

# **Introduction to statistics**

**By- Dr. Ekta Khare**

# What is Statistics?

Statistics is a branch of applied mathematics which deals with *the collection, classification, analysis and interpretation of data*. The word statistics is derived from the Latin word 'status' means a 'political state' or 'government'.

- The field of statistics has two main areas:
  - mathematical statistics and
  - applied statistics.
- Mathematical statistics concerns the development of new methods of statistical inference and requires detailed knowledge of abstract mathematics for its implementation.
- Applied statistics involves applying the methods of mathematical statistics to specific subject areas, such as economics, psychology, and public health

# What is Biostatistics?

Biostatistics is a branch of biological science which deals with the study and methods of collection, presentation, analysis and interpretation of data of biological research. Biostatistics is also called as biometrics since it involves many measurements and calculations. In biostatistics, the statistical methods are applied to solve biological problems. Basic understanding of biostatistics is necessary for the study of biology particularly doing research in biological science. *The statistics will help the biologist to: (1) understand the nature of variability and (2) helps in deriving general laws from small samples.*

**Francis Galton** is called as the '**Father of Biostatistics**'. He created the statistical concept 'correlation'. Sir Galton for the first time used statistical tools to study differences among human population. He also invented the use of questionnaires and surveys for collecting data on human communities.

- Morphometrician and anthropologist Prasanta Chandra Mahalanobis is considered as the father of statistics in India. He was one of the founding members of the erstwhile Planning Commission of India and founded Indian Statistical Institute, Kolkata.

### ***Steps in biostatistics:***

A biostatistical investigation is carried out through the following sequential steps.

- (1). Collection of data (variable)
- (2). Classification of the collected data
- (3). Analysis of data
- (4). Interpretation of data

# Scope of Biostatistics

- The scope of biostatistics is extensive and cover almost the whole of biology that deals with generation and analysis of numerical data.
- Biostatistics is used right from designing scientific experiments through the data analysis.
- The scope includes:
  - principles of scientific methodology,
  - defining various types of data and studies,
  - levels of measurements,
  - descriptive statistics,
  - inferential statistics and hypothesis testing, and correlation.
- The field also includes various predictive methods and curve/model-fitting including regression analysis, maximum likelihood, Bayesian Inference and Principal Component Analysis.

## *Importance of Statistics in Biological Science*

Biostatistics has applications in all the branches of life sciences. Few applications of biostatistics are summarized below.

### **(1). Research**

- Ø Research is incomplete without the statistics
- Ø Every result (data) in the research need to be statistically validated.
- Ø For the design of experiments
- Ø Selecting the method of collection of data
- Ø Deriving logical conclusions from the data
- Ø Deriving single values from a group of variables

### **(2). Medical and Pharmaceutical Science:**

- Ø For checking the efficiency of drugs
- Ø To find out the possible side effects of drugs
- Ø For conduction of drug treatment trials

**(3). Genetics:**

- Ø Study the inheritance patterns of genes
- Ø Essential for the study of Mendelian genetics
- Ø For the studying the genetic structure of a population
- Ø Studying the behavior of genes in a population (**Population Genetics**)
- Ø For the mapping chromosomes

**(4). In numerical taxonomy (taxonomy with numbers)**

**(5). For monitoring the community and public health**

**(6). Demography (study of human population)**

**(7). Statistics helps to implement policies by the government**

**(8). Used in weather forecasting**

### *Limitations of Statistics*

- (1). Statistical laws are true on average. Single observation is not a statistic.
- (2). Statistics cannot be applied to single / individual data.
- (3). Statistical methods are best applicable on quantitative data.
- (4). Statistical method cannot be applied to highly heterogeneous data.
- (5). If sufficient care is not exercised in collecting, analyzing and interpreting the data, the statistical results might be misleading.
- (6). Only a person expert in statistics can handle the statistical tools efficiently.
- (7). There are too many methods to study a single problem in statistics
- (8). Statistics does not depict the entire story or the phenomenon.
- (9). Statistical results are not always beyond doubt.
- (10). Some errors are possible in the statistical decisions.
- (11). We do not know whether an error has been committed or not.