

## Statistics

Statistics is a very broad subject, with applications in a vast number of different fields. Statistics is the science of dealing with uncertain phenomenon and events. Statistics in practice is applied successfully to study the effectiveness of medical treatments.

### What is Statistics?

In generally, statistics is the methodology for collecting, analysing, interpreting and drawing conclusions from information and collected data.

**or**

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation and presentation of numerical data.

### What is a Statistician?

At a high level, statisticians are professional who apply statistical methods and models to real- world problems. Statisticians make sense of information collection about the world.

### Why Statistics?

- To develop an appreciation for variability and how effect product, process and system.
- It is estimating the present, predicting the future.
- Study methods that can be used to solve problems, build knowledge.
- Statistics make data into information.
- Develop an understanding of some basic ideas of statistical.
- Reliability, stochastic process (probability concepts).
- Statistics is very important in every aspects of society (govt. people or business).

### What is Biostatistics?

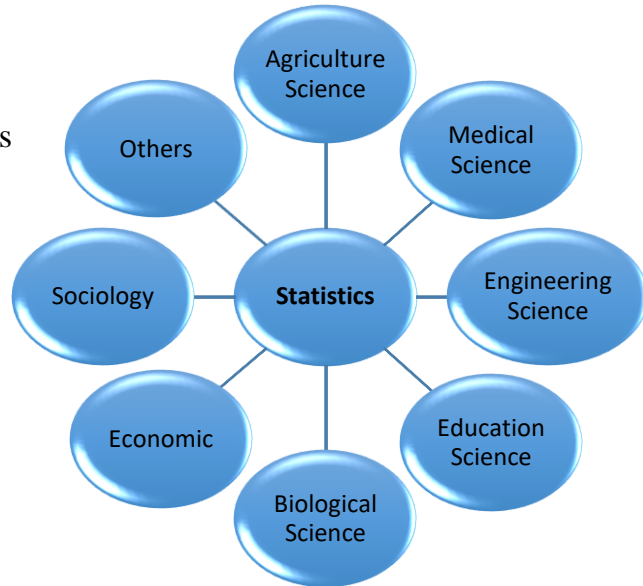
**Biostatistics** is the branch of statistics responsible for the proper interpretation of scientific data generated in the biology, agriculture, public health and other health sciences (i.e., the biomedical sciences).

**Biostatistics** is the application of statistics to a wide range of topics in biology. The science of biostatistics encompasses the design of biological experiments, especially in medicine, pharmacy and agriculture; the collection, summarization, and analysis of data from those experiments; and the interpretation , and inference from, the results.

## Relationship between statistics and other sciences

The sciences without statistics that means nothing. There are a significant relation between the statistics and most or all other sciences which can be expressed as follow

Figure: Relationship between statistics and other sciences



Statistics is the science of gaining information from numerical and categorical data. Statistical methods can be used to find answers to the questions like:

- What kind and how much data need to be collected?
- How should we organize and summarize the data?
- How can we analyse the data and draw conclusions from it?

### Functions of Statistics

1. Definiteness
2. Condensation
3. Comparison
4. Formulating and testing of hypothesis
5. Prediction or Forecasting
6. Planning and Decision

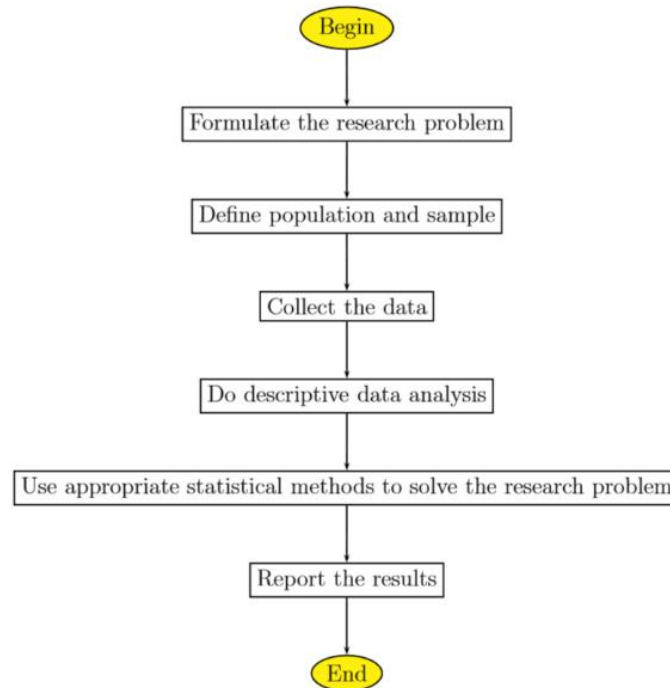
There are two major types of statistics:

1. Descriptive statistics
2. Inference statistics

The branch of statistics devoted to the summarization and description of data is called descriptive statistics and the branch of statistics concerned with using sample data to make an inference about a population of data is called inferential statistics.

## Statistical data analysis

The goal of statistics is to gain understanding from data. Any data analysis should contain following steps:



## Variables and organization of the data

### Variables

**Variable:** as x, y, ... z or character which shows variability among individuals called variable like student's length, student's weight, Tree's height, number of flowers per plant...etc. The variable can be divided in to two types:

#### A- Qualitative or (categorical) variable:

A variable which can't be measured in quantitative form. But can only be identified by name or categories or it is a variable which its observations result in non-numerical form like: place of birth, types of drug, stages of breast cancer (I, II, III, or IV), degree of pain (minimal, moderate, severe) and Gender of child (male, female)

#### B- Quantitative variable:

It is a variable in which its observations are in numerical measurements like, blood pressure value, Serum cholesterol, plant height, germination%. Number of oranges /tree, number of leaves /plant number of branches /tree. Quantitative variable or characters may also be classified to:

**1. Continuous Variables:** Means that all possible values are observing on continuous scale like the blood pressure of 5 students are (11, 12,13,14 and 15 cm Hg).

**2. Discontinuous (discrete) Variables:** Means that all possible values are not observing on a continuous scale. like the blood cholesterol of 5 students are (179, 181, 189, 191, 190). Or the organic matter% content of 5 forest soils are :3 ,4 ,6 ,7 and 8.

## Population and Sample

**Population:** Is the number of individuals or subjects having one or more variables in common. Like the students of Salahaddin university, the population that is being studied is called target population. There are two types of population:

**1-Finite population:** It is a population which its number can be count like the number of students in Salahaddin university.

**2-Infinite population:** It is a population which its number cannot be count like the number of fishes in Dukan lake.

**Sample:** It is a part of population. A sample that represent the characteristics of population called representative sample.

**Estimates:** Are the measures calculated from samples such as sample mean ( $\bar{X}$  =  $\bar{x}$  -) and sample variance ( $S^2$ ).

**Parameters:** Are the measures determined from populations such as population mean ( $\mu$ ).

**Data:** is a set of observations. One observation called datum or measurement.

**Grouped data:** Is data after arrangement.

**Raw data:** is data without or before arrangement.

## There are some symbols of statistics

X, Y, ...Z (Capital): Variable

$X_i$  : Individual observation

$\bar{X}$  : Sample mean, is pronounced as (x-bar)

$\mu$ : Population mean is pronounced as (mu)

n : Sample Size

N : Population size

$S^2$ : Sample variance

$\sigma^2$ : Population variance

S : Sample standard deviation

$S_x$  : Standard error or standard deviation of mean

$\sigma_x$  : population standard error

$\Sigma$  : Summation (Sigma)

$M_e$  : Median

$M_o$  : Mode

R : Range

C.V.:Coefficient of variation

$f_i$ : Frequency

$\alpha$ : Alpha

SS : Sum of squares

ANOVA: Analysis of Variance

df : degree of freedom

r : Regression

b : Correlation