UNIT – I

ORIGIN OF STATISTICS:

The word Statistics came into existence in the **18th** century. The English word Statistics had its origin from the **Latin** word **status** meaning **"political state"**. Words **"Stato"** and **Statistik"** were also used in similar sense in the Italian and German languages.

In the ancient days, the word statistics was used only for facts and figures required by the state. For eg: collecting information relating to crimes, military strength, census, birth rate etc. and they use to call statistics as **" science of king"** or "**science of state craft"** or **"political arithmetic"**.

But now statistics has been used everywhere in fields like in business decision-making, schools, colleges etc. In short, statistics means aggregate of facts. The word statistics is basically used in two senses:

➢ Statistics data

➢ Statistical method

MEANING OF STATISTICS:

The set of quantitative data relating to phenomena, whether concerning state or otherwise are called statistics. The word statistics is used in different senses by different persons.

To a layman, statistics has become **"synonymous with figures**".

DEFINITION OF STATISTICS:

(i) Statistics can be defined as "the science which deals with collection, presentation, analysis and interpretation of numerical data".

- **Croxton and Cowden**

(ii) "Statistics are the collection of noteworthy facts concerning state both historical and descriptive".

- **Gotfried and Achenwell**

(iii) "Statistics is an aggregate knowledge brought together for a practical end namely the knowledge of concrete states".

- **Wappans**

CHARACTERISTICS OF STATISTICS:

(i) Aggregate of facts:

Statistics must relate to aggregate of facts. A single figure even though numerically expressed cannot be called statistics because such figures are unrelated and cannot be compared.

(ii) Affected by multiplicity of causes:

Generally speaking, statistical facts are not traceable to single factor. For eg: the prices of particular commodity are affected by multiple factors such as supply, demand, import, income distribution, taste etc,

(iii) Numerically expressed:

It is only the numerical data which constitute statistics. Any qualitative expression such as rich, poor, bad, intelligent etc do not constitute statistics, and this is outside the scope of study of statistics unless a numerical equivalent value is assigned to each expression.

(iv) Enumerated or Estimated:

Statistical data relating to any subject could be collected either by counting or by estimation. For eg: number of students appearing in an examination in a particular class will be counted; but to find out the number of persons attending a particular public meeting will be estimated.

(v) Reasonable degree of accuracy:

Statistical data are collected only with a reasonable degree of accuracy. The degree of accuracy of estimated value largely depends on the nature and the purpose of the enquiry.

(vi) Systematic manner:

The collection of data must be in a careful and systematic manner because data collected in haphazard and unsystematic may lead to erroneous, misleading fallacious conclusions.

(vii) Pre-determined purpose:

It is essential that one must be clear in advance about the objectives and purpose of enquiry or data to be collected before actually going for it.

(viii) Placed in relation to each other:

The comparability is one of the important attributes for the numerical facts to be known as statistics. The numerical data may be compared period-wise or region-wise.

IMPORTANCE AND FUNCTIONS OF STATISTICS:

(i) Preciseness and definiteness:

Statistics presents the facts in a precise and definite form. It is generally agreed that facts which are more precise and definite form.

(ii) Simplification in understandable form:

Statistics presents the large volume of complex data in meaningful and understandable form.

(iii) Facilitates comparison:

Statistics enable us to understand the behavior of data a time period or at a point and has certain reasons for it. Averages, standard deviations, coefficient of variation, time series, index number, correlation, graphs, and diagrams are some of the statistical techniques which make the comparison of particular phenomenon possible.

(iv) Enrich individual knowledge:

Statistics is a science which provides opportunities to individuals to enrich their know knowledge and experience. It has provided a master key to a mankind that we can use it anywhere and can study any problem and on right times.

(v)   Formulation of policies:

Statistics helps in the formulation of various economic, business and other policies at state, national or global level.

(vi)   Business forecasting:

Forecasting means formal process of predicting future events that will significantly affect the functioning of enterprise. Government and business units needs forecasting for designing better plans policies and sound decision making in order to use existing man and material resources optimally. To achieve this end, statistics offers various forecasting techniques such as time-series, extrapolation in order to make reliable predictions on future.

USES OF STATISTICS:

➢ Nowadays the problems of business firms are becoming complex because of the growing size and ever increasing competition. So they are using more and more statistics in "**Planning and decision- making**".

➢ Statistical data and statistical methods are the important tools of an economist in proper understanding of economic problems like poverty, unemployment etc. and in the "**formulation of economic policies**".

➢ Statistics help not only in the formulation of policies but also in evaluating their effect.

➢ Statistical methods help in forming suitable decision in production, sales and finance departments.

- Statistical tables and charts are used by the sales manager to present numerical facts of sales, production etc.

- Sampling methods are used in making survey of consumer preferences.

- Time series and business forecasting enable the businessman to estimate the future sales, profits etc.

LIMITATIONS:

- It does not deal with the qualitative characteristics like efficiency, honesty, intelligence, blindness and deafness.

- It is based on aggregate facts. And it does not deal with individual measurements. For example, the wage earned by an individual worker at any one time taken by it is not a statistical datum. But the wages of workers of a factory can be used statistically.

- Statistical results are true only on an average

- Statistics is only one of the methods of studying a problem. i.e. Statistical tools do not provide the best solution under all circumstances

- If statistical results are based on incomplete information, then the statistics can be misused.

  "W.I.King, "Statistics are like clay from which you can make a god or a devil as you please.

SOURCE OF DATA COLLECTION:

Collection of data refers to systematic recording of results either by counting or by enumeration. The entire structure of statistical analysis for any enquiry is based on systematic collection of data. Broadly speaking, there are two types of data;

- ➢ **Primary data**
- ➢ **Secondary data**

PRIMARY DATA:

It is the data, which is collected for the first time by investigators or enumerators working under his supervision to serve a particular purpose. Such a data is of original in nature. In India, the organizations such as CSO (Central Statistical Organization), Census of India, NSS (National Sample Survey), and RBI (Reserve Bank of India) collect and publish the primary data and so they are primary sources of data.

METHODS OF COLLECTING PRIMARY DATA:

Primary data may be obtained by applying any of the following methods:

- ➢ Direct personal interviews,
- ➢ Indirect oral interviews,
- ➢ Information from correspondents,
- ➢ Mailed questionnaire method
- ➢ Schedule sent through enumerators and
- ➢ Telephonic survey.

DIRECT PERSONAL INTERVIEW:

In this method, there is a face-to-face contact between interviewer and the informants. The interviewer asks them questions related to the survey and collect the required information. The information thus obtained is first-hand or original in character.

MERITS:

➢ Response is more encouraging as most people are willing to supply information when approached personally.

➢ The information obtained by this method is likely to be more accurate.

➢ Questions about which the informant is likely to be sensitive can be carefully sandwiched between other questions by the interviewer. He can twist the questions keeping in mind the informant's reactions.

➢ The language of communication can be adjusted to the status and educational level of the person interviewed, thus avoiding inconvenience and misinterpretation on the part of the informant.

DE-MERITS:

➢ It may be very costly where the number of persons to be interviewed is large and they are spread over a wide area.

➢ There are greater chances of personal prejudice and bias when comparing with other methods.

➢ Untrained or poorly trained interviewers may spoil the entire work.

➢ The time required to collect information by this method is more than other methods.

INDIRECT ORAL INTERVIEWS:

Under this method of collecting data, the investigator contacts the third parties called witness capable of supplying the necessary information. Enquiry committees and commissions appointed by the governments generally adopt this method to get people's views. For eg: clues about thefts or murders are obtained by the police by interrogating third parties who are supposed to have knowledge about the case under investigation. This method is very popular in practice. However, the correctness of information obtained depends upon a number of factors, such as:

➢ The type of persons whose evidence is being recorded. If the people do not know the full facts of the problem under investigation or if they are prejudiced it will not be possible to arrive at correct conclusions.

➢ The ability of the interviewers to draw out the information from witnesses by means of appropriate questions and cross examination.

INFORMATION FROM CORRESPONDENTS:

The investigator appoints local agents or correspondents in different places to collect information. The correspondents collect and transmit the information to the central office where the data are processed. For eg: Newspaper agencies generally adopt this method. The correspondents may be paid or honorary persons but generally they are paid. This method is adopted in various departments of government. Also correspondents in different places supply information relating to such events as accidents, strikes etc, to head office.

MAILED QUESTIONNAIRE METHOD:

Under this method, a well-designed questionnaire is mailed to the informants with a request to fill it up and return the same within the specific time schedule. It is pointed out in the

covering letter to the respondents that information supplied by them in the questionnaire will be kept strictly confidential. The investigators send questionnaire along with self addressed envelope to respondents for quick and better response. This type of method is very popular among research workers, private agencies, etc.

MERITS:

➢ It is most economical method in terms of time, money and manpower.

➢ It may be used more effectively in cases where scope of enquiry is very wide and extensive.

➢ Direct information from respondents, leaves less scope for personal bias.

DE-MRITS:

➢ This method does not work where majority of respondents are literate.

➢ There is low degree of reliability of the information supplied by informants.

➢ This method is not flexible.

SCHEDULE OR QUESTIONNAIRE SENT THROUGH ENUMERATORS:

Under this method of enquiry, the trained enumerators personally visit the informants and explain the objectives of enquiry ask questions and record their replies. This method is commonly used by big business houses, research institutions and large public undertakings.

MERITS:

➢ It enables to extract information from all type of respondents literate and illiterate.

➢ The information collected by this method is more accurate and reliable.

➢ This method is ideal for extensive surveys.

➢ In the case of poor response, the enumerators can personally persuade the informants to supply information.

DE-MERITS:

➢ It is the most expensive method because the enumerators who collect information are paid workers.

➢ It is time consuming

➢ It is hectic and tiring

➢ The success of this method depends on skill of enumerators, object of enquiry, etc.

TELEPHONIC SURVEY:

Under this method investigator instead of presenting himself before the informants contact them on telephone and collect the desired information.

MERITS:

➢ It is convenient and less time consuming.

DE-MERITS:

➢ It is subjective in nature.

SECONDARY DATA:

It refers to the data which is originally collected and published by the authorities other than who require it. Such data is already available from the government publications, research study, journals or newspapers. Main sources of secondary data may be classified in the following two categories:

➢ **Published sources**

➢ **Un- published sources**

PUBLISHED SOURCES:

Published sources of secondary data basically refers to various national / international organizations / agencies which collect data and publish the statistical data relating to business,

trade, labour, price , consumption, production, investment, savings, population, unemployment, banks and financial institutions, corporations etc, are some of the important sources of secondary data are:

- Publications of the central and state governments, of foreign governments and international bodies like IBRD, IMF, ILO, WTO, and WHO etc.

- Publications of semi-government organization eg: Reserve Bank of India Bulletin, reports on currency and finance.

- Publications of CSO and other research bodied (ICAR & IARI, Delhi).

- Publications of various chambers of commerce, trade associations and co-operative societies.

UNPUBLISHED SOURCES:

It covers all those sources of secondary data where records are maintained by private agencies or business firms for their own use and are restrictedly available for use of general public. Data collected by research institutions are also included in the category of unpublished sources of secondary data.
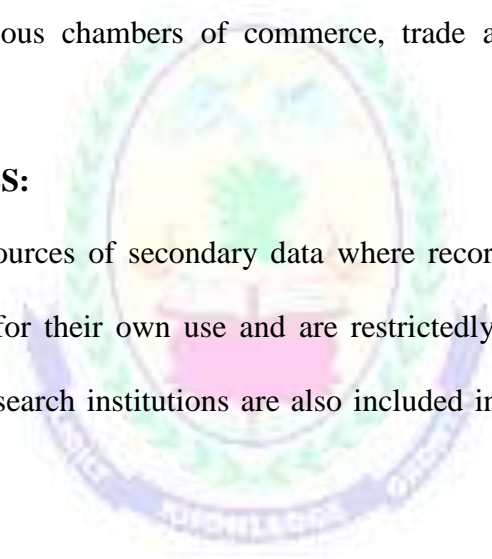
DIFFERENCE BETWEEN PRIMARY AND SECONDARY DATA:

PRIMARY DATA	SECONDARY DATA
Primary data is first-hand information and original in nature.	Secondary data is in the form of compilation of existing data or already published data.
The collection of primary data involves huge resources in terms of money and time, finance and energy.	Secondary data is relatively less costly.
Primary data is usually collected by keeping in mind the purpose for which it is collected so its suitability will be more.	Secondary data may or may not suit the purpose.
Primary data may be used as it is in its original form.	The use of secondary data requires lot of care and precaution.
Primary data are more reliable, accurate and adequate.	Secondary data are not always, reliable, accurate and adequate.

REQUISITES OF A GOOD QUESTIONNAIRE:

➢ A covering letter should be enclosed with every questionnaire.

➢ The questions should be simple and short to understand and the number of questions should be small.

➢ Questions should be arranged logically.

➢ Ambiguous questions should be avoided.

➢ As far as possible, personal questions about income, sales-tax paid, etc., should be avoided.

➢ The necessary instructions to the informants should be included.

➢ Objective type questions and 'yes or no' questions should be included.

➢ A questionnaire should be pre-tested and cross-checked before mailing it.