

INTRODUCTION TO STATISTICS

'Statistics' means numerical information expressed in quantitative terms, which may relate to objects, subjects, activities, information, phenomena, or regions of space.

Generally, there are two types of information i.e., quantitative and qualitative information. Statistics is the part of mathematical science which pertains to the collection, classification, tabulation, analysis, interpretation and presentation of data.

The word 'Statistics'

The word of statistics has been derived from the 'status', which is Latin word OR 'Statista' which is Italian word. In the 18th century, **Prof. G. F. Achenwall** has been used it first time. These words were used for political state of the region during early period. The Italian word 'Statista' was used to maintain the records of census or data related to wealth of a state/ nation. Successively, the meaning and usage of statistics extended and there onwards its nature also changed.

■ MEANING OF STATISTICS

Singular Sense

In this sense, 'Statistics' refers to what is called statistical methods, which deals with the collection, classification, tabulation, analysis and interpretations of statistical data. So, it is described as a branch of science which deals with numerical facts and make decision as well. Every statistical process should pass through these stages.

Plural Sense

In plural sense, 'Statistics' means a set of numerical information known as statistical data, i.e., information on population features, climatic variable, enrolment of students in a class, and so on. According to Websters "Statistics are the classified facts representing the conditions of the people in a State specifically those facts which can be stated in number or in tables of number or classified arrangement".

■ TYPES OF STATISTICS

There are the **two** broad ways of classifying statistics,

- one is the **on the basis of function** and
- another is the **on the basis of distribution of data**.

■ **Statistics Based on Function:** There are three types of statistics on the basis of subject matter/ function. The types of statistics based on function are -

1. **Descriptive Statistics** - Descriptive Statistics is the branch, which deals with descriptions of obtained data. It is a summary statistic which summarizes features/characteristics from a collection of information. Moreover, it includes classification, tabulation, measurement of central tendency as well as variability. The researchers use of these measures to understand about the tendency of data/scores. Which further enhance the ease in description of the phenomena.
2. **Correlational statistics** - In correlational statistics, the obtained data are disclosed for their inter correlations and it includes various types of methods to compute the relationship (correlations) among data. It also provides description about sample or population for their further analyses purpose to explore the significance of sampling and population averages.
3. **Inferential statistics** - Statistical inference (SI) is the process of data analysis to deduce properties of probability distribution. Inferential statistical analysis infers properties of a population or census through the testing hypotheses and deriving estimates which is based on the primary assumption i.e., the observed data set is sampled from a larger population. It is also deals with the drawing of conclusions about population/census. Moreover, it provides technique to compute the probabilities of future behaviour of the subjects/areas.

▪ **Statistics Based on Distribution of Data:** There are two types of statistics i.e., parametric and nonparametric statistics on the basis of distribution of data.

1. **Parametric Statistics** - It is defined to have an assumption of normal distribution for its population / census under subject / study and refers to those statistical methods / techniques that have been developed on the assumption that the data are of a certain type i.e., It follows the normal probability curve (NPC). The measure under parametric statistics should be an interval scale and the scores should be drawn from a normal distribution (NPC)". There are certain basic assumptions of parametric statistics. The property of normal distribution is the very first characteristic of parametric statistics. Some important parametric statistics used for data analysis are T-test, F-test, r-test, and Z-Test.

2. **Nonparametric statistics** - The non-parametric statistics are not based on the assumption of normal distribution of population (NPC) and it is also known as distribution free statistics. It means that the non-parametric statistics is not follows the normal probability curve. The nonparametric statistics are not bound to be used with interval scale data. When the nature of data is non-continuity and difficult to maintain the assumption of normal distribution, then we used the nonparametric statistics. Moreover, the nonparametric statistic scan be used even for nominal data along with the ordinal data. Some important nonparametric statistics are chi square, Spearman's rank, Kendall's ranked and Mann-Whitney U test.

▪ IMPORTANCE OF STATISTICS

• **Policy Planning:** To finalise a government or individual policy, it requires some relevant data from previous documents or expected environment that the policy can be effectively utilised with maximum favourable benefits / results.

• **Behavioural and Social Sciences:** In social sciences specially in Economics, the both types of information i.e., quantitative and qualitative are used to analysis and draw policy recommendations. Moreover, statistics helps the academicians/ researchers to alter the information in a comprehensive way to analysis and predict the patterns of behaviour or trends.

• **Pure and Mathematical science:** The tools of statistical are also used to have precise measures in pure and mathematical sciences and to see differences on different occasions in various conditions.

• **Education Sector:** The statistical tools or instruments is also used in the area of education. Statistics used to create patterns and trends of variables on the basis of past and present conditions and hence showing the direction of development in education sector. Further, these trends help to create the policies and planning of the education.

• **Management and Commerce:** Statistics is very useful tool in management and commerce. It organises the various aspects of work and wellbeing of the employees. It is also a very useful instrument for account, which is the branch of commerce. Moreover, it also keeps an eye on the progress trend of the organisation.

• **Industries and Service:** Statistics is a basic tool to analysis the progress of industry as well as service sector and it also helps to make further strategies for the development of these sectors.

• **Problem Solving Technique:** Statistics provides the problem-solving tool between two or more variables. To find out the best applicable solution to a problem situation, we can use statistical technique and it is possible because of statistics.

• **Theoretical Researches:** On the basis of statistical analysis, we can establish the significance relationship of those facts for a particular paradigm or phenomena, which theories evolve the facts obtained from the field.

▪ LIMITATIONS OF STATISTICS

• The first serious limitation of Statistics is that, it deals with aggregate of facts and not with single observation. Therefore, the methods of statistics do not give any recognition to an individual person/ object/ event.

• The next limitations of statistics is that it deals mostly with numerical data. So, statistics is more applicable to those phenomena which can be measured quantitatively not qualitatively. However, qualitative phenomenon can be applied indirectly through some statistical techniques.

• The third important limitation of statistics is that, its conclusions are true only on the average or aggregates. Therefore, statistical inferences may not be considered as exact like conclusions based on Mathematical laws or Mathematical sciences.

▪ DISTRUST AND MISUSE OF STATISTICS

Sometimes, the statistical tools have been used by the irresponsible and inexperienced people to fulfil their self-reason / motive. The various misuses of statistical tools sometimes called an unscrupulous science. There are the following misgivings about Statistics;

- “*Statistics can prove anything*” — Bowley
- “*Statistics is an unreliable science*”
- “*Facts are stubborn things, but statistics are pliable.*” — Mark Twain
- “*An ounce of truth will produce tons of Statistics*”
- “*Figures do not lie, liars figure.*”
- “*There are three types of lies, namely, lies, damned lies, and statistics.*” — Benjamin Disraeli

So, we should be taken care and precautions for the interpretation of statistical data and its results.

“Statistics should not be used as a blind man uses a lamp-post for support instead of illumination” - Andrew Lang

Selected Readings / Reference

- ‘Introduction to Statistics’, Dr. Sanjeev Kumar, CCU, Meerut
- Gupta, S.C.(1990) Fundamentals of Statistics. Himalaya Publishing House, Mumbai
- Jay L. Devore: Probability and Statistics for Engineering and the Sciences, Cengage learning,
- Murray R. Spiegel : Theory & Problems of Statistics, Schaum’s publishing Series.