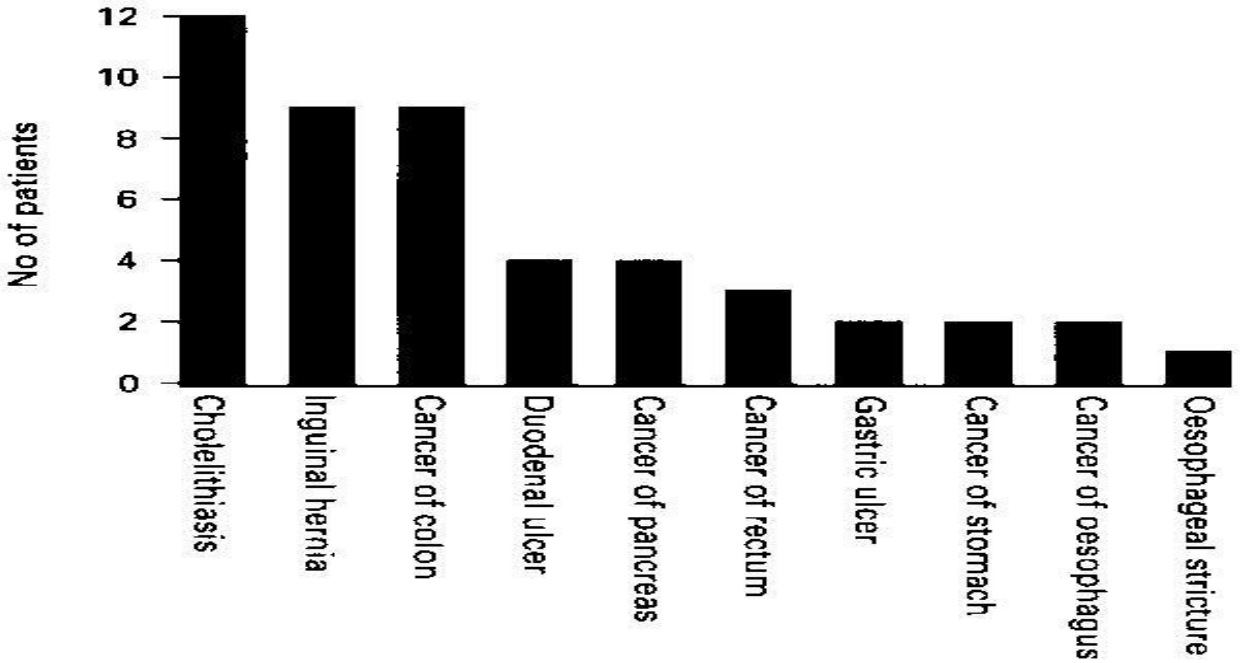
- **A- Ordinal (could be ranked or ordered)**
- **B- Nominal**

Classify these variables

1. SBP
2. Blood cholesterol
3. No. of blood transfusions in a series of renal transplant patients
4. Smoking habit (ever smoked, never smoked etc)
5. Educational attainment (primary, secondary etc)
6. Blood group

Summarising qualitative data

- Frequency count
- Bar chart



Summarising quantitative data

- Measures of central tendency
    1. Mean = average = $\Sigma x/n$
    2. Median = middle value
    3. Mode = most frequent
- Measures of Dispersion
    1. Range
    2. Interquartile range (25^{th}-75^{th})
    3. Standard deviation

B) Methods Of Presentation Of Data

1- Numerical

2- Graphical

3- Mathematical

NUMERICAL PRESENTATION OF DATA

1. **Simple Numerical presentation.**

(*un- grouped, un- classified*)

e.g. The weight of 5 children (8,7,9,4,3,5)

2. **Tabular presentation of data.**

The best and most convenient method for summarization of a large mass of data is using a table.

a) Simple frequency distribution table

1- for qualitative variables.

2- for quantitative variables.

b) Table of an association

Simple frequency distribution table for qualitative variables

(distribution of 5-10 years children with measles admitted to al-karak hospital during the year 2004 according to sex)

sex	No. of patient
Male	2550
Female	1550
total	4100

Simple frequency distribution table

- **Qualitative Variable**
- The left column is the variable, the right one is the frequency(number of observation for each category)
- Have a total at the bottom
- **Every table should have a title which answers three questions (what, who and when).**
- Source should be written below the table
- Example

Distribution of the third year medical students according to their degree of success in may 2007 in Mutah university faculty of medicine

Degree of success	No of students
Excellent Very good Good fair	120 200 350 180
total	850

Source: records of the faculty of medicine 2007

Quantitative variable
(continuous or discrete)

1. The smallest observation is chosen as the lower limits of the intervals.
2. The width or the size of the interval is 5,10 or 15
3. The number of interval counted to be in the permissible range of 4-12. to achieve the main advantage of tabular presentation namely summarization of a large mass of data.
4. The main disadvantage of tabular presentation is loss of precision in the presentation.
5. If less than 4 (loss of precision)we can reconstruct the table using a width less than 5
6. If > 12 (result in missing the summarization value of the table) we can reconstruct the table using a width > than 5

Duplication of the limits of intervals should be avoided by one of the 4 methods

A	B	C	D
10 to less than 15 15 to less than 20 20 to less than 25 25 to less than 30	10-14 15-19 20-24 25-29	10- 15- 20 25-30	10-14.9 15-19.9 20-24.9 25-29.9

1. Forms A, and C can be used in both continuous and discrete quantitative variables.

2. Form B can be used in discrete quantitative variables

3. While form D is only used in continuous quantitative variables.

Example for continuous variable:

- Distribution of 100 medical student (1985) according to their Hb %

Hb %	No of patients
65- 70- 75- 80- 85- 90- 95<100	22 15 19 11 17 12 4
total	100

Each Interval In Table Has The Following.

1. Lower limit
2. Upper limit
3. Width
4. Midpoint
5. Number of observation
6. Real limits (± 0.5)

First Interval:

Lower limit (L.L)= 65

Upper limit (U.L)= 69.999, practically= 70

Width(continuous quantitative variables)=

U.L - L.L= 70 – 65 = 5

$$\text{Mid point} = \frac{\text{L.L} + \text{U.L}}{2} = \frac{65 + 70}{2} = 67.5$$

Number of observations = 22

Hb %	No of patients
65- 70- 75- 80- 85- 90- 95<100	22 15 19 11 17 12 4
total	100

***Example for discrete quantitative variable:**

Distribution of patients in hospital (X) according to their family size in 1984.

Family size	No. of patients
3-4 5-6 7-8 9-10 11-12	28 55 32 18 12
Total	**145**

Last interval:

L.L =11

U.L =12

Width (discrete quantitative variables) = (U.L – L.L + 1) =

(12 – 11 + 1) = 2 **(No. of counts within the interval e.g.**

(11 & 12 = 2)

Mid point = $\frac{11+12}{2}$ = 11.5

Number of observation = 12

N.B: in construction of the table, we should avoid open ended table I.e. lower limit of first interval and/or upper limit of last interval are missed or unknown

Table Of An Association Or Contingency Table.

A) Two by two table:

i.e Two columns by two rows, it is used to show relation between a condition and characteristic e.g relation between smoking and lung cancer

Example: relation between smoking and lung cancer

Smoking Status	Lung cancer		Total
	Yes	No	
Smoker	35	15	50
None smoker	65	85	150
Total	100	100	200

B) c x r table: i.e. more than two columns by two or more than two rows

Example: Different types of treatment of disease (X) and outcome.

Treatment	outcome			Total
	Cured	Improved	died	
A	35	12	3	50
B	30	10	10	50
C	33	12	5	50
Total	98	34	18	150

Comparing Frequency Distribution Table

i.e distribution of two different groups according to one variable.

N.B : For comparison, the total groups should have the same total frequencies otherwise calculate the percent of total for each frequency.

Examples: distribution of low birth weight (L.B.Wt)

babies and normal babies according to their mother's age.

Age of mother (years)	L.B.Wt	Normal	Total
20- 25- 30- 35- 40<45	22 13 15 20 30	30 39 18 8 5	52 52 33 28 35
Total	100	100	200

Two Way Table Or Two Way Classification

i.e One group is classified according to two variables e.g weight and height or age and blood pressure to find correlation between these two variables.

Example: Distribution study by weight and height

Weight (KG)	Height (cm)					Total
	150-	155-	160-	165-	170<175	
60-	18	5	2	-	-	25
65-	15	8	7	2	-	32
70-	11	15	8	2	2	38
75-	7	15	10	1	3	40
80<85	-	2	5	2	7	15
Total	**51**	**45**	**32**	**10**	**12**	**150**