Introduction:

In the modern world of computers and information technology, the importance of statistics is very well recogonised by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

Origin and Growth of Statistics:

The word ‘ Statistics’ and ‘ Statistical’ are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of scientific method is of comparatively recent growth. Research particularly into the mathematical theory of statistics is rapidly proceeding and fresh discoveries are being made all over the world.

Meaning of Statistics:

Statistics is concerned with scientific methods for collecting, organising, summarising, presenting and analysing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis. Statistics is concerned with the systematic collection of numerical data and its interpretation.

The word ‘ statistic’ is used to refer to

1. Numerical facts, such as the number of people living in particular area.
2. The study of ways of collecting, analysing and interpreting the facts.

Definitions:

Statistics is defined differently by different authors over a period of time. In the olden days statistics was confined to only state affairs but in modern days it embraces almost every sphere of 2 human activity. Therefore a number of old definitions, which was confined to narrow field of enquiry were replaced by more definitions, which are much more comprehensive and exhaustive. Secondly, statistics has been defined in two different ways – Statistical data and statistical methods.

The following are some of the definitions of statistics as numerical data.

1. Statistics are the classified facts representing the conditions of people in a state. In particular they are the facts, which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement.
2. Statistics are measurements, enumerations or estimates of natural phenomenon usually systematically arranged, analysed and presented as to exhibit important interrelationships among them.

Definitions by A.L. Bowley:

Statistics are numerical statement of facts in any department of enquiry placed in relation to each other. - **A.L. Bowley**

Statistics may be called the science of counting in one of the departments due to Bowley, obviously this is an incomplete definition as it takes into account only the aspect of collection and ignores other aspects such as analysis, presentation and interpretation. Bowley gives another definition for statistics, which states ‘ statistics may be rightly called the scheme of averages’ . This definition is also incomplete, as averages play an important role in understanding and comparing data and statistics provide more measures.

Definition by Croxton and Cowden:

Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis. It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one.

According to this definition there are four stages:

1. Collection of Data: It is the first step and this is the foundation upon which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of correct method.

2. Presentation of data: The mass data collected should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

3. Analysis of data: The data presented should be carefully analysed for making inference from the presented data such as measures of central tendencies, dispersion, correlation, regression etc.,

4. Interpretation of data: The final step is drawing conclusion from the data collected. A valid conclusion must be drawn on the basis of analysis. A high degree of skill and experience is necessary for the interpretation.

Scope of Statistics:

Statistics is not a mere device for collecting numerical data, but as a means of developing sound techniques for their handling, analysing and drawing valid inferences from them. Statistics is applied in every sphere of human activity – social as well as physical – like Biology, Commerce, Education, Planning, Business Management, Information Technology, etc. It is almost impossible to find a single department of human activity where statistics cannot be applied. We now discuss briefly the applications of statistics in other disciplines.

1. **Statistics and Industry:**

Statistics is widely used in many industries. In industries, control charts are widely used to maintain a certain quality level. In production engineering, to find whether the product is conforming to specifications or not, statistical tools, namely inspection plans, control charts, etc., are of extreme importance. In inspection plans we have to resort to some kind of sampling – a very important aspect of Statistics.

2. **Statistics and Commerce:**

Statistics are lifeblood of successful commerce. Any businessman cannot afford to either by under stocking or having overstock of his goods. In the beginning he estimates the demand for his goods and then takes steps to adjust with his output or purchases. Thus statistics is indispensable in business and commerce.

As so many multinational companies have invaded into our Indian economy, the size and volume of business is increasing. On one side the stiff competition is increasing whereas on the other side the tastes are changing and new fashions are emerging. In this connection, market survey plays an important role to exhibit the present conditions and to forecast the likely changes in future.

3. **Statistics and Agriculture:**

Analysis of variance (ANOVA) is one of the statistical tools developed by Professor R.A. Fisher, plays a prominent role in agriculture experiments. In tests of significance based on small samples, it can be shown that statistics is adequate to test the significant difference between two

sample means. In analysis of variance, we are concerned with the testing of equality of several population means.

For an example, five fertilizers are applied to five plots each of wheat and the yield of wheat on each of the plots are given. In such a situation, we are interested in finding out whether the effect of these fertilisers on the yield is significantly different or not. In other words, whether the samples are drawn from the same normal population or not. The answer to this problem is provided by the technique of ANOVA and it is used to test the homogeneity of several population means.

4. Statistics and Economics:

Statistical methods are useful in measuring numerical changes in complex groups and interpreting collective phenomenon. Nowadays the uses of statistics are abundantly made in any economic study. Both in economic theory and practice, statistical methods play an important role. Alfred Marshall said, " Statistics are the straw only which I like every other economist have to make the bricks". It may also be noted that statistical data and techniques of statistical tools are immensely useful in solving many economic problems such as wages, prices, production, distribution of income and wealth and so on. Statistical tools like Index numbers, time series Analysis, Estimation theory, Testing Statistical Hypothesis are extensively used in economics.

5. Statistics and Education:

Statistics is widely used in education. Research has become a common feature in all branches of activities. Statistics is necessary for the formulation of policies to start new course, consideration of facilities available for new courses etc. There are many people engaged in research work to test the past knowledge and evolve new knowledge. These are possible only through statistics.

6. Statistics and Planning:

Statistics is indispensable in planning. In the modern world, which can be termed as the "world of planning", almost all the organisations in the government are seeking the help of planning for efficient working, for the formulation of policy decisions and execution of the same. In order to achieve the above goals, the statistical data relating to production, consumption, demand, supply, prices, investments, income expenditure etc and various advanced statistical techniques for processing, analysing and interpreting such complex data are of importance. In India statistics play an important role in planning, commissioning both at the central and state government levels.

7. Statistics and Medicine:

In Medical sciences, statistical tools are widely used. In order to test the efficiency of a new drug or medicine, t - test is used or to compare the efficiency of two drugs or two medicines, ttest for the two samples is used. More and more applications of statistics are at present used in clinical investigation.

8. Statistics and Modern applications:

Recent developments in the fields of computer technology and information technology have enabled statistics to integrate their models and thus make statistics a part of decision making procedures of many organisations. There are so many software packages available for solving design of experiments, forecasting simulation problems etc. SYSTAT, a software package offers mere scientific and technical graphing options than any other desktop statistics package. SYSTAT supports all types of scientific and technical research in various diversified fields as follows

1. Archeology: Evolution of skull dimensions
2. Epidemiology: Tuberculosis

3. Statistics: Theoretical distributions
4. Manufacturing: Quality improvement
5. Medical research: Clinical investigations.
6. Geology: Estimation of Uranium reserves from ground water.

Limitations of statistics:

Statistics with all its wide application in every sphere of human activity has its own limitations. Some of them are given below.

1. Statistics is not suitable to the study of qualitative phenomenon: Since statistics is basically a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements. As a matter of fact, qualitative phenomenon like honesty, poverty, beauty, intelligence etc, cannot be expressed numerically and any statistical analysis cannot be directly applied on these qualitative phenomenons. Nevertheless, statistical techniques may be applied indirectly by first reducing the qualitative expressions to accurate quantitative terms. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.

2. Statistics does not study individuals: Statistics does not give any specific importance to the individual items, in fact it deals with an aggregate of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.

3. Statistical laws are not exact: It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.

4. Statistics table may be misused: Statistics must be used only by experts; otherwise, statistical methods are the most dangerous tools on the hands of the inexpert. The use of statistical tools by the inexperienced and untraced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. As King says aptly ‘ statistics are like clay of which one can make a God or Devil as one pleases’ .

5. Statistics is only, one of the methods of studying a problem: Statistical method do not provide complete solution of the problems because problems are to be studied taking the background of the countries culture, philosophy or religion into consideration. Thus the statistical study should be supplemented by other evidences.

Measurement Scales

Nominal scale. The values of a *nominal variable* cannot be ordered. Examples are the gender of a person (male female) or the status of an application (pending–not pending).

Ordinal scale. The values of an *ordinal variable* can be ordered. However, the differences between these values cannot be interpreted in a meaningful way. For example, the possible values of education level (none–primary education–secondary education–university degree) can be ordered meaningfully, but the differences between these values cannot be interpreted. Likewise, the satisfaction with a product (unsatisfied–satisfied–very satisfied) is an ordinal variable because the values this variable can take can be ordered, but the differences between "unsatisfied– satisfied" and "satisfied–very satisfied" cannot be compared in a numerical way.

Continuous scale. The values of a *continuous variable* can be ordered. Furthermore, the differences between these values can be interpreted in a meaningful way. For instance, the height of a person refers to a continuous variable because the values can be ordered (170 cm, 171 cm, 172 cm, …), and differences between these values can be compared (the difference between 170 and 171cm is the same as the difference between 171 and 172 cm).
Sometimes, the continuous scale is divided further into subscales :

***Interval scale*.** Only differences between values, but not ratios, can be interpreted. An example for this scale would be temperature (measured in ◦C): the difference between −2 ◦C and 4 ◦C is 6◦C, but the ratio of 4/−2 = −2 does not mean that −4 ◦C is twice as cold as 2 ◦C.

***Ratio scale*.** Both differences and ratios can be interpreted. An example is speed: 60 km/h is 40 km/h more than 20 km/h. Moreover, 60 km/h is three times faster than 20 km/h because the ratio between them is 3.

***Absolute scale*.** The absolute scale is the same as the ratio scale, with the exception that the values are measured in "natural" units. An example is "number of semesters studied" where no artificial unit such as km/h or ◦C is needed: the values are simply 1, 2, 3, ...

Categories of data:
Any statistical data can be classified under two categories depending upon the sources utilized.
These categories are,
1. Primary data 2. Secondary data
Primary data:
Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.
For example: If a researcher is interested to know the impact of noonmeal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.
The primary data can be collected by the following five methods.
1. Direct personal interviews.

2. Indirect Oral interviews.
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.

1. Direct personal interviews:

The persons from whom informations are collected are known as informants. The investigator personally meets them and asks questions to gather the necessary informations. It is the suitable method for intensive rather than extensive field surveys. It suits best for intensive study of the limited field.

Merits:

1. People willingly supply informations because they are approached personally. Hence, more response noticed in this method than in any other method.
2. The collected informations are likely to be uniform and accurate. The investigator is there to clear the doubts of the informants.
3. Supplementary informations on informant's personal aspects can be noted. Informations on character and environment may help later to interpret some of the results.
4. Answers for questions about which the informant is likely to be sensitive can be gathered by this method.
5. The wordings in one or more questions can be altered to suit any informant. Explanations may be given in other languages also. Inconvenience and misinterpretations are thereby avoided.

Limitations:

1. It is very costly and time consuming.
2. It is very difficult, when the number of persons to be interviewed is large and the persons are spread over a wide area.
3. Personal prejudice and bias are greater under this method.

2. Indirect Oral Interviews:

Under this method the investigator contacts witnesses or neighbours or friends or some other third parties who are capable of supplying the necessary information. This method is preferred if the required information is on addiction or cause of fire or theft or murder etc., If a fire has broken out a certain place, the persons living in neighbor-hood and witnesses are likely to give information on the cause of fire. In some cases, police interrogated third parties who are supposed to have knowledge of a theft or a murder and get some clues. Enquiry committees appointed by governments generally adopt this method and get people' s views and all possible details of facts relating to the enquiry. This method is suitable whenever direct sources do not exists or cannot be relied upon or would be unwilling to part with the information. The validity of the results depends upon a few factors, such as the nature of the person whose evidence is being recorded, the ability of the interviewer to draw out information from the third parties by means of appropriate questions and cross examinations, and the number of persons interviewed. For the success of this method one person or one group alone should not be relied upon.

3. Information from correspondents:

The investigator appoints local agents or correspondents in different places and compiles the information sent by them. Informations to Newspapers and some departments of Government come by this method. The advantage of this method is that it is cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are

likely to be negligent, prejudiced and biased. This method is adopted in those cases where informations are to be collected periodically from a wide area for a long time.

Mailed questionnaire method:

Under this method a list of questions is prepared and is sent to all the informants by post. The list of questions is technically called questionnaire. A covering letter accompanying the questionnaire explains the purpose of the investigation and the importance of correct informations and request the informants to fill in the blank spaces provided and to return the form within a specified time. This method is appropriate in those cases where the informants are literates and are spread over a wide area.

Merits:

1. It is relatively cheap.
2. It is preferable when the informants are spread over the wide area.

Limitations:

1. The greatest limitation is that the informants should be literates who are able to understand and reply the questions.
2. It is possible that some of the persons who receive the questionnaires do not return them.
3. It is difficult to verify the correctness of the informations furnished by the respondents.

Schedules sent through Enumerators:

Under this method enumerators or interviewers take the schedules, meet the informants and filling their replies. Often distinction is made between the schedule and a questionnaire. A schedule is filled by the interviewers in a face-to-face situation with the informant. A questionnaire is filled by the informant which he receives and returns by post. It is suitable for extensive surveys.

Merits:

1. It can be adopted even if the informants are illiterates.
2. Answers for questions of personal and pecuniary nature can be collected.
3. Non-response is minimum as enumerators go personally and contact the informants.
4. The informations collected are reliable. The enumerators can be properly trained for the same.
5. It is most popular methods.

Limitations:

1. It is the costliest method.
2. Extensive training is to be given to the enumerators for collecting correct and uniform informations.
3. Interviewing requires experience. Unskilled investigators are likely to fail in their work.

Before the actual survey, a pilot survey is conducted. The questionnaire/Schedule is pre-tested in a pilot survey. A few among the people from whom actual information is needed are asked to reply. If they misunderstand a question or find it difficult to answer or do not like its wordings etc., it is to be altered. Further it is to be ensured that every question fetches the desired answer.

Secondary Data:

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency. According to W.A.Neiswanger, ‘ A primary source is a publication in which the data are published by the same authority which gathered and analysed them. A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible’ .

Sources of Secondary data:

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. The sources of secondary data can broadly be classified under two heads:

1. Published sources, and
2. Unpublished sources.

1. Published Sources:

The various sources of published data are:

1. Reports and official publications of

(i) International bodies such as the International Monetary Fund, International Finance Corporation and United Nations Organisation.

(ii) Central and State Governments such as the Report of the Tandon Committee and Pay Commission.

2. Semi-official publication of various local bodies such as Municipal Corporations and District Boards.

3. Private publications-such as the publications of –

(i) Trade and professional bodies such as the Federation of Indian Chambers of Commerce and Institute of Chartered Accountants.

(ii) Financial and economic journals such as ‘ Commerce’ , ‘ Capital’ and ‘ Indian Finance’ .

(iii) Annual reports of joint stock companies.

(iv) Publications brought out by research agencies, research scholars, etc.

It should be noted that the publications mentioned above vary with regard to the periodically of publication. Some are published at regular intervals (yearly, monthly, weekly etc.,) whereas others are ad hoc publications, i.e., with no regularity about periodicity of publications.

Note: A lot of secondary data is available in the internet. We can access it at any time for the further studies.

2. Unpublished Sources

All statistical material is not always published. There are various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc. Such sources can also be used where necessary.

Classification:
The collected data, also known as raw data or ungrouped data are always in an un organised form and need to be organized and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible and assimilable form. The process of grouping into different classes or sub classes according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data.
Thus classification is the first step in tabulation. For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore,
Mumbai etc.,

Tabulation:

Tabulation is the process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information.
A table is a systematic arrangement of classified data in columns and rows. Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious.Classification and ‘ Tabulation’ , as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation data are classified and then displayed under different columns and rows of a table.

Advantages of Tabulation:
Statistical data arranged in a tabular form serve following objectives:
1. It simplifies complex data and the data presented are easily understood.
2. It facilitates comparison of related facts.
3. It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
4. It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided. Moreover, the needed information can be easily located.
5. Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

Preparing a Table:
The making of a compact table itself an art. This should contain all the information needed within the smallest possible space. What the purpose of tabulation is and how the tabulated information is to be used are the main points to be kept in mind while preparing for a statistical table. An ideal table should consist of the following main parts:
1. Table number
2. Title of the table
3. Captions or column headings
4. Stubs or row designation
5. Body of the table
6. Footnotes
7. Sources of data

Table Number:

A table should be numbered for easy reference and identification. This number, if possible, should be written in the centre at the top of the table. Sometimes it is also written just before the title of the table.

Title:

A good table should have a clearly worded, brief but unambiguous title explaining the nature of data contained in the table. It should also state arrangement of data and the period covered. The title should be placed centrally on the top of a table just below the table number (or just after table number in the same line).

Captions or column Headings:

Captions in a table stands for brief and self explanatory headings of vertical columns. Captions may involve headings and sub-headings as well. The unit of data contained should also be given for each column. Usually, a relatively less important and shorter classification should be tabulated in the columns.

Stubs or Row Designations:

Stubs stands for brief and self explanatory headings of horizontal rows. Normally, a relatively more important classification is given in rows. Also a variable with a large number of classes is usually represented in rows. For example, rows maystand for score of classes and columns for data related to sex of students. In the process, there will be many rows for scores classes but only two columns for male and female students.

A model structure of a table is given below:

Table Number **Title of the Table**

Sub Heading	Caption Headings Caption Sub-Headings	**Total**
Stub Sub-Headings	**Body**	
Total		

Foot notes:

Sources Note:

Body:

The body of the table contains the numerical information of frequency of observations in the different cells. This arrangement of data is according to the discription of captions and stubs.

Footnotes:

Footnotes are given at the foot of the table for explanation of any fact or information included in the table which needs some explanation. Thus, they are meant for explaining or providing further details about the data, that have not been covered in title, captions and stubs.

Sources of data:

Lastly one should also mention the source of information from which data are taken. This may preferably include the name of the author, volume, page and the year of publication. This should also state whether the data contained in the table is of ‘ primary or secondary’ nature.

Requirements of a Good Table:

A good statistical table is not merely a careless grouping of columns and rows but should be such that it summarizes the total information in an easily accessible form in minimum possible space. Thus while preparing a table, one must have a clear idea of the information to be presented, the facts to be compared and he points to be stressed. Though, there is no hard and fast rule for forming a table yet a few general point should be kept in mind:

1. A table should be formed in keeping with the objects of statistical enquiry.
2. A table should be carefully prepared so that it is easily understandable.
3. A table should be formed so as to suit the size of the paper. But such an adjustment should not be at the cost of legibility.
4. If the figures in the table are large, they should be suitably rounded or approximated. The method of approximation and units of measurements too should be specified.
5. Rows and columns in a table should be numbered and certain figures to be stressed may be put in ‘ box’ or ‘ circle’or in bold letters.
6. The arrangements of rows and columns should be in a logical and systematic order. This arrangement may be alphabetical, chronological or according to size.
7. The rows and columns are separated by single, double or thick lines to represent various classes and sub-classes used. The corresponding proportions or percentages should be given in adjoining rows and columns to enable comparison. A vertical expansion of the table is generally more convenient than the horizontal one.
8. The averages or totals of different rows should be given at the right of the table and that of columns at the bottom of the table. Totals for every sub-class too should be mentioned.
9. In case it is not possible to accommodate all the information in a single table, it is better to have two or more related tables.