RESEARCH AND BIOSTATISTICS

CONTENT

- What is Bio-Statistics?
- Definition of Statistics
- Distinction between
  - Data, Information, Statistics
- Types of data
- Statistical symbols
- Collection of data
- Organization of data
- Presentation of data
- Analysis of data
- Scopes and Applications

What is Biostatistics ? (biometrics or biometry)

- Development and application of statistical techniques to scientific research relating to life:
  - **Human, Plant and Animal**
- Here the focus is on Human life and Health.

Thus the areas of application relates to:

- **Pharmacology**,
- **medicine**,
- **epidemiology**,
- **public health**,
- **Physiology and anatomy**,
- **Genetics**

Essence of Bio-Statistics?

As the knife is to surgery so is the Bio-statistics to Medical Research

What is Statistics? (Definition)

- The Statistics is defined differently by different authors from time to time.
- The reasons being:

  Its scope and utility has widened considerably over time.
- It is defined in two ways:

  as “statistical data” and

  “statistical methods”

DEFINITION OF STATISTICS AS A BODY OF SCIENCE

American Heritage Dictionary® defines statistics as: "The mathematics of the Collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling.“

The Merriam-Webster’s Collegiate Dictionary® definition is: "A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data."

A Simple but Concise definition by Croxton and Cowden:

"Statistics is defined as the

- Collection,
- Presentation,
- Analysis and
- Interpretation of numerical data."

DATA, INFORMATION AND STATISTICS

DATA

"Facts or figures from which conclusions can be drawn".

- Data can take various forms, but are often numerical.
- Data can relate to an enormous variety of aspects, for example:

• the daily weight measurements of each individual in your classroom;
• Blood pressure and pulse recording of patients at regular interval

INFORMATION:

"Data that have been recorded, classified, organized, related, or interpreted within a framework so that meaning emerges".

Information, can take various forms:

- The number of persons in a group in each weight category (20 to 25 kg, 26 to 30 kg, etc.);
- Number of Adolescents having menarch cycle within <20 days, 20-25 days, 25 – 30 days, etc.
- Number of patients responding to the drug within 3-5, 5- 7 & more than 7 days of therapeutic dose

- *Just as trees are the raw material from which paper is produced, so too, can data be viewed as the raw material from which information is obtained.*

STATISTICS

 A statistics is "a type of information obtained through mathematical operations on numerical data".

- The average weight of people in your classroom;
- The average number of in-door patients per day in a month for 12 months.
- The minimum and maximum blood pressure observed each hour of critical patient.

Data collected on the weight of 20 individuals in your classroom

Data	Information	Statistics
20kg	5 individuals in the 20-to-25-kg range	Mean weight = 22.5 kg
25kg		
28kg	15 individuals in the 26-to-30-kg range	Median weight = 28 kg
30kg		
….		
etc		

TYPES OF DATA

- **Qualitative Data: They are Nonnumeric.**

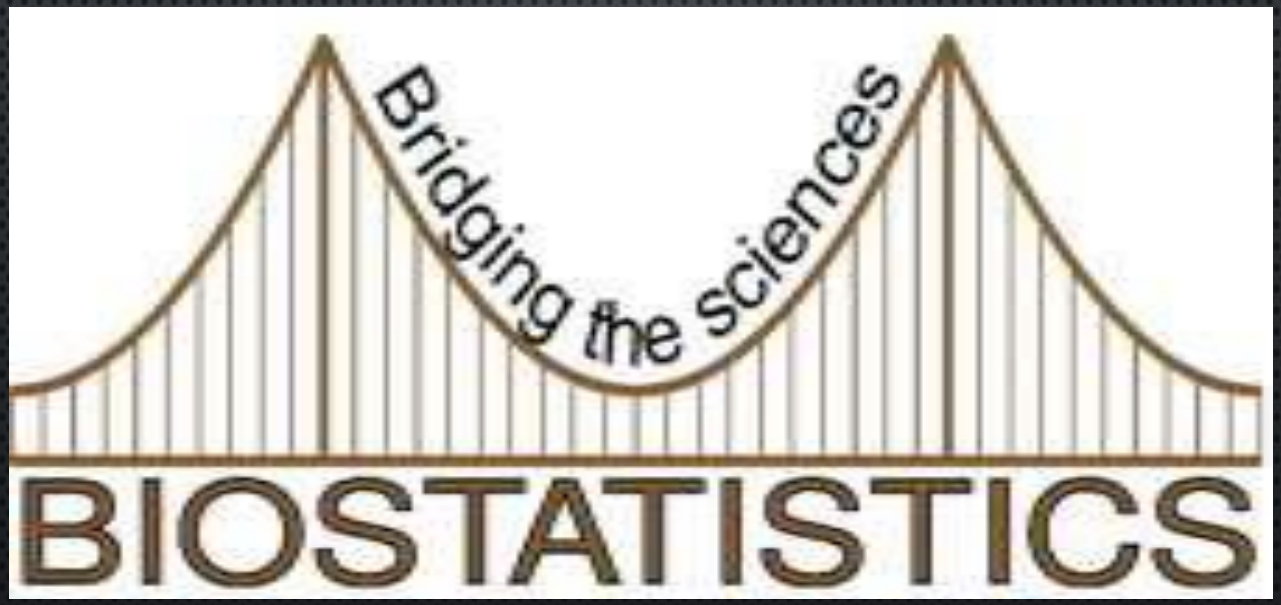
In form of words such as

- **Description of events: *Verbal Autopsy***
- **Transcripts of interview: *Opinion of CDMOs on Efficacy of ASHA***
- **Written documents: Guidelines on Roles and Responsibility of ASHA**
- **Characteristic not capable of Quantitative measurement: *Sex of a patient, Colour, Odour***

- **Quantitative Data: They are numeric.**

The characteristic capable of quantitative measurement: Height, Weight, Blood pressure etc.

Statistical Symbols



Sample statistic	Population parameter	Description
n	N	number of members of sample or population
$\bar{x}$ "x-bar"	μ "mu" or μ	mean
M or Med or $\tilde{x}$ "x-tilde"	none	Median
s (TIs say Sx)	σ "sigma" or σ	Standard deviation For variance, squared symbol (s^2 or σ^2).
r	ρ "rho"	coefficient of linear correlation
$\hat{p}$ "p-hat"	p	proportion
z t χ^2	(n/a)	calculated test statistic

COLLECTION OF DATA

This is the first and most important stage of statistical investigation.
Since the data are the raw material to any statistical investigation, utmost care is necessary to be taken for

collection of reliable and
accurate data.

The first hand collection of data is very difficult.
Therefore the investigator should see if the data intended for the purpose is already collected by some other agencies. This would help
save time,
money and
duplication of effort.

ORGANISATION OF DATA

The data collected from a published secondary source is in organised form. The large mass of data collected from a survey need to be organised.

The first stage in organising data is editing.
Editing for completeness,
Editing for inconsistencies,
Editing for homogeneity,
Editing for accuracy,
Editing for reliability,

The 2nd stage in organising data is classification of data.
That means arranging data according to certain common characteristics of the units on which data are collected.

The final step is the tabulation.
The purpose is to present data in rows and columns so that there is absolute clarity in data to be presented

PRESENTATION OF DATA:

The presentation of data is an art and science. It facilitate

statistical analysis,
comparison and
appreciation.
The data can be presented through:
Diagrams, Graphs, Pictures etc.

ANALYSIS OF DATA:

Analysis of data are done through;
Simple observation,
Application of simple and

highly sophisticated statistical techniques like;
Measures of central tendency,
variation, correlation, regression,
seasonality, trend, computation of indices etc.

Functions of Statistics

- It presents facts in a definite form
- It simplifies mass of figures
- It facilitate comparison
- It helps in formulating and testing hypothesis.
- It helps in prediction
- It helps in formulation of suitable policies.

SCOPE AND APPLICATIONS OF BIOSTATISTICS

•**In Physiology And Anatomy**
O To define what is normal or healthy in a population.
O To find the limits of normality in variables such as weight and pulse rate etc. in a population.

To compare the efficacy of a particular drug, operation or line of treatment – for this, the percentage cured, relieved or died in the experiment and control groups, is compared and difference due to chance or otherwise is found by applying statistical techniques.

• **To find an association between two attributes such as cancer and smoking or filariasis and social class**–an appropriate test is applied for this purpose.

•Deficiency of iodine as an important cause of goiter in a community is confirmed only after comparing the incidence of goiter cases before and after giving iodized salt

➢ The methods used in dealing with statistics in the fields of medicine, biology and public health for planning, conducting and analyzing data which arise in investigations of these branches.

➢ The most common areas where one can find an extensive applications of statistical methods in human genetics is
* Human Genome Project
* Linkage Analysis
* Sequencing

When you can measure what you are speaking about and express it in numbers, you know something about it. but when you can't measure it,
when you can't express it in numbers your knowledge is a meager and unsatisfactory one.

Lord Kelvin

REFERENCES

- Biostatistics – A foundation for Analysis in the Health Sciences: Wayne W. Daniel, Seventh Edition, Wiley Students Analysis.
- A First Course in Statistics with Application: A.K. P. C Swain, Kalyani Publishers
- Methods in Biostatistics sixth edition : BK Mahajan
- Fundamentals of Biostatistics : Sanjeev BS Rao KV. Biostatistics: A manual of statistical methods for use in health, nutrition and anthropology. 2nd ed. New Delhi: Jaypee Brothers Medical Publisher (P) ltd; 2007.


Thank you...