# Statistics - Scope and Development

1.1 History of Statistics  
1.2 Definition of Statistics  
1.3 Functions of Statistics  
1.4 Scope and importance of Statistics  
1.5 Limitations of Statistics  
1.6 Some applied areas of Statistics  
1.7 Official Statistics

## Collection of Data

2.1 Data Collection  
2.2 Variables  
2.3 Types of Data  
2.4 Questionnaire and Schedule  
2.5 Methods of Primary Data Collection  
2.6 Sources of Secondary Data
Introduction

Decision making is one of the highest forms of human activities. Every day we make decisions that may be personal, business related or of some other kind. Usually these decisions are made under conditions of uncertainty. Many times the situations or problems we face in the real world have no precise or definite solution. Statistical methods help us to make scientific and intelligent decisions in such situations. In recent years, the growth of statistics has made itself felt in almost every phase of human activity. Statistics no longer consists merely of collection of data and their presentation in charts and tables. It is now considered the science of inferences on observed data and the entire problem of making decisions in the face of uncertainty. This covers considerable ground since uncertainties are met when we flip a coin, when a dietician experiments with food additives, when an actuary determines life insurance premiums, when a quality control engineer accepts or rejects manufactured products, when a teacher compares the abilities of students, when an economist forecast trend, when a newspaper predicts an election result and so forth.

It would be presumptuous to say that statistics in its present state of development can handle all situations involving uncertainties, but the new techniques are constantly being developed and modern statistics can, provide the framework for taking at these situations in a logical and systematic fashion. The beginning of mathematics of statistics may be found in mid-eighteenth century studies in probability motivated by interest in game of chance. Thus the scholars began to apply probability theory to actuarial problems to some aspects of social
sciences. By this century it found application in all phases of human endeavour that in some way involve an element of uncertainty or risk.

Like almost all fields of study, statistics has two aspects, Theoretical and Applied. Theoretical or Mathematical Statistics deals with development, derivations and proof of statistical theorems, formulæ, rules and laws. Applied statistics involves the application of those theorems, formulæ, rules and laws to solve real world problems. Broadly speaking, applied statistics can be divided into two areas, Descriptive Statistics and Inferential Statistics. Descriptive statistics consists of methods for analysis of data and the area that deals with decision making procedure is referred to as inferential statistics.

1.1 History of Statistics

The word Statistics have been derived from Latin word “Status” or the Italian word “Statista”. The meaning of these words is “Political State” or a Government. Shakespeare used a word Statist in his play Hamlet (1602). In the past, the statistics was used by rulers for official purposes. Even though application of Statistics was very limited, the rulers and kings needed information about lands, agriculture, commerce, population of their states to assess their military potential, their wealth, taxation and other aspects of Government.

Gottfried Achenwall used the word ‘statistik’ at German University in 1749 which means political science of different countries. In 1771, W. Hooper (Englishman) used the word ‘statistics’ in his translation of Elements of Universal Erudition written by Baron B.F Bieford. In his book, statistics has been defined as the science that teaches us what is the political arrangement of all the
modern states of the known world. There is a big gap between the old statistics and the modern statistics, but old statistics is also used as a part of the present statistics.

During the 18th century English writers have used the word statistics in their works. A lot of work has been done in the end of the nineteenth century.

At the beginning of the 20th century, William S Gosset developed the methods for decision making based on a small set of data. During the 20th century several statisticians were active in developing new methods, theories and application of statistics. The advent of electronic computers is certainly a major factor in the development of modern statistics. Sir Ronald Aylmer Fisher is known as father of modern statistics.

### 1.2 Definition of Statistics

1. “Statistics can be defined as the collection, presentation and interpretation of numerical data.” - Croxton and Cowden.

2. “Statistics are measurement, enumerations or estimates of natural or social phenomena, systematically arranged to exhibit their inner relation.” - Conner.

3. “The science of Statistics is essentially a branch of applied mathematics and can be regarded as a mathematics applied to observational data.” - R.A Fisher.

4. “Statistics means aggregate of facts affected to marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.” - Horace Secrist

This definition points out some essential characteristics of statistics. These
characteristics are:

(i) **Statistics are the aggregates of facts.** It means, a single figure is not statistics. For example, national income of a country for a single year is not statistics but the same for two or more years is statistics.

(ii) **Statistics are affected by a number of factors.** For example, sale of a product depends on a number of factors such as its price, quality, competition, the income of the consumers, and so on.

(iii) **Statistics must be reasonably accurate.** Wrong figures, if analysed, will lead to erroneous conclusions. Hence, it is necessary that conclusions must be based on accurate figures.

(iv) **Statistics must be collected in a systematic manner.** If data are collected in a haphazard manner, they will not be reliable and will lead to misleading conclusions. It is collected with a pre-determined purpose.

(v) **Statistics should be placed in relation to each other.** If one collects data unrelated to each other, then such data will be confusing and will not lead to any logical conclusions. Data should be comparable over time and over space.

### 1.3 Functions of Statistics

1. **Statistics simplifies complexity**
   
   The complex mass of data are made simple and understandable with the help of statistical methods.

2. **Statistics presents facts in a definite and precise form**
   
   Statistics presents statements of facts in a precise, quantitative and definite form.

3. **Statistics provides comparison**
1.4 Scope and importance of Statistics

Statistics provides a number of suitable methods of comparison between present and past values and hence able to predict future.

4. Statistics enlarges human knowledge and experience

Statistics makes most of our vague and indefinite opinions, clear and definite.

5. Statistics helps in formulating policies, testing of hypotheses and forecasting future events

Important policies, decision making and forecasting in business, economics, finance, industry, etc are taken on the basis of statistical methods.

1.4 Scope and importance of Statistics

1. Statistics and Planning: Statistics is an indispensable tool in planning the modern age. Because of the complexities and uncertainties, planning is essential for solving the complex problem in various walks of life.
2. **Statistics and Economics:** Statistical data and techniques of statistical analysis such as time series analysis and demand analysis are immensely useful in solving problems in Economics.

3. **Statistics and Industry:** In industry, Statistics is widely used in quality control. In production engineering, to find out whether the product is confirming to the specifications or not. Statistical tools, such as inspection plan, control chart, etc. are highly useful.

4. **Statistics and Mathematics:** Statistics is intimately related to Mathematics. Statistical techniques are the outcomes of wide applications of Mathematics.

5. **Statistics and Medical Science:** In medical science the statistical tools for collection, presentation and analysis of observed facts relating to causes and incidence of disease and the result of application of various drugs and medicines are of great importance.

6. **Statistics, Psychology and Education:** In Education and Psychology, Statistics has found wider applications such as, determining (or to determine) the reliability and validity of a test, measuring intelligence quotient, factor analysis, etc.

7. **Statistics and Management Studies:** Statistical analysis is frequently used in providing information for making decisions in the field of marketing, production, finance, banking, investment, purchase and accounting.

**Activity**
Prepare a report regarding the functions and importance of statistics in daily life by reading the features and reports in newspapers and magazines.

### 1.5 Limitations of Statistics

(i) There are certain phenomena or concepts where Statistics cannot be used. For example, beauty, intelligence and courage cannot be quantified. Statistics has no place in all such cases where quantification is not possible.
(ii) Statistics reveals the average behaviour, the normal or the general trend. Statistics does not study individual items but deals with aggregate. For example, one may be misguided when told that the average depth of a river from one bank to the other is four feet. There may be some points in between where its depth is far more than four feet.

(iii) Since statistics are collected for a particular purpose, such data may or may not be relevant or useful in other situations or cases. For example, secondary data (i.e., collected by a person) need not be useful for another person.

(iv) Statistics are not 100 per cent precise as in Mathematics. Those who use Statistics should be aware of this limitation.

Misuse of Statistics

The misuse of Statistics is the main cause of discredit to this science and has led to public distrust in Statistics. The various reasons of misuse are:

(i) Sources of data not given.

(ii) Defective data.

(iii) Unrepresentative sample.

(iv) Inadequate sample.

(v) Unfair Comparisons.

(vi) Unwanted conclusions.

(vii) Inappropriate statistical tools.
1.6 Some applied areas of Statistics

Actuarial Science

Actuarial science is the discipline that applies mathematical and statistical methods to assess risk in the insurance and the finance sectors. Actuaries are professionals who are qualified in this field. In many countries, actuaries must demonstrate their competence by passing a series of rigorous professional examinations. Actuarial science includes a number of interrelating subjects, including Probability, Mathematics, Statistics, Finance, Economics, Financial Economics, and Computer Programming. Historically, actuarial science used deterministic models in the construction of tables and premiums. The science has gone through revolutionary changes during the last 30 years due to the proliferation of high speed computers and the union of stochastic actuarial models with modern financial theory (Frees 1990).

Biostatistics

Biostatistics (sometimes referred to as Biometry or Biometrics) is the application of Statistics to a wide range of topics in Biology. The science of Biostatistics encompasses the design of biological experiments, especially in medicine, agriculture and fishery; the collection, summarization, and analysis of data from those experiments; and the interpretation of, and inference from, the results. A major branch is Medical Biostatistics, which is exclusively concerned with medicine and health.
Agricultural Statistics

The agricultural investigations are based on the application of statistical methods and procedures which are helpful in testing hypotheses using observed data, in making estimations of parameters and in predictions. The application of statistical principles and methods is necessary for effective practice in resolving various problems that arise in the many branches of agricultural activity. Because of the variability inherent in biological and agricultural data, knowledge of statistics is necessary for their understanding and interpretation. Numerous activities in agriculture are very different from each other, resulting in different branches of agricultural science like: field crop production, vegetable production, horticulture, fruit growing, plant protection, livestock, veterinary medicine, agricultural mechanization, water resources, agricultural economics, etc.

Activity

List out the various branches of statistics related to different disciplines.

1.7 Official Statistics

Official Statistics are statistics published by government agencies or other public bodies such as international organizations. They provide quantitative or qualitative information on all major areas of citizens’ lives. Official Statistics make information on economic and social development accessible to the public, allowing the impact of government policies to be assessed, thus improving accountability.

The Ministry of Statistics and Programme Implementation (MOSPI) came into
Statistics - Scope and Development

existence as an Independent Ministry in 1999 after the merging of the Department of Statistics and the Department of Programme Implementation.

The Ministry has two wings, Statistics and Programme Implementation.

- SDRD - Survey Design and Research Division
- FOD - Field Operations Division
- DPD - Data Processing Division
- CPD - Co-ordination and Publication Division

The Statistics Wing called the National Statistical Office (NSO) consists of the Central Statistical Office (CSO), the Computer Centre and the National Sample Survey Office (NSSO).

Central Statistical Office (CSO)

The Central Statistical Office which is one of the two wings of the National Statistical Organisation (NSO) is responsible for co-ordination of statistical activities in the country and for evolving and maintaining statistical standards. Its activities include compilation of National Accounts; conduct of Annual Survey of Industries and Economic Censuses, compilation of Index of Industrial Production as well as Consumer Price Indices. It also deals with various social statistics, training, international cooperation, Industrial Classification, etc.
The CSO is headed by a Director-General who is assisted by 5 Additional Director-Generals looking after the National Accounts Division, Social Statistics Division, Economic Statistics Division, Training Division and the Coordination and Publication Division.

CSO is located in the Sardar Patel Bhawan, Parliament Street, New Delhi. The Industrial Statistics Wing of CSO is located in Kolkata. The Computer Centre also under the CSO is located in R K Puram, New Delhi.

**National Sample Survey Office (NSSO)**

The National Sample Survey Organisation, now known as National Sample Survey Office, is an organization under the Ministry of Statistic of the Government of India. It is the largest organisation in India, conducting regular socio-economic surveys. It was established in 1950.

NSSO has four divisions:

1. Survey Design and Research Division (SDRD)
2. Field Operations Division (FOD)
3. Data Processing Division (DPD)
4. Co-ordination and Publication Division (CPD)

The Programme Implementation Wing has three Divisions, namely,

(i) Twenty Point Programme
(ii) Infrastructure Monitoring and Project Monitoring
(iii) Member of Parliament Local Area Development Scheme.
Besides these three wings, there is National Statistical Commission created through a Resolution of Government of India (MOSPI) and one autonomous Institute, viz., Indian Statistical Institute declared as an institute of National importance by an Act of Parliament.

Know your progress
Discuss the important statistical organizations (offices) in India.

Indian Statistical Institute (ISI)

Indian Statistical Institute (ISI), a unique institution devoted to the research, teaching and application of statistics, natural sciences and social sciences. Founded by Prof. Prasanta Chandra Mahalanobis in Kolkata on 17th December, 1931. He is known as father of Indian statistics. The Indian Statistical Institute publishes Sankhya, the Indian Journal of Statistics.

In recognition of the notable contributions made by Prof. P.C. Mahalanobis in the fields of economic planning and statistical development in the post independent era, the Govt. of India has decided to designate 29th June every year, coinciding with his birth anniversary, as Statistics Day in the category of special day to be celebrated at the national level. The Day is celebrated by holding seminars, discussions and competitions to highlight the importance of official statistics in national development.
Economics & Statistics Department

The Directorate of Economics & Statistics, Government of Kerala is the nodal agency of the State responsible for the systematic collection, compilation, analysis, interpretation and dissemination of statistics relating to various sectors of Kerala Economy.

The Directorate of Economics & Statistics is the nerve centre of the State statistical system. Director is the technical and administrative head of the Department. Being the statistics authority of the State the director functions as the authority for the collection, processing and dissemination of all statistical data relating to the State economy.

Besides the Directorate there are 14 District Offices, each headed by a Deputy Director with the exception of Wayanad. The Deputy Director in the District Offices is assisted by one District Officer, one or more Additional District Officers, one Price Supervisory Officer and one or two Research Officers. At taluk level, there is a Taluk Statistical Office, which is the lowest statistical unit in the State. There are at present 61 Taluk Statistical Offices, each under the control of a Taluk Statistical Officer.

Activity

Visit the nearest economics and statistics department and prepare a detailed report regarding their functions.

Let us sum up

Statistics are all around us. Without statistics we couldn’t plan our budgets, pay our taxes, enjoy games to their fullest, evaluate classroom performance, etc. In this chapter we discussed the history, importance, development, scope and some definitions of statistics. Statistics is applied in all walks of life. Various branches of statistics are explained here. We have seen the functions and roles of the ministry of statistics and programme implementation and the famous Indian Statistical Institute in Kolkata. We also introduced the Department of Economics and Statistics of the state.
Learning outcomes

After transaction of this unit, the learner:-

• explains the history, definitions and scope of Statistics.
• recognises the importance of Statistics in various fields.
• compares different branches of Statistics.
• illustrates the functions of MOSPI, CSO, NSSO, ISI and Department of Economics and Statistics of Kerala.

Evaluation Items

1. “Statistics can be defined as the collection presentation and interpretation of numerical data”. This definition is given by:
   a) R.A Fisher          b) Horace Secrist
   c) Croxton and Crowden d) Conner.

2. Who is known as father of Modern Statistics?
   a) Conner          b) R.A Fisher
   c) Mahalanobis     d) Gosset

3. The discipline that applies mathematical and statistical methods to assess risk in the insurance and finance industries is called
   a) Bio statistics b) Agricultural statistics
   c) Actuarial Statistics d) Production Statistics

4. The Central Statistical Office is located in:
   a) Mumbai          b) Kolkatta
   c) New Delhi       d) Chennai

5. The largest organisation in India conducting regular socio-economic surveys is
   a) CSO              b) NSSO
   c) ISI             d) NASA
6. Indian Statistical Institute (ISI), is founded by:
   a) P.C. Mahalanobis    b) R.A Fisher
   c) Horace Secrist      d) C. R. Rao

7. The Indian Statistical Institute (ISI) is situated in:
   a) Kolkata           b) Bangalore
   c) Chennai           d) Pune

8. National Statistics Day is celebrated on:
   a) June 1            b) June 29
   c) July 4            d) July 29

9. .......... is known as the father of Indian Statistics.
   a) R.A Fisher        b) S.P. Gupta
   c) C.R. Rao          d) P.C. Mahalanobis

10. The journal published by Indian Statistical Institute (ISI) is
    a) Statistica        b) Sankhya
    c) Sample surveys    d) Census

11. Name the nerve centre of Kerala state statistical system.

12. “In this century statistics found application in all phases of human endeavor”. Comment on the statement.

13. How will you critically approach the definition of statistics given by Horace Secrist?

14. Examine the scope of statistics in various fields.

15. “The number of accidents is lower in foggy weather than on clear days. Hence it is safe to drive in fog.” Do you agree with the statement? Why?

16. Explain the importance of Statistics in the following branches of study
    a) Actuarial science  b) Bio statistics  c) Agricultural Statistics

17. Write short notes on the following:
    a) CSO  b) NSSO  c) ISI
18. What are the Divisions of NSSO?


Answers:

1) c  2)b  3)c  4)c  5)b  6)a  7)a  8)b  9)d  10)b
Introduction

In chapter 1, we discussed Statistics as the study of collection, organization, analysis, interpretation and presentation of data. For studying statistics the first step is collection of data, which we will discuss in detail in this chapter.

2.1 Data Collection

In our day to day life we deal with different types of data collection situation. A teacher might collect information regarding the test score of a student, a journalist might collect information regarding the recent social issues, a politician collects information on how voters plan to vote in the upcoming election, etc. Data collection is the systematic gathering of data for a particular purpose from various sources.

Data is the plural of the term datum, which means any measurement, result, fact or observation which gives information. Statistical surveys are the most popular devices for obtaining the desired data. Before dealing with statistical surveys, we have to familiarize with the following terms.

Statistical Investigation

Statistical investigation includes collection, classification, presentation, analysis and interpretation of data according to well defined procedures. The person authorized to make investigation is known as Investigator. In a statistical investigation the investigator formulates the problem, suggests the data collection methods, organises various steps in an appropriate way, analyses the data and interpret the result. Usually, the investigators depute some persons to collect the data from the field. These persons are known as Enumerators. The enumerator may not be aware of the investigation procedures completely.
His/her duty is to collect the data for the investigator. It is the duty of the investigator to train and supervise the work of the enumerator. The process of data collection by the enumerator is known as **Enumeration**.

### Population and Sample

A **population** consists of all elements, individuals, items or objects whose characteristics are being studied. For a politician, while considering voters plan for the next election, all registered voters in the specified constituency determines the population. If data are collected from each and every unit of the population, the investigation is called **census**. Based on the number of objects in a population, we can classify the population as finite and infinite. A population is said to be **finite**, if the number of individuals involved in the population is finite. All students of Kerala University for the year 2013-14 constitute a finite population. A population which is not finite or extremely large is **infinite**. The population comprises of all people in the world above 18 years of age is considered as an infinite population.

If the population is infinite or is of extremely large size, it is not feasible or practicable to access the entire population for study. As a result, it is apt to take a representative part as a substitute for the entire population. This representative part of the population is known as **sample**. The method of collecting data from the sample is known as **sampling** or **sample survey**. Various sampling designs and their selections are discussed in the last chapter.

The origin of descriptive statistics can be traced to data collection methods used in censuses taken by the Babylonians and Egyptians between 4500 and 3000 BC. In addition, the roman Emperor Augustus (27 BC to 17 AD) conducted surveys on births and deaths of the citizens of the empire, as well as the number of livestock each owned and the crops each citizen harvested yearly. In India about 2000 years ago we had an efficient system of collecting administrative statistics, particularly,during the regime of Chandra Gupta Maurya (324 to 300 B.C.). The system of collecting data related to births and
2.2 Variables

Consider a group of people in a locality. The members of the group are found to be varying in many factors like sex, age, eye colour, intelligence, height, weight, blood pressure etc. The factors which can vary from one object to another are called variables. Among these variables, sex, eye colour and intelligence which cannot be numerically measured are called qualitative variables or attributes. A qualitative variable is one that can be identified by noting its presence or identified with different categories of the factor. The other variables height, weight, age and blood pressure which are numerically measured are called quantitative variables. A quantitative variable consists of numerical values. Depending on the values taken by a quantitative variable, it is further classified as discrete variable and continuous variable. If the variable takes specific values only, it is called discrete variable. The variable, number of children in a family, does not take values other than 0, 1, 2, 3, etc. That is, there is a specified
break between the successive values. This is an example of a discrete variable. A continuous variable takes any value within the defined range of values. Between any two values of a continuous variable, an indefinitely large number of values may occur. Height, weight, time etc are examples of continuous variables. Depending on the type of variables involved, data may also be classified as discrete or continuous.

**Know your progress**

1. Write examples for variables.
2. Write examples for discrete and continuous variables.
3. Give examples other than those presented in this section of a qualitative variable, discrete quantitative and continuous quantitative variable.

**Levels of Measurement - Nominal, Ordinal and Cardinal Data**

S.S.Stevens (1906 -1973) described the data into different scales of measurement as nominal, ordinal and cardinal data. This classification is based on the data under consideration. A nominal scale of measurement is used to name categories such as gender, nation, etc. For example the categorization like male, female is a nominal data. While filling Higher Secondary single window application form for admission we give different codes to represent data such as Thiruvanathapuram 01, Kasaragod 14 etc. This is also a nominal data. Here the number is merely a label, does not have a quantitative significance. The only effect of such a measurement is that we can count how many objects fall in each category like 10000 belongs to Thiruvanathapuram.
9700 belongs to Kasaragod, etc. In the ordinal scale of measurement, we can put an order to the data according to the relation among the values of the variables. While considering the educational qualification of a group of people, we can categorise them as Secondary, Higher Secondary, Graduate, Postgraduate, etc. Here we can rank Secondary -1, Higher Secondary -2, Graduate -3, Postgraduate -4, etc. Here code 3 is surely higher than code 1 as a graduate is at a higher level than a secondary student based on educational qualifications. That is in ordinal scale, there is a specific order or rank for the codes given to each category. The data regarding a quantitative variable is a cardinal data. The height of students in a class, monthly salary of school teachers, marks of students in a class etc. are examples of cardinal data.

Know your progress

1. Compare Nominal, Ordinal, Cardinal data.
2. Give some examples Nominal, Ordinal and Cardinal data.

2.3 Types of Data

Primary and Secondary Data

Data means the raw facts and figures that have been collected. Data can be gathered by looking through existing sources, conducting experiments or by conducting surveys. Based on these sources of collections, statistical data may be classified as primary or secondary. Primary data are those which have to be collected by the investigator for the first time for his/her own purpose. It is fresh in nature. In the words of Wessel data originally collected in the process of investigation are known as primary data. It is collected by using appropriate survey techniques. Data obtained from existing sources which may be published or unpublished are known as secondary data. Secondary data may not be in the required form. These data are obtained by other persons and are being used now at second hand. According to M. M. Blair, “secondary data
those which are already in existence, and which have been collected for some other purpose than the answering of question in hand”

Comparison between Primary and Secondary data

<table>
<thead>
<tr>
<th>Primary Data</th>
<th>Secondary Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is original in nature</td>
<td>It is not original in nature</td>
</tr>
<tr>
<td>It is in the form of raw materials</td>
<td>It is in the form of finished products</td>
</tr>
<tr>
<td>Collection involves more money and time</td>
<td>Less time and money are needed</td>
</tr>
<tr>
<td>Trained persons are required for data collection</td>
<td>The investigator should be vigilant while collecting secondary data</td>
</tr>
<tr>
<td>Primary data, after use become secondary data</td>
<td>Secondary data cannot be converted to primary data</td>
</tr>
</tbody>
</table>

2.4 Questionnaire and Schedule

Questionnaires and schedules are series of questions arranged in a logical order so as to collect information for a specified purpose. The purpose may be single or multiple.

Questionnaire

A questionnaire is usually mailed by post or by email to selected informants. The informants are allowed a specified time to fill up the questionnaire and have to return to the investigator. Here the quality of the obtained data depends on the quality of the questions and the honesty of the informants. As the informants are to fill up the data, they should be literate. This method is suitable in cases where the informants are widely scattered. One of the main disadvantages of this method is that the chance of getting incomplete information is large.
Schedule

If the group of informants are not widely scattered, or if they are not literate, the enumerator himself/herself can personally approach the informants with the set of question and collect information. These questions may not be in detailed manner as questionnaire. It may not contain explanatory foot notes or explanations of terms used. These set of questions used for data collection is termed as schedule. In some cases questionnaire itself can be used as a schedule.

Comparison between Questionnaire and Schedule

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is often sent by post</td>
<td>Enumerators carry the schedule personally to the informant</td>
</tr>
<tr>
<td>Answers are filled by the respondents</td>
<td>Answers are filled by the enumerators</td>
</tr>
<tr>
<td>Informants are to be literate</td>
<td>Informants need not be literate</td>
</tr>
<tr>
<td>Success depends on the quality of questions and sincerity of the informant</td>
<td>Success depends on the honesty and competence of the enumerator</td>
</tr>
<tr>
<td>Chance of getting incomplete information is more</td>
<td>Chance of getting incomplete information is less as enumerators explain the questions</td>
</tr>
</tbody>
</table>

Requisites of good questionnaire

While preparing a questionnaire we have to keep in mind the following points.

1. Questions should be capable of generating all required information

2. The language and wording of the questions should be convenient to the informant
3. Question should not contain technical terms and words with uncommon meaning, such questions leads to different information from different informants

4. Yes or No questions or multiple answer choice questions should be preferred.

5. Personal questions should be avoided

6. Necessary foot notes should be provided

7. Usually the number of questions should be 20 to 25

8. Questions should have a logical order

9. Questions should be self explanatory

10. Questionnaire should be attractive so as to impress the informant

11. Questions should be unambiguous

Once the questionnaire is ready, it is advisable to conduct a pre-test with the questionnaire for a small group. This is known as pilot survey. It helps the investigator to measure the worthiness and reliability of these questions.

**Drafting of Questionnaire**

A sample questionnaire is prepared below for studying the socio economic status of people residing in a village.
QUESTIONNAIRE TO COLLECT DATA ABOUT THE SOCIO-ECONOMIC STATUS OF PEOPLE

1. Name:
2. Address:
3. Members in the family: No. of Males □ No. of Females □
4. Age of family members: (write number of members in each category in Corresponding boxes)
   a) below 10 □  b) 10-20 □  c) 20-50 □  d) 50 and above □
5. Residing in:
   a) Own house □  b) Rental □
6. Type of House:
   a) Temporary □  b) Structured □
7. Toilet facility:
   a) Proper □  b) Improper □
8. Water Facility:
   a) own well □  b) water provided by panchayat □  c) other Sources □
9. Electrified Home:
   Yes □  No □
10. Details of electronic items in house: (write number of items in each boxes)
    a) Bulbs □  b) Refrigerator □  c) Fan □  d) Television □
    e) Tubes □  f) Mixer grinder □  g) Others □
11. Mode of cooking:
    a) Wood □  b) Kerosene □  c) Gas □  d) Electricity □
12 Occupation:
   a) Govt. service □ b) Non Govt. Service □ c) Own business □
   d) Agriculture □ e) Others □

13 Monthly income of family (in rupees):
   a) Below 8000 □ b) 8000-20000 □
   c) 20000-50000 □ d) above 50000 □

14 The amount you monthly spend for education of children:
   a) below 500 □ b) 500-2000 □
   c) 2000-4000 □ d) above 4000 □

15 Do you own a private vehicle?:
   Yes □ No □

16 If Yes, give the number of vehicles in each category:
   a) Two wheeler □ b) Three wheeler □
   c) Car □ d) Others □

17 The approximate monthly expenditure:
   a) Below 5000 □ b) 5000-15000 □
   c) 15000-20000 □ d) above 20000 □

18 Would you prefer an outing occasionally with your family?
   Yes □ No □

19 If Yes, what will be your budget for a single trip?
   a) below 2000 □ b) 2000-5000 □ c) above 5000 □

20 Are you able to properly maintain your standard of living with your actual income?
   Yes □ No □

21 Any other information regarding your family (Give details in one or two sentences)
   (** This data will be used for study purpose only)
2.5 Methods of Primary Data Collection

Direct Personal Interview

If the field of investigation is small, the investigator or enumerator can access all the informants personally and conduct spot enquiry. This method is called direct personal interview. The success of this method depends on the efficiency of the enumerator. The enumerator should be tactful to get all the required information. In this method the enumerator can collect all the supplementary information required for interpretation of data.

Indirect Oral Investigation

Consider a situation in which the investigator wants to collect data about a resident in a city. In this case, the investigator may approach a third party, called witness, who is capable of giving sufficient information about the resident. This is a case of indirect oral investigation. Indirect oral investigation is applicable in cases where the informant is reluctant to give information or when the informant is not available. The disadvantage of this method is that the reliability of the information heavily depends on the quality and honesty of the witness or intermediate person.

Direct Observation

This method is widely used by mass media for collecting information or by journalists. Consider a situation in which the investigator wants to report the current situation in an area due to heavy rain and flood. He does not have a predetermined set of questions to collect data. The investigator collects data from what he observes. This may not be in a well defined manner. The investigator must be well equipped so as to collect maximum information from the place. The quality of the information depends mainly on the honesty of the investigator to report it to the maximum extend.
Telephone interview

In some cases the informant may be reluctant to give answer in a face to face personal interview. In such cases it is better to select another method for data collection. Telephone interview is one such method. In this case the investigator collects data from the informant indirectly but personally. This is less time consuming and cheaper than direct personal interview. The disadvantage is that it will not worked in some rural areas were telephonic connection is very low.

Mailed Questionnaires and Schedules

Questionnaires and schedules are one of the most popular methods for collecting primary data. As the questionnaires are usually mailed to the respondents, it is known as mailed questionnaire method. The only difference between the questionnaire and schedule is that in questionnaires the answers are filled by the respondents themselves but in schedule, the answers are filled by the enumerators.
Focus Group Discussion

A Focus Group Discussion (FGD) is a small group discussion guided by a trained leader. It is used to collect more opinions about a specified topic in order to take better decisions in future plans. For example, Mrs. Rema wants to start a preschool in an area. Her aim is to provide childcare as much as possible. For that she invites the parents who have children under four years from that locality and arranges Focus Group Discussion. The parents have their own ideas about childcare and other locally adopted programmes. Their suggestions will help Mrs. Rema to start the institution in a well equipped manner. Here Mrs. Rema plays the role of the investigator. She collects information through focus group discussion for her specified purpose.

2.6 Sources of Secondary Data

Any published or unpublished data which are reliable for the current situation is a source of secondary data. While collecting secondary data the investigator must be aware of the following points.

- The geographical area of the collected data.
- The time at which the data was collected.
- The terms and definitions involved in the data.
- The person who collected the data and the purpose for which they are collected.

Some sources of secondary data are listed below

- Government publications.
- Office records in panchayats, municipalities etc.
- Survey reports of various research organizations.
- Survey reports in Journals, Newspapers and other publications.
- Websites.
Let us sum up

For studying statistics, the first step is collection of data. It is systematic gathering of informations. Statistical surveys are tools of data collection. Data means information regarding a variable. The variable may be qualitative and quantitative or it may be discrete or continuous. The two survey techniques are Census and sampling. Depending on the source of information, data can be classified as primary or secondary. The important primary data collection methods are direct personal investigation, indirect oral investigation, direct observation, telephone interview, mailed questionnaire or schedule sent through enumerators and focus group discussion. Any published or unpublished data which are collected by a third party, and now used by the investigator for his purpose is a secondary data. Better trained persons are required for converting the secondary data to the required form.

Learning outcomes

After transaction of this unit, the learner:-

- differentiates population and sample.
- recognises investigator, investigation, enumerator and enumeration.
- classifies variables and constants.
- distinguishes qualitative variables and quantitative variables.
- differentiates discrete and continuous variables.
- compares primary and secondary data.
- identifies questionnaire and schedule.
- constructs/drafts questionnaire.
- explains different methods of data collection.
- recognises the sources of secondary data.

Evaluation Items

1. Data that can be classified according to colour. They are measured on ........ scale
   a) Nominal  b) Ordinal  c) Cardinal
2. The group of all subjects under study is called ………

3. The representative part of a population is called ………

4. The number of days of absence of a worker per year is ……… type of data
   a) Nominal   b) Qualitative   c) Discrete   d) Continuous

5. The pre-test with the questionnaire before conducting a survey is called ………

6. The blind population of India constitute:
   a) a hypothetical population   b) a sample   c) an infinite population   d) a finite population

7. Which of the following represents data
   a) a single value   b) only two values in a set   c) a group of values in a set   d) all the above

8. Statistics deal with
   a) qualitative information   b) quantitative information   c) both (a) and (b)   d) none of (a) and (b)

9. Data taken from a publication Agricultural situation in India will be considered as
   a) primary data   b) secondary data

10. Mailed questionnaire method of enquiry can be adopted if respondents
    a) live in cities   b) have a high income   c) are literate

11. A study based on complete enumeration is known as:
    a) sample survey   b) pilot survey   c) none of the above

12. Statistical data are collected for.
    a) no purpose   b) a given purpose   c) any purpose

13. A statistical population may consists of
    a) an infinite number of items   b) a finite number of items   c) either of (a) and (b)   d) none of (a) and (b)
14. Compare Primary and Secondary data

15. Distinguish between a questionnaire and a schedule.

16. What kind of data you receive when you are told about
   (a) blood type b) household c) heights of waterfall

17. What are points to be remembered while drafting a questionnaire?

18. What are the points to be remembered while collecting secondary data?

19. Explain the various primary data collection methods.

20. What are the important sources of secondary data?

21. Find the discrete data and continuous data from the following list
   a. Number of shares sold each day in a stock market
   b. Temperature recorded every half an hour at a weather bureau
   c. Lifetime of television tubes reduced by a company
   d. Yearly income of employees in a company
   e. The age of an individual
   f. Number of petals a flower has.

22. Categorize the data obtained in the following situations as quantitative
    and qualitative
   a. Political preference of a group of people
   b. Family size (Number of members of a family) of hundred families in a
      township
   c. IQ score of plus one students undergoing state syllabus in Thiruvananthapuram
      district
   d. Academic qualification of a group of unemployed youth in a city

23. Which of the following constitute finite or infinite population
   a. Population consisting of odd integers.
   b. Weight of 200 new born babies in a hospital.
   c. Height of fifteen year old children in a school.
   d. Number of head and tail when a coin is tossed.
24. Categorize the following under cardinal, nominal or ordinal
   a. Telephone Numbers
   b. Roll numbers given to students in a class
   c. Ranks given to a class after a test
   d. Respondents attitude towards a newly designed project in an institution on a five point scale such as 1=strongly opposed, 2= may be opposed, 3= not strongly favoured, 4= may be favoured and 5= very strongly favoured
   e. The quantity of water in a reservoir measured in every half an hour.
   f. The price of furniture exhibited in a shop.

25. It is proposed to conduct a survey to obtain information on the study habits of Higher Secondary students in Kannur district and also the facilities available to them. Prepare a questionnaire for this purpose

26. A survey is to be carried out amongst school children to study how they spent time after school hours. Prepare a questionnaire for that purpose

27. Indicate whether the following statements are true or false. If false correct the statements
   a. Secondary data are generally used in those cases where the primary data do not provide an adequate basis for analysis.
   b. Secondary data does not need much scrutiny and should be accepted at its face value
   c. The task of editing secondary data is a highly specialised one
   d. The questionnaire requires a pre-testing before putting into practice

28. Which type of study do you prefer in the following cases? (Census or Sampling). Give reason
   a. The effect of a medicine
   b. To study about the wage distribution of 250 employees in a company
   c. A study on the roll of media in the marketing of a face cream
   d. A study of a patients heart beat who is admitted to ICCU of a hospital
   e. A study on the number of petals of a flower of a special kind
f. The life span of an electric bulb.

29. Which of the primary data collection method do you suggest in the following situations?. Give reasons
   a. You are appointed as marketing manager of a company. The company introduces a washing machine with many options. You are asked by your employer to prepare a datasheet regarding the opinion of your customers about the new equipment
   b. To prepare a report for a media on Nehru Trophy Vallamkali in the current year.
   c. To introduce shift sessions in an institution

30. As a reporter of a certain media, you got an opportunity to interview an IAS topper. Which type of data collection method will you use?. List out other primary data collection methods.