Unit

1 Introduction to Statistics

1.1 Introduction

Statistics is not a new discipline. The origin of statistics had been started with the origin of the development of human society. It is as old as the human society as human beings used statistics even unknowingly in preliminary phase of development of human civilization. In ancient time, statistics was regarded as the science of statecraft and it was used to collect the data of age and sex-wise population as well as property and wealth of state by the governments for framing military and fiscal policies. The historical evidences such as census of population during the construction of 'Pyramid' in Egypt by Pharaohs, the counting and recording of losses occurred during Napoleonic war in Britain, censuses held in England and Germany in the middle ages are regarded as the initiation of development of statistics. But nowadays it embraces almost each and every sphere of natural and human activity.

By the works of French gambler Chevalier de-Mere (Science of Probability), De-Moivre (Normal Probability), Gauss (Principle of Least Square and Normal Laws of Errors), Markov (Markov chains), Liapounoff (Central Limit Theorem) etc have made the outstanding contributions to modernize statistics. Similarly, Francis Galton and Karl Pearson pioneered the study of regression analysis and correlation analysis which has been widely used in various field of modern world. Sir Ronald A. Fisher applied statistics to diversified fields such as genetics, biometry, psychology and education, agriculture etc. So R. A. Fisher is regarded as the father of statistics.

1.2 Meaning and Definition of Statistics

The word statistics have been derived from the Latin word 'status' or Italian word 'statista' or German word 'statistik' or French word 'statistique', each of which means a political state. The word 'statistics' is used in singular as well as plural sense. Thus it is usually defined in two different senses, one is singular sense and other is plural sense. In singular sense, it means the statistical methods and techniques for dealing numerical data. The acts of dealing data are collection, presentation, analysis and the interpretation of the numerical figures. On the other hand, statistics means systematic collection of quantitative information of facts (or simply data) in plural sense.

i. In singular sense

Statistics means the science of statistical methods embodying the theory and techniques used for collecting, analyzing and drawing inferences from the numerical data. And it is defined in singular senses as follows:

"*Statistics is the science of the measurement of social organism, regarded as a whole in all its manifestations*". ***–A.L. Bowley***

"*The science and art of handling aggregate of fats, observing, enumeration, recording, classifying and otherwise treating them*". ***–Harlow***

"*Statistics may be regarded as the science of collection, presentation, analysis and interpretation of numerical data*". ***- Croxton and Cowden***

In singular sense, statistics means the various methods and techniques adopted for the collection, presentation, analysis and interpretation of the figures.

ii. In Plural Sense

Statistics is defined in plural sense as follows:

"*Statistics are the classified facts representing the conditions of the people in a state craft. Specially those facts which can be stated in number or in tables of numbers or in any tabular or classified arrangement*". ***–Webster***

"Statistics may be defined as the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard or accuracy, collected in a systematic manner, for a predetermined purpose and placed in relation to each other." ***- Prof. Horace Secrist***

In plural sense, statistics mean the quantitative information or numerical facts collected systematically.

But, for a layman, statistics is simply a mass of figures.

Hence, statistics is a science which studies the combination of the numerical data for analysis and interpretation as well as the methods and principles applied in collecting, presenting, analysis and interpreting the data under the study.

1.3 Division of Statistics

Nowadays, the word statistics is used in two contexts. In one context, singular sense, it refers to a subject of study that deals with various scientific methods, which are essential from the initial stage of data collection to the final stage of data presentation. In other context, plural sense (plural of the statistics), it refers to the numerical results obtained by applying statistical methods to a set of data. All the published numerical data on business, finance, population, health, environment etc. constitute statistics in the plural sense.

The subject of statistics is divided into following parts:

a. Mathematical statistics. b. Applied statistics.

c. Descriptive statistics. d. Inferential statistics.

Mathematical statistics deals with the development of statistical theory and methods based on certain principles and mathematics, while the applied statistics deals with the applications of statistical methods to the data. In this context, business statistics is considered as applied statistics. The statistics is also divided into descriptive and inferential statistics. Descriptive statistics is used to summarize or present the data, either numerically or graphically. Numerical descriptors of data include the followings:

i. One-way or two-way frequency tables.

ii. Various kinds of summary measures, such as mean, variance, correlation coefficient and so on.

ii. Statistical models.

While graphical summarizations include various kinds of charts and graphs, such as pie chart, line graph, scatter plot and so on. The main objective of this book is to describe descriptive statistics.

Inferential statistics is used to draw inferences about the population from sample data drawn from the population. These inferences may take the form of answers to yes/no questions (hypothesis testing), estimates of numerical characteristics (estimation). Inferential statistical methods involve quite advanced methods and will not be considered in this book.

Learning of the statistics will require the knowledge of basic mathematics. Like in other disciplines, statistics has its own vocabularies. An important and most frequently used term in statistics is variable. Whenever such terminologies are felt necessary to use we shall explain them.

If we observe or measure a characteristic we find that it takes on different values in different persons, places, or things, and it is customary to label the characteristics a variable. Some examples of variables include salary, age, sex, post, and educational degree of employees of a company, amount of daily sales of a store and marks obtained by the students in an examination. Data are considered as the values of a variable (in case of univariate data) or values of several variables (in case of multivariate data). The values of variable are generated by measuring, asking or observing the characteristics of similar subjects under the study.

Data generated by quantitative variables (such as salary, age, amount of sales, and so on) are called quantitative data, while the data generated by qualitative variables (such as sex, post, educational degree and so on) are called qualitative data or categorical data.

1.4 Functions of Statistics

Statistics is applied in everywhere therefore no nation exists without numerical fact in figure. The functions of statistics are as follows:

a. **Statistics simplifies the complexity:** The function of statistics is to present the huge mass of figures into a simple, presentable and understandable form. Using various statistical techniques, the complexity can be reduced in simplest form of the information obtained from the study.

b. **Statistics present the facts in a definite firm:** Another important function of statistics is to present the information of facts in a quantitative form. By definition, statistics presents any kind of information under the investigation in figure and number, so the conclusion stated in numerical figure is definite.

c. **Statistics provides techniques of comparison:** Different statistical methods and procedures facilitate comparison of the relevant features of several data. The statistical methods such as average, measure of dispersion, ratio etc. help in comparison between phenomena which enable to draw conclusion.

d. **Statistics helps in forecasting:** In business and industry, forecasting the future based on past experience and analyzing the historical tendencies is the most important task. This can be done with the help of several statistical techniques such as regression analysis, analysis of time series, index number etc. Prediction of future trends obtained from the application of statistical techniques in an investigation related to the business and management is important and is more convincing in framing plans and policies.

e. **Statistics helps in formulating policies:** The policy formulation of any firm, organization, business agency, bank, and nation will be suitable, if it is framed on the basis of statistical analysis and information. Making future policy is challengeable task. Statistics provides basic requirements for framing the future policies.

f. **Statistics gives the idea about possibilities of certain events:** Probability theory is one of the major areas of statistics. We can find the occurrence and non-occurrence of events with the help of probability laws and rules.

g. **Statistics helps in formulating and testing hypothesis:** To draw conclusion and to develop new theories in economics and business, formulation and testing the hypothesis is one of the important tasks. Formulating and testing the hypothesis without certain statistical techniques are incomplete. Testing hypothesis with the absence of statistical procedures and methods mislead the conclusion.

h. **Statistics helps to draw valid conclusion:** It is difficult task to enumerate each and every members of population (universe) under the study of some phenomena. The statistical technique such as sampling offers the best and scientific idea to study the sample group and generalize the conclusion so obtained in

the universe. An investigator or researcher can draw inferences about the whole universe by applying the different statistical techniques. Hence statistics helps to draw valid conclusion of the study.

i. **Statistics provides techniques for organizing data scientifically:** Collected data are in the raw form. They need to organize scientifically. Statistics provides techniques that help us to organize data scientifically. Nowadays it is customary to organize data in computer. Organization or management of the data is very essentials in modern research, since well organized data eases the work of data analysis and helps for drawing desired information very quickly.

1.5 Scope and Limitation of Statistics

Statistics is not only viewed as the device for collecting data but as a mean of technique for their handling and analysis as well as drawing inferences from them. Likewise, it is not the by-product of administrative set up of the state but it embraces all kinds of sciences such as social, physical and natural. Because of its widespread uses and applications in various diversified fields such as agriculture, industry, sociology, biometry, planning, economics, business, management, psychometry, insurance, accountancy etc, it is rather impossible to think any sphere of human activity where statistics does not creep on. Such wide use of statistics in the fields of the human activity shows the scope and importance of statistics. The scopes of statistics in different disciplines are discussed as follows:

i. **Statistics in planning**

To achieve the expected goal and objective, planning is the first constraint in any fields of the universe. Especially in the field of business and management, for the efficient working and formulating policy and decisions, the planning is restored. The statistical information related to production, consumption, prices, demand, supply, investment, income, expenditure etc, as well as the advanced statistical techniques such as index number, analysis of time series and regression analysis all are used in policy formulation and future planning of business organization, industry and state also. Now-a -days, efficient planning in every field is compulsory. Thus this modern age is also known as "*age of planning*".

ii. **Statistics in economics**

Statistical data and advanced techniques of statistical analysis solve the varieties of economic problems such as production, consumption, distribution of income and wealth, wages, prices, profits, savings, expenditure, unemployment, investment, poverty etc. Statistical techniques have been used in determining the measure of Gross National Product and Import-Export analysis. Furthermore, the advanced statistical techniques have been successfully used in the analysis of cost functions, production functions and consumption functions. Use of statistics in economics has led to the formulation and establishing the economic theory and laws such as Engel's law of consumption, Samuelson's Revealed Preference Analysis, Use of Analysis of Time Series, Index Number and Demand Analysis in Economic Planning, Development of New Discipline; Econometrics etc. Thus, the interaction between statistics and economics is the effective use of statistics in formulation of economic theories and economic policies. In fact, statistics got so much integrated with economics that it led to the development of a new subject called econometrics which basically deals with economic issues involving use of statistics.

iii. **Statistics in business and management**

It is universally accepted that statistical data and the powerful statistical tools such as probability theory, expectation, sampling techniques, and tests of significance, estimation theory, analysis of time series, index number, forecasting techniques and so on play indispensable role in decision making. The use of statistical data and techniques is indispensable in almost all branches of business. It is difficult to have success in business if careful study of the market is not made. Statistics helps in formulating

policies, forecasting the future based on the past experience and records the analysis of time series used in business for the study of trend in order to obtain the estimates of the probable demand of goods and seasonal phenomenon for determining '**Business Cycle**' which may also termed as the four phase cycle composed of prosperity, recession, depression and recovery. Likewise, another important statistical tool index number (economic barometer) enables the businessman to have an idea about the purchasing power of money.

Statistical techniques have been used by business organization and management in marketing decisions, investment, personal administration, credit policy inventory control, accounting and sales a control. The use and important of statistics in business and management is reflected by the visions of the following statisticians.

"*Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainty*" ***– Wallis and Roberts***

"*Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risks*" ***– Prof. Ya- Lun-Chou***

The uses of statistics in management are follows:

i. It helps for making policies, plans and programs.

ii. Data required for correct managerial decisions come from statistics.

iii. It helps to identify the factors and events responsible for the overall development of organizations.

iv. It provides a framework for the subject matter of investigation related to management.

v. It helps to find out the relationship between variables related to business and management.

And the uses of statistics in business are as follows:

i. In business, it can be used to understand the reasons why the share or commodity markets fall and rise. Based on such understandings, one can forecast future market and invest accordingly.

ii. It can be used to understand the reasons of changing behaviors of consumers from one brand of commodity to another or even the future demand of customers. A shrewd businessman can take advantage out of such understandings.

ii. Feasibility study of the market before launching a new product in the market is essential. Statistics can help to carry out such study.

iv. Executives can take data-driven decisions with the help of statistics.

Statistics in Computer Application

Statistics is not only viewed as the device for collecting data but as a mean of technique for handling and analysis as well as drawing inferences for them. Likewise, computers are ideally suited for data analysis concerning large research projects.

The computer is used for the following data analysis as following steps:

i. Data organization and coding.

ii. Storing the data in the computer.

iii. Selection of appropriate statistical measures/techniques.

iv. Selection of appropriate software package.

v. Execution of the computer program.

Statistics helps for creating mathematical model, logical consideration on theoretical bases and algorithms for computer science.

Hence, by computers, resolved data stored as well as problem of data processing for process of converting data into meaningful information becomes easy, fast and sage. The sufficient use of internet facilities allows electronic mailing device easy for the communication of data or information. Computer alone is not sufficient to resolve all the problems that arise during the time of data analysis and interpretation of information. So, statistics is inseparable in such cases.

1.5.1 Limitations of Statistics

The wide application of statistics signifies the importance of statistics in each and every field of sciences in the universe. However, it has some limitations. They are as follows:

a. **Statistics does not deal individually and works on aggregate level only:** Statistical findings are usually interpreted in terms of average, which may not be true to every individual. It deals with a group of individuals and aggregate of facts that indicates characteristics of the whole group. So, individual recognition is not mentioned.

b. **Statistics does not directly deal with qualitative phenomena:** Statistical method and technique is applicable for only the data expressed in numerical figure. It deals with the quantitative information under the investigation. It deals with qualitative characteristics such as intelligence, beauty, aptitude, knowledge etc, by changing them into numerical figure with the help of several tools.

c. **Statistical laws are not exact:** It is not an absolute measured itself. Most of the statistical analyze of data based on the statistical measures which are not absolute in nature. For example, correlation coefficient, skewness, kurtosis etc. Similarly, when one coin is tossed probability of getting a head is half but while tossing a coin six times probability of getting a head may not be half as before.

d. **Statistics may be misused:** Only a person who has got an expert knowledge of statistics can handle statistical data efficiently. If sufficient attention is not paid in collecting, analyzing and interpretation the data, statistical result might be misleading.

e. **Statistics cannot prove anything deductively**: The logic employed in inferential statistics is inductive in nature [drawing inference from small part (sample) to a larger part (population)], which is opposite to deductive logic used in mathematics. The deductive logic or argument does not prove anything.

f. **Statistical results are sometimes distrusting**: Sometimes conflicting nature of statistical statements are available in the literature, particularly in the medical sciences. For example, statements like "doing X reduces high blood pressure' and also statements like "doing X actually worsens high blood pressure". However, many readers may fail to notice these distinctions, or the media may oversimplify this vital contextual information and the public's distrust of statistics is thereby increased.

g. Statistical methods or techniques may be faulty in case of heterogeneous data.

h. Some errors are possible in statistical decision. Non- Statistical person do not know whether an error has been committed or not.

Distrust of Statistics

The improper use of statistical tools by the normal people having no detail idea of statistics led to the public distrust in statistics. If irresponsible, inexperienced and dishonest persons use statistical data and techniques, it loses public belief, faith and confidence. Thus, in the science of statistics, several distrusts may arise because of the following reasons:

i. Figures are innocent and believable.

ii. Figures put forwarded for arguments may be inaccurate and incomplete and thus distorting the truth.

iii. Though the figures are accurate, the dishonest person may mould and manipulate for personal and selfish motives. If so, public will be misinformed.

Hence, utmost care and precautions as far as possible should be taken for the interpretation of statistical data in all its manifestation.

1.5.2 Data Collection

This is the age of information and technology. The numerically expressed information is known as 'Datum'. So, Data is one of the main sources of information. The process of getting necessary information from the units under investigation is called collection of data. Collection of data is the first step in statistical investigation. The data collected constitutes the foundation of statistical analysis. Therefore, care must be taken while collecting data, otherwise the conclusion drawn can never be reliable. In the process of collecting data, the person from whom the information are collected known as '**Respondents**' and the person who conducts the statistical inquiry is known as '**Investigator**'.

It is important to note that information obtained from the data will able to answer our subject of enquiry; consequently, utmost care must be taken to collect as reliable and relevant data as possible. For this purpose, there are several crucial steps that need to be followed during the process of data collection in order to ensure that the data collection process and measurement systems are reliable. Incorporating these steps into a data collection plan will improve the likelihood that the data and measurements can be used to support the resulting analysis.

The investigator has to carry out preliminary analysis of the problem in question and also has to have clear cut decision taken on following points before starting to work of data collection; statement of the problem, scope of enquiry, sources of information, methods of data collection, unit of data collection, degree of accuracy and nature and type of enquiry.

Types and Sources of Data

For any statistical inquiry, the basic problem is to collect facts and information relating to a particular phenomenon under the study. Data are the raw materials for statistical analysis to draw a conclusion. Data may be either quantitative or qualitative in nature. The person who conducts the inquiry or collects the data for study is known as investigator and the person who gives the information to the investigator is known as respondents. The process of counting or enumeration together with the systematic recording of the information is called the collection of data.

It is accepted that the data collection is the first step for any type of statistical investigation. Thus, accuracy and preciseness of the study are based on the collected data for the study. This means the entire structure of the statistical analysis and interpretation is based upon the systemic way of collection of data which is reliable and adequate. This is why, before collecting the data for the statistical investigation, some points should be examined carefully. They are termed as preliminaries of data collection. The preliminaries of data collection are as follows:

a. **Objective and scope of inquiry**: It is essential to define the objective or purpose of inquiry clearly. This will enable investigator (researcher) to collect information properly. In the absence of objective of the inquiry, irrelevant information omitting important information may be collected and this may lead to fallacious conclusion of the study.

Scope of the inquiry means the coverage with respect to the type of information, subject matter and geographical area. It determines the size and type of sampling, selection of universe of the investigation and procedures. Thus the decision about the type of inquiry to the conducted research or investigation depends upon the objectives and scope of the inquiry.

b. **Statistical units to be used:** A well- defined and identifiable object or a group of objects with which the measurements or counting any statistical investigation are associated is called a statistical unit. For example, an individual person, a family, a shop of locality etc. in a social survey. A very important step before the collection of data is to define clearly the statistical units on which data are to be collected.

c. **Source of information:** For any statistical inquiry, the investigator may collect the information first or he/she may use the data from other published sources. The data collected originally by the investigator for the first time for the study, the data are known as primary data and if he/she uses the data, which had already been collected such as publications and reports of government/ semi-government/non-government organizations, magazines, newspapers, research journals etc., the data are known as secondary data.

d. **Method of data collection:** The next thing is to decide method of data collection. If primary data are to be collected, a decision has to be made whether census method or sample technique is to be used for data collection. In case of primary data, a choice of census method and sample method depends upon the objectives and scope of the study, the limitations of resources in terms of time, money, manpower etc, and in case of secondary data, testing and editing the reliability, suitability, adequacy and accuracy of the data are to be carefully kept in mind of investigator.

e. **Degree of accuracy aimed in the final results:** The information gained from any already completed sample study on the subject in the precision achieved for a given sample size may serve as a useful guide in this matter provided there is no fundamental reason to this empirical basis. In any statistical enquiry, perfect accuracy in final results is practically impossible to achieve because of errors in measurement, collection of data, analysis of the data and interpretation of the results. It should not be understood to imply that one should sacrifice to conduct the enquiry at low costs.

f. **Types of enquiry:** Another point has to be kept in mind before collecting the data is to decide the type of inquiry. Several types of enquiry are

i. Official, Semi-official or Unofficial
ii. Initial or Repetitive
iii. Confidential or Open
iv. Direct or Indirect
v. Regular or Ad-hoc
vi. Census or Sample
vii. Primary or Secondary

1.5.3 Types of Data

For any investigation or enquiry, the collection of data is most important because data are the raw materials of the enquiry for its final conclusion. All the information (quantitative or qualitative) collected from the respondents and to use for the purpose of statistical analysis, it is termed as data. Based on the sources, the data generally are classified in two types.

They are

i. Primary data
ii. Secondary data.

1.6 Primary Data and Methods of Primary Data Collection

1.6.1 Primary Data

The data which are originally collected by an investigator for the first time for any statistical analysis are known as primary data. The data are fresh, first hand and original in nature. The primary data are collected for the certain purpose of study or investigation. The source of this type of data is called primary source of data. For example, if an investigator wishes to study the average marks of students in

statistics in a college then the data collected for this purpose by the investigator himself/herself are primary data.

1.6.2 Methods of Collection of Primary Data

The following methods are the commonly used for the collection of primary data. They are as follows:

1. Direct personal interview
2. Indirect personal investigation method
3. Mailed questionnaire method
4. Schedules sent through enumerators
5. Information from local correspondents

1. Direct personal interview method: It is used to collect the data by the investigator from the respondents directly. In this method, the investigator meets the respondents personally and makes the necessary inquiries and extracts the required data from them. Thus, it is suitable if the enquiry is intensive rather than extensive. Since such investigations require personal attention of the investigator, the information, gathered from such investigation is original in nature. Investigator can collect additional information as per his/her need.

Merits of direct personal interview method are as follows:

i. Information collected using this method is accurate and original

ii. When the audience is approached personally by the investigator, the response is likely to be more encouraging.

iii. The data collected using this method is reliable.

iv. This method is flexible.

v. The investigator can extract proper information from the respondents talking to them at their academic level in their language of communication.

vi. The personal biases in responding can be detected asking cross questions to the respondents.

Demerits of direct personal interview method are follows:

i. This method is only suitable in intensive studies.

ii. It is not useful for the wider area of inquiry.

iii. It is expensive in terms of time, money and manpower.

iv. This method requires intelligent, skillful, trained, tactful and courageous manpower. Otherwise the inquiry cannot be reliable valid and satisfactory.

v. The respondent may give the biased information which leads the wrong conclusion of the investigation.

2. Indirect personal investigation method: This method is used when the informants are reluctant to give the definite information. Information regarding the property, income, personal habits like smoking habits, drug addicts, alcoholism, girl trafficking and disease like HIV/AIDS etc., the respondents hesitate to provide true information. In such cases indirect oral investigation is more practicable and suitable. In this method, third person is used to collect the information, they are called **witness**. The police report is one of the examples of the indirect oral investigation method.

Merits of indirect personal investigation are:

i. This method is less expensive than the direct personal interview method

ii. The expert views and suggestions on the problem can be solicited.

iii. This method is convenient gathering sensitive information with the help of witness.

iv. A wide area can be covered for investigation.

v. This method is appropriate in the investigation if the respondents are reluctant to give information.

Demerits of indirect personal investigation are:

i. Due to lack of direct supervision and touch, the information may be inaccurate and unreliable.

ii. If wrong and improper witness is selected, the information given by them will be biased.

iii. There will be lack of interest and willingness of the witness to give the information.

3. Mailed questionnaire method: A set of questions relating to the subject of inquiry is known as questionnaire. A space for the answers/response to be filled by the respondents is provided. A questionnaire is mailed to the respondents with request for quick response within the specified time and return to the investigator. Respondents must be educated in this method. When the field of investigation is large and the investigator requires quick result at low cost, this method is more suitable in practice than other methods.

Merits of mailed questionnaire method are as follows:

i. This method is economic in terms of time, money and manpower.

ii. This method is used for extensive inquiries covering a very wide area.

iii. It is assumed that educated persons never lie. So, the information obtained is original and authentic.

iv. Errors due to personal biases of the investigators are eliminated.

Demerits of mailed questionnaire method are as follows:

i. This method of data collection is not applicable for the uneducated respondents.

ii. In this method, the informants may feel fear to response questionnaire, so there is high degree of non-response error.

iii. Respondents may reply the wrong answers (information) to questionnaire.

iv. Some questions contained in the questionnaire may affect the feelings of respondents.

4. Schedules sent through enumerators: This method is distinct from the questionnaire method in process gathering information. As we discussed questionnaire method above, information or response to the questions are filled by the respondents themselves and returns to the investigator. And in this method investigator select (appoint) the enumerator or agent and give training to collect information from the field of inquiry. Investigator sent enumerators with schedule (a list of questions) to the respondents and they ask questions to the respondents and record their replies. This method is more practicable when the respondents are illiterate. This method is generally applied in population census conducted by the government.

Merits of schedules through enumerator method are as follows:

i. Non- response error can be minimized using this method.

ii. Information through this method is more accurate than mailed questionnaire method.

iii. This method can be used even when the respondents are illiterate.

iv. The enumerators can check the accuracy of the information by asking cross questions.

v. It is suitable in wide area of investigation.

Demerits of schedule sent through enumerator method are as follows:

i. This method is expensive regarding time, cost and man power.

ii. Enumerators must be well -trained. Otherwise information collected by them may not be correct that results fallacious conclusion of the investigation.

iii. The enumerator may not be responsible well in collecting information.

iv. The respondents may not believe the enumerators.

5. **Information through correspondent or local agency method:** In this method, the information is not formally collected by the investigator or the enumerators. The investigator appoints the correspondent or local agent/agency in different places for the inquiry to collect the information. These correspondents or local agencies/agents collect the information in accordance with appropriate/suitable ways and then submit their reports to the central or head office where the data are processed for the final analysis. This method of data collection is usually used by the media agency.

Merits of information through correspondent method are as follows:

i. This method of data collection can cover wide area of inquiry.

ii. It is cheapest method.

iii. This method is more appropriate to get regular information.

Demerits of information through correspondent method are follows:

i. Due to the personal prejudice and biasness of correspondent, the information may not be accurate and reliable.

ii. It is inconvenient and time consuming to use in the absence of facility of communication such as telephone, internet etc.

1.6.3 Problems of Primary Data Collection

While collecting primary data, various problems have to be faced. The problems or difficulties which arise during of primary data collection and secondary data collection are discussed below:

Problems in collecting primary data are as follows:

i. Dishonest and irresponsible respondents may show non- response character.

ii. Real information may be collected if the respondents do not understand questions clearly sent by the investigator.

iii. A problem may arise due to lack of transportation facility.

iv. There is high degree of non- response error from the uneducated respondents.

v. Personal biasness and prejudice of enumerators cause the fallacious conclusion of investigation.

vi. It is required that the expert, knowledgeable, skillful, trained and intelligent manpower for preparing questionnaire and collecting primary data which is not possible and accessible for all cases.

vii. If the scope of inquiry is wide, money, time and manpower should be sufficiently available. Otherwise, inadequate and inaccurate data may get.

1.7 Secondary Data and Sources of Secondary Data Collection

1.7.1 Secondary Data

The data which have already been collected and processed by some agency or person and taken over from there and used by any other agency or person for their statistical analysis are known as secondary data. Such data may not be original in nature. Thus secondary data are less accurate than primary data. In

a case of some inquiry, collection of primary data is not always practicable due to availability of time, money and manpower. There is a lot of published and unpublished information from which further studies can be made.

1.7.2 Sources of Secondary Data Collection

The sources of secondary data mainly classified in two types. They are:

i. Published sources

There are number of national and international organizations or agencies who collect statistical data relating to business, trade, education, health, population, poverty, consumption, import, export, money exchange etc. and publish their findings in reports. These publications serve as the source of secondary data.

Published sources of secondary data are as follows:

i. Reports and publications of ministries, departments of the government, semi- government offices, NRB, FNCCI etc.

ii. Reports and publications of worldwide reputed INGO'S such as UNDP, UNO, UNESCO, WTO, WHO, SAARC, World Bank etc.

iii. Reports and publications of reputed NGO's, research journals, periodicals, dissertations etc.

iv. Reports of various committees and commissions appointed by the Government.

ii. Unpublished sources:

All the statistical data need not be always published.

There are various sources of unpublished statistical data. They are

i. Records maintained by government offices.

ii. Researches carried out by the individual research scholars faculty members in the universities.

iii. Records updated by the departments, institutions for their internal purpose.

iv. Records maintained by private firms or business enterprises which they do not like to publish.

1.8 Merits and Demerits of Secondary Data

1.8.1 Merits of secondary data are as follows:

a. It saves time and money.

b. The scope of inquiry can be increased in terms of area and time period.

c. If investigator is expert and skillful to filter and gather the required information, the quality of secondary data is better.

d. Investigation using secondary data can have more references to consult.

Demerits of secondary data are as follows:

a. Data may not be in the exact form of the requirement of the investigator.

b. The data may be outdated.

c. When secondary data are gathered from two different sources, it may not be comparable in terms of definition, units and time period covered.

d. The degree of accuracy and reliability of the investigation should rely on previous investigation or researches.

e. Exact definition of units and terms used in secondary data may be unknown.

Precautions in Using Secondary Data

Before using the secondary data, the investigator must be careful on some factors of the data to the given problem under investigation. They are as follows:

i. Reliability of Data: Actually, reliable data used in study ensure the reliability of the conclusion to be drawn. The result of investigation or study to the given problem will mislead if the data are not reliable. It is essential that the agency or organization should be unbiased in collecting data in sense that it had no personal motives and interest. So, the investigator should be careful about the following points:

i. The reliability, integrity of the organization or institutions.
ii. The reliability of sources of information.
iii. The methods used for the collection of the data and
iv. The techniques and procedures used to analyze the data.

ii. Suitability of Data: Even if the data are reliable, but may not be suitable for the purpose of inquiry under study. Thus it is confirmed first before to use that whether the data is suitable for the study or not. For this, it is necessary to observe the homogeneity in terms of objective, nature, scope, condition, terms and units used in the original inquiry and investigation in hand.

iii. Adequacy of Data: The data collected for a study which is suitable to use in another investigation and reliable but even one thing is to be considered that whether the data is adequate for the study or not. Suitable and reliable data of a study may not be adequate and sufficient for other study or investigation. To draw a valid conclusion and to obtain the reliable result of study using secondary data, the data must be checked by the investigator whether the data is reliable or not first, suitable or not in second and at last and right before to use, it should be examined the adequacy of the data. Otherwise, objective of the study may not be gained.

Problems in Collecting Secondary Data

The problems of collecting secondary data are as follows:

a. Suitable, reliable and adequate data to the inquiry under the investigation can be rarely obtained. Sometimes, the unpublished data cannot be obtained as the person hesitates to give it.
b. Irrelevant or duplicate data may be collected.
c. The data collected could be flawed or misinterpreted.
d. The secondary data which are suitable and reliable for a study but that may not be even adequate for the study.

Differences between Primary Data and Secondary Data

As we have discussed above, the data which is primary for one investigator is treated as secondary for the other. Mainly they are different in terms of method and mode of collection of data. Some of the inquiries require primary data only, likewise secondary data are sufficient and adequate in several inquiries and both type of data may be essential for some investigation. However primary data and secondary data are different with each other. They are as follows:

Primary data	Secondary data
i. Primary data are accurate and original. ii. Methods of primary data collection are more expensive in terms of time, cost and manpower.	i. Secondary data may not be accurate and original in the sense that they are collected by other.

iii. Primary data may be influenced by personal biasness and prejudice of the investigator. iv. It is mostly used in establishment of new theory. v. It is collected as per the objective and scope of investigation to be carried out by investigator.	ii. Methods of secondary data are less expensive. iii. Secondary data may not be influenced by personal biasness and prejudice of investigator. iv. It is mostly used in statistical investigation and analysis. v. It might have been collected with the different objectives.

Exercise 1.1

Theoretical Questions

1. Define statistics and its types. Discuss its functions and limitations.

2. Define statistics and explain its uses and applications in computer application.

3. Why the people distrust statistics? Justify with reasons.

4. Define primary data and explain the problems of collecting primary data.

5. Define secondary data. What are the problems of collecting secondary data? Explain.

6. Differentiate between primary data and secondary data.

7. Discuss the methods of primary data collection.

8. Explain questionnaire method with its merits and demerits.

9. Differentiate direct personal interview and indirect personal investigation method.

10. What are the sources of secondary data? What precautions should be kept in mind before to use the secondary data?

11. Explain application off statistics in the field of computer application?

Exercise 1.2

Multiple Choice Questions: Circle (O) the correct answer.

1. How many types of data are there on the basis of sources of data collection?
(a) 1 (b) 2 (c) 3 (d) 4

2. The statement, "Statistics is both a science and an art", was given by:
(a) R.A. Fisher (b) Tippet (c) L.R. Connor (d) A.L. Bowley

3. Who stated that statistics is a branch of applied mathematics which specializes in data?
(a) Horace Secrist (b) R.A. Fisher (c) Ya-lun-chou (d) L.R. Connor

4. The word "statistics" is used as:
(a) Singular (b) Plural (c) Singular and plural both
(d) None of the above

5. "Statistics provides tools and techniques for research workers", was stated by:
(a) John I. Griffin (b) W.I. King (c) A.M. Mood (d) A.L. Boddington

6. Data taken from the publication, 'Agricultural Situation in Nepal' will be considered as:
(a) primary data (b) secondary data (c) primary and secondary data

(d) neither primary nor secondary data

7. Which of the following represents data?

(a) a single value
(b) only two values in set
(c) a group of values in a set
(d) none of the above

8. Statistics deals with:

(a) qualitative information
(b) quantitative information
(c) both (a) and (b)
(d) none of (a) and (b)

9. Statistical results are,

(a) cent per cent correct
(b) not absolutely correct
(c) always incorrect
(d) misleading

Answer Key

1. (b)	2. (b)	3. (b)	4. (c)	5. (c)	6. (b)	7. (c)	8. (b)	9. (b)