
UNIT 1 BASIC CONCEPTS

Structure

- 1.0 Objectives
- 1.1 Introduction
- 1.2 Meaning of Statistics
 - 1.2.1 Statistics in Plural Sense
 - 1.2.2 Statistics in Singular Sense
 - 1.2.3 Meaning of the Word 'Statistic'
- 1.3 Importance of Statistics
 - 1.3.1 Statistics and Economics
 - 1.3.2 Statistics and Business
 - 1.3.3 Statistics and Physical Sciences
 - 1.3.4 Statistics and Mathematics
 - 1.3.5 Statistics and other Social Sciences
- 1.4 Misuses of Statistics
- 1.5 Limitations of Statistics
- 1.6 Let Us Sum Up
- 1.7 Key Words
- 1.8 Some Useful Books
- 1.9 Answers or Hints to Check Your Progress Exercises

1.0 OBJECTIVES

After going through this Unit you will be able to:

- appreciate the importance of statistics in our life;
- explain some basic concepts used in the study of statistics;
- define statistics in both singular as well as the plural sense; and
- identify the uses and misuses of statistics.

1.1 INTRODUCTION

Now-a-days the word 'statistics' has become a household word, although different people comprehend it in different senses. The modern educated person has to be a person of statistics, broadly understanding its meaning and applying it to his/her life in different ways. For example, everyday we come across different types of quantitative information in both print as well as electronic media on topics like population, exchange rate fluctuations, inflation rate, day and night temperatures (being below or above normal; lowest or highest in the century or in the last thirty years or so), etc. In order to improve our understanding of the world around us, it is necessary to:

- a) measure what is being said,
- b) express it numerically, i.e., in numbers/quantities like weights in so many kilograms, eggs in so many dozens, etc., and

- c) utilize quantitative information or expression to draw conclusions and suggest policy measures.

Needless to say that if we cannot measure and express, in terms of numbers what is being said, then our knowledge will remain insufficient and far from being satisfactory. Statistics thus involves some sort of numerical information called “numerical data” or simply “data”. For example, one may give a statement that he/she has studied statistics (that is quantitative information) on absenteeism among the educated and the uneducated workers in Indian industries and found that incidence of absenteeism is more among the latter. He/she is referring to the numerical figures or numerical information technically called data.

Other examples of data are:

- a) India is suffering from population explosion, annual growth of population being around 2%.
- b) Students of XIIA have shown a better result than those of XIIB because the average marks of the former are 25% more than the average marks of the latter.
- c) Foreign exchange reserve of the country has been the highest so far since independence and stood at \$ 110 billion.
- d) As per the 2001 census population of India was 1027 million.

Many more such examples can be found and the students are expected to go through this exercise on their own.

History of Statistics

The word Statistics is the modern form of the word *statistik* which in turn has been derived from the Italian word “statista” meaning “statesman”. Professor Gott Fried Achenwall used it in the 18th century. It was Dr. E.A.W. Zimmerman who introduced the word statistics into England.

Early government records show statistical information on some aspects of population, land records, military strength of different wings, mortality during epidemics and so on. Perhaps it was because of this that Statistics was called the science of kings. But as the humanity developed, the usage as well as understanding of Statistics increased and now it is difficult to imagine a field of knowledge which can do without statistics. In fact, it has become an important tool of analysis.

1.2 MEANING OF STATISTICS

Let us look into the meaning of the word ‘Statistics’. It conveys different meaning to different people. A common man may simply interpret it as a mass of figures, graphs or diagrams relating to an economic, business or some other scientific activity. However, for an expert, it may also imply a *statistical method of investigation* in addition to a mere mass of figures. Let us discuss each of these.

1.2.1 Statistics in Plural Sense

Statistics in plural sense means the mass of quantitative information called ‘data’. For example, we talk of information on population or demographic features of India available from the Population Census conducted every ten years by the Government of India. Similarly, we can have statistics (quantitative data or simply data) on

enrollment of students in a particular university, say, over the last ten years. Further, data are collected by almost all ministries of the Government of India relating to their activities.

Also referred to as Statistical Data, Horace Secrist describes statistics in plural sense as follows:

“By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other.”

This definition of statistics in plural sense highlights the following features:

- a) *Statistics are numerical facts*: In order that information obtained from an investigation can be called as statistics or data, it must be capable of being represented by numbers. The collected data may be obtained either by the measurement of characteristics (like data on heights, weights, etc.) or by counting when the characteristics (like honesty, smoking habit, beauty, etc.) is not measurable.
- b) *Statistics are aggregates of facts*: Single and unrelated figures even though expressed as quantities are not statistics. For example, in a university examination Mr. Sharma secures 65% marks does not make statistics or data. However, if we find that out of 3 lakh university students whose average marks were 55%, Mr. Sharma secured 65% marks, then these figures are statistics. So no single figure in any sphere of statistical inquiry, say production, employment, wage and income constitutes statistics.
- c) *Statistics are affected to a marked extent by multiplicity of causes*: In physical sciences it is possible to isolate the effect of various forces on a particular event. But in ‘Statistics’ facts and figures, that is, the collected information, are greatly influenced by a number of factors and forces working together. For example, the output of wheat in a year is affected by various factors like the availability of irrigation, quality of soils, method of cultivation, type of seed, amount of fertilizer used, etc. In addition to this there may be certain factors which are even difficult to identify.
- d) *Statistics are numerically expressed*: Statistics are statements of facts expressed numerically or in numbers. Qualitative statements like “the students of a school ABC are more intelligent than those of school XYZ” cannot be regarded statistics. Contrary to this the statement that “the average marks in school ABC are 90% compared with 60% in school XYZ, and that the former had 80% first division compared with only 50% in the latter”, is a statistical statement.
- e) *Statistics are enumerated or estimated with a reasonable degree of accuracy*: While enumerating or estimating statistics, a reasonable degree of accuracy must be achieved. The degree of accuracy needed, in an investigation, depends upon the nature and objective of investigation on one hand and upon the time and resources on the other. Thus it is necessary to have a reasonable degree of accuracy of data, keeping in mind the nature and objective of investigation and availability of time and resources. The degree of accuracy once decided must be uniformly maintained throughout the investigation.
- f) *Statistics are collected in a systematic manner*: Before the collection of statistics, it is necessary to define the objective of investigation. The objective

of investigation must be specific and well defined. The data are then collected in systematic manner by proper planning which involves finding of answers to questions such as: Whether to use sample or census investigation, how to collect, arrange, present and analyse data, etc. This will be discussed in Unit 2 in greater detail.

- g) *Statistics should be placed in relation to one another*: Only comparable data make some sense. Unrelated and incomparable data are no data. They are just figures. For example, heights and weights of students of a class do not have any relation with the income and qualification of their parents. For comparability, the data should be homogeneous; that is, it should belong to the same subject or class or phenomenon. For example, pocket money of the students of a class is certainly related to the income of their parents. Prices of onions and potatoes in Delhi can certainly be related to their prices in other cities of India.

Thus, it will not be wrong to say that “*all statistics are numerical statements of facts but all numerical statements of facts are not statistics*”.

1.2.2 Statistics in Singular Sense

In the singular sense, Statistics refers to what is called *statistical methods* which means the ever-growing body of techniques for collection, condensation, presentation, analysis and interpretation of statistical data/quantitative information. In simple language, it means the subject of Statistics like any other subject such as Mathematics or Economics.

We can now take up definitions given by some famous statisticians.

A. L. Bowley gave a few definitions but none of them was complete and satisfactory. However, his two definitions make some sense even though incomplete. For example, he says, “*Statistics may be called the science of counting*”. Here he is emphasizing on enumeration aspect of statistics, which no doubt is important. At another place he describes statistics as “*the science of measurement of the social organism...*”. He is also of the view that “*Statistics may rightly be called the science of average*”. Although measurement, enumeration and averages (Arithmetic, Geometric and Harmonic means; Mode and Median which we will discuss in the next Block) are important, yet they are not the only concern of Statistics, as we shall study in the subsequent units.

Croxtan and Cowden have put forward a very simple and precise definition of Statistics as “*Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data*”.

This definition lays emphasis on five important aspects, which in fact, constitute the very scope of the subject called **Statistics** or **Statistical Methods**. These are:

- A) *Collection of data*: In any statistical inquiry, the collection of data is the first basic step. They form the foundation of statistical analysis, and therefore utmost care should be taken in collecting data. Faulty data will certainly lead to misleading results and can do more harm than good. The data can be drawn from two sources:

- a) *Primary source* where data are generated by the investigator himself through various methods discussed in detail in the next Unit, Section 2.4.

- b) *Secondary source* where data are extracted from the existing published or unpublished source, that is, from the data already collected by others. It saves a lot of time, effort and money of the investigator; but then he has to be conscious and judicious in their use. A detailed discussion is available in our Unit 2, Section 2.5.
- B) *Arrangement of Data*: Data from the secondary source are already arranged or organised like population data from Census of India. A minor rearrangement to suit our needs can be undertaken. However, primary data are in a haphazard form and need some arrangement so that it makes some sense. The steps involved in this process are: –
- a) *Editing*: This involves the removal of omissions and inconsistencies involved in the collected information.
 - b) *Classification of data*: It follows editing. It involves arranging data according to some common characteristic/s. Normally the raw information received from the respondents is put on the master sheets. For example, we may conduct a survey on, say, metal based engineering industries of Orissa, from where information are collected on capital structure, output of different types of products, employment of unskilled, semi-skilled and skilled workers, cost and price structure, technology aspects, etc. All this information can be put on master sheets. For more details refer to Unit 3.
- C) *Tabulation*: It is the last step in the arrangement process. From the master sheets (or coded sheets) information is tabulated in the form of frequency distributions or tables, where information is arranged in columns and rows. For more details refer to Unit 3.
- D) *Presentation of Data*: After the data have been arranged and tabulated, they can now be presented in the form of diagrams and graphs to facilitate the understanding of various trends as well as the process of comparison of various situations. Two different types of presentation of data are normally used, detailed study of which will be made in Unit 3.

These are:

- a) Statistical tables
 - b) Graphs including line graphs.
- E) *Analysis of Data*: It is the most important step in any statistical inquiry. A major portion of this course in Statistics is devoted to the methods used for analysing the collected data to derive some policy conclusions. The tools of analysis will be discussed in details in later units. For the time being we can summarize them as follows:

TOOLS OF STATISTICAL ANALYSIS

I) Theoretical Statistics

- a) *Uni-variate analysis*.
 - i) Measures of Central tendency: This includes mathematical averages such as arithmetic mean (\bar{X}) geometric mean (G) and harmonic mean (H) and positional averages such as mode (M_o) and median (M_d), and other partition values which include quartiles (Q), octiles (O), deciles (D) and percentiles (P).

- ii) Measures of dispersion: These include crude measures such as range (R), quartile deviation (QD), Mean Deviation (d), standard deviation (s), etc.
- iii) Measures of Skewness (S_k) — Karl Pearsons, Robert Bowley's, and Moment based (b_1 coefficient) measures.
- iv) Measures of Kurtosis (b_2 coefficient) based on moments.
- v) Probability and probability distributions such as Binomial, Poisson and Normal.

b) *Bi-variate analysis*

It includes analysis using two variables like amount of fertilizers (x) and the amount of yield (y) where it is known that yield (y) is affected by the amount of fertilizers (x) used. In this context we will discuss linear correlation (r_{xy}) and regression analysis in Units 7 and 8.

II) Applied Statistics

Here we use the tools developed in I to analyse some very useful aspects of our daily life. These include:

- a) Time series
- b) Index numbers
- c) Vital Statistics
- d) Inferential statistics, e.g., testing of hypothesis, etc.
- F) *Interpretation of Data*: It is the last but very crucial stage of a statistical inquiry or investigation. It is a job in itself which requires high degree of aptitude, skill and experience. In case of faulty interpretation, the very purpose of the investigation is lost. Our policies and actions later on depend very much on how soundly and correctly we interpreted our data. On this basis Wallis and Robert (Statistics - A New Approach) have rightly remarked that statistics may be regarded as "*a body of methods for making wise decisions in the face of uncertainty*".

1.2.3 Meaning of the Word 'Statistic'

You must have been buying a few kilograms of wheat every month for your family consumption. How do you judge the quality of wheat contained in a bag of 100 kgs? Theoretically two methods are open to you:

- a) *Census method* where each and every grain of wheat is examined. You will study in Unit 2 how this method is costly, time consuming, boring and at the same time unnecessary because almost same results can be obtained from a sample inquiry.
- b) *Sample method* where one or more samples, each containing few grains, are selected and examined. If you are satisfied with the sample/s, you buy the grains, assuming that all grains in the bag are of similar quality.

The statistical values of the characteristics of a population (such as mean height of students in a university for) are known as *parameters*. On the other hand, the mean, standard deviation, etc. of the sample taken from the population, are known as *statistic* and are the estimators of the parameter values.

Check Your Progress 1

- 1) Are the following statements correct? Give reasons in two or three lines:-
 - a) Statistics has no use for a modern man.
 - b) Statistics and statistic imply the same thing.

- c) Statistics in singular sense implies statistical methods.
- d) Statistics may rightly be called the science of averages.
- e) Statistics need not be numerically expressed.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2) Give five examples of the use of Statistics in daily life.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3) Use the words Statistics, statistics and statistic in three separate sentences to bring out the difference in them.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

4) How would you judge the quality of potatoes from a bag of 100 kg. from which you want to buy 5 kg. Use the words population, sample, statistic and parameters in your explanation. (Restrict your answer to six lines).

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.3 IMPORTANCE OF STATISTICS

We have seen in Section 1.1, that Statistics has a very wide application in our daily life. It is required in every field of inquiry; its knowledge has become necessary to study and understand our day to day problems. A Statistician H. G. Wells had rightly pointed out that *“statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write”*. Further, Statistics has acquired universal application. Even A. L. Bowley remarked that, *“Statistics cannot be confined to any one science”*. Thus, statistics is/are applied and interpreted in different ways in different fields of knowledge.

1.3.1 Statistics and Economics

The relationship between Statistics and Economics is very old. In 1690 Sir William Petty wrote a book titled *“Political Arithmetic”* using Statistics in Economics. In the late 19th century Alfred Marshall had observed that *“Statistics are the straw out of which I, like every other economist, have to make bricks”*.

In the 20th century, economists largely based their theories on statistical inquiry — on empirical evidence of human behavior rather than on deductive methods of analysis. J.M. Keynes, V. Pareto and others used statistics very extensively. Recently Statistics and Economics have been so intermixed that a new branch known as *Econometrics* has developed. Mean, Standard deviation, Regression analysis, Normal distribution, Sampling theory, etc. are being used extensively in economic analysis. In addition to this, following are the other important uses of statistics. The list is only illustrative and not exhaustive.

- 1) Estimation and analysis of national income.
- 2) Input-output analysis.
- 3) Empirical analysis of production function.
- 4) Financial statistics as contained in Reserve Bank of India Bulletins.
- 5) Statistical studies of population or demographic features like death rate, birth rate, life expectancy, etc.
- 6) Statistical studies of market structures like oligopolies and monopolies, etc.
- 7) Macro-economic variables like price level, employment, money supply, etc.
- 8) It would be impossible to understand and steer the growth process in underdeveloped economies without the availability of sufficient and reliable statistical information. Economic planning is just not possible without the availability of sufficient and reliable data.

1.3.2 Statistics and Business

Statistics helps in business too. For a progressive business concern, analysis of costs, revenue, profits, labour and capital, marketing, etc. are essential. Business planning involves business forecasting based on market surveys on demand, availability of substitute brands, opinions of consumers regarding different brands, consumers preferences, etc. Using time series analysis, one may isolate the effects of secular trend, seasonal variations, cyclical factors and irregular factors, on a business activity (see Unit 10).

Statistical methods are useful to business in formulating its business policies and activities in the field of production, finance, personnel, accounting and quality control. Modern business firms make extensive use of graphs, charts, and diagrams in their sale promotion efforts and display of their production achievements.

1.3.3 Statistics and Physical Sciences

Statistics has proved to be useful in physical sciences like Physics, Geology, Astronomy, Biology, Medicine, etc. A modern doctor relies heavily on the information on various parameters of a patient in diagnosing his disease. These include his body temperature behaviour, blood pressure and blood sugar level, ECG, etc. Doctor needs this information all the more when performing surgery.

Further, before introducing a new drug, data are collected and analysed for its effects on rats, monkeys, rabbits, etc. If found statistically satisfactory, the experiments are then conducted on human beings. The efficacy of the medicine is studied statistically. For example, researchers may be interested in finding whether quinine is still effective in the control of malaria with a new strain of mosquito. They may conduct the experiment on, say, 1000 patients selected at random. If the percentage of success is quite high, researchers may declare that quinine is still effective in the control of malaria.

Similarly, statistical studies are conducted in other physical sciences. Perhaps, it will not be an exaggeration to say that there is hardly any scientific study where use of statistical methods is not undertaken. The Gaussian "*Normal Law of Errors*" was used to study the movements of stars and planets. Thus, as Bowley pointed out, statistics can "*prove useful at any time under any circumstance*".

1.3.4 Statistics and Mathematics

The relation between Statistics and Mathematics is known to exist since the 17th century. The theory of probability has bearing on various statistical methods. In the last 100 years or so Statistics and Mathematics have come very close to each other to evolve a new subject called *Mathematical Statistics*.

1.3.5 Statistics and other Social Sciences

Similarly, scholars are increasingly using Statistics in Education, Political Science, Geography, Psychology, Anthropology, etc. All public opinion polls are based on Statistics. Other fields where Statistics is useful are all types of insurance, war/defense preparedness, index numbers and dearness allowances formulae, etc.

1.4 MISUSES OF STATISTICS

Although Statistics is indispensable in almost all fields of learning as pointed out above, yet it is likely to be misused and misinterpreted by the vested interests. These interests, like a ruling party, can always manipulate figures to arrive at the predetermined favourable results. Because of the various misuses, Statistics is sometimes called an *unscrupulous science*. Various facts can be twisted, distorted and presented with an evil design. This becomes easy when the state or the other vested interests have the monopoly of collecting and presenting statistics.

All this, no doubt therefore, has produced various misgivings about Statistics such as:

- a) "Statistics can prove anything"
- b) "Statistics are the lies of the first order".
- c) "There are three kinds of lies, namely, lies, damned lies and statistics"
- d) "Statistics is the rainbow of lies."

Statistical conclusions may be misinterpreted and hence can be disastrous. A story goes that a mathematician finding average height of his family members higher than average depth of a stream, decided to cross it safely. But on the other bank of the stream he found that except himself, all other members were drowned because his abnormal height had pulled up the average.

1.5 LIMITATIONS OF STATISTICS

As mentioned earlier in Section 1.3 of this Unit, and according to H.G. Wells, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write". However, statistics are not like Alladin's lamp which can perform all tricks. The following list of limitations is worth mentioning:

Firstly, statistical analysis depends upon the type of variable under consideration. For qualitative data such as beauty, health, goodwill and honesty, attempts have been made indirectly in the form of Analysis of Association of Attributes. Here we do not measure things like honesty but count their number.

Secondly, statistics deal only with aggregates. That is no significance is attached to individual items which make this aggregate. For example, one state of India may be richer than other states, but some people may be much poorer in the rich state than some people of the poorer states. Averages sometimes may be misleading.

Thirdly, statistical conclusions are not mathematically exact. It is possible that with wrong samples, taken knowingly or unknowingly, the results may be favourable by fluke.

Fourthly, as mentioned in Section 1.4, statistical measures like averages may be misinterpreted and hence can prove disastrous.

Check Your Progress 2

- 1) Comment on the following statements in 3-4 sentences
 - a) Statistics are confined only to Economics and Business.
 - b) There are lies, damned lies and Statistics.

- c) If the per capita income of a country is Rs. 4050, it means that everybody is getting that income.
- d) Econometrics is an intermix of Economics and Mathematics.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2) Give five examples of the use of Statistics in Economics and Business.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

3) Mention four fields where Statistics is being used prominently.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

4) Explain the following terms:

- a) Mathematical Statistics
- b) Statistics as an unscrupulous science
- c) Statistics as Alladin's lamp

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.6 LET US SUM UP

A modern man must possess knowledge of Statistics like that of reading and writing. The word Statistics in singular sense implies statistical methods aimed at collecting, arranging, presenting, analysing and interpreting data. In the plural sense, it mean mass of quantitative information like population data.

The word statistic (as against statistics) means an estimator obtained from a sample with a purpose to infer about the population value called parameter.

Statistics has utility in almost all branches of knowledge. In Economics and Business it has a special utility. Combination of Economics, Statistics and Mathematics has led to a new subject called *Econometrics*.

In spite of immense utility, some unscrupulous persons have misused statistics driving it to the level that is worse than damned lies. Because of this, sometimes, it has been termed as unscrupulous science.

1.7 KEY WORDS

Statistics: In plural sense, it means a set of numerical figures commonly known as statistical data.

Statistics: In singular sense, it means scientific methods for collection, presentation, analysis and interpretation of data.

Statistic: It is measure, like arithmetic mean, median, geometric mean, standard deviation, etc., calculated from sample. It is also termed as estimator in the theory of estimation.

Parameter: It is a measure like arithmetic mean, median, geometric mean, standard deviation, etc., calculated by using all values of population.

Quantitative data: These are information on measurable characteristics. Such data are available in the form of numerical figures.

Qualitative data: These are information on a non-measurable characteristics like honesty, beauty, color, caste, etc.

Population: The totality of all the units falling under the scope of an investigation.

Sample: A sample is a fraction of the population used to study its one or more characteristics.

Census: A method of investigation in which information is collected from all units of the population.

Sampling: A method of investigation in which information is collected from sampled units only.

1.8 SOME USEFUL BOOKS

Elhance, D.N. and V. Elhance, 1988, *Fundamental of Statistics*, Kitab Mahal, Allahabad.

Nagar, A.L. and R.K. Dass, 1983, *Basic Statistics*, Oxford University Press, Delhi

Mansfield, E., 1991, *Statistics for Business and Economics: Method and Applications*, W.W. Norton and Co.

Yule, G.U. and M.G. Kendall, 1991, *An Introduction to the Theory of Statistics*, Univeristy Books, Delhi.

1.9 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

- 1) (a) false, (b) false, (c) true, (d) true, (e) false.
- 2) Refer Section 1.1
- 3) Refer Sub-Sections 1.2.1 to 1.2.3
- 4) Refer Sub-Section 1.2.3

Check Your Progress 2

- 1) a) Refer Section 1.3
b) and (c) refer Section 1.4
d) Refer Sub-Section 1.3.1
- 2) Refer Sub-Sections 1.3.1 and 1.3.2
- 3) Refer Section 1.3
- 4) a) Refer Sub-Section 1.3.4
b) Refer Section 1.4
c) Refer Section 1.5