

Unit 2**Statistical Survey****Structure:**

- 2.1 Introduction
 - Objectives
 - Relevance
 - Statistics in practice
 - Definition of statistical survey
- 2.2 Stages of Statistical Survey
 - Planning a statistical survey
 - Execution of statistical survey
- 2.3 Basic Terms used in Statistical Survey
 - Units or individuals
 - Population or universe
 - Sample
 - Quantitative characteristic
 - Qualitative characteristic
 - Variable
- 2.4 Collection of Data
 - Primary data
 - Secondary data
 - Pilot survey
- 2.5 Measurement Scales
 - Qualitative (categorical) data
 - Quantitative (numerical) data
- 2.6 Scrutiny and Editing of Data
- 2.7 Summary
- 2.8 Glossary
- 2.9 Terminal Questions
- 2.10 Answers
- 2.11 Case Study

2.1 Introduction

In the previous unit, 'Introduction to Statistics', we have been introduced to the definition and functions of statistics. We also studied the broad divisions of statistics. We now have an idea about the characteristics of statistics and

the limitations of statistics. In this unit, we will study about statistical survey and the collection and analysis of numerical data.

When the population is large, it is hard to conduct a survey. In such situations, a sample is drawn and studied to determine the characteristics of the entire population. The primary purpose of conducting a sample survey is to obtain certain information about the population.

We define the term 'survey' as a measurement procedure to gather people's opinions. Surveys differ from each other as their purpose, field of study, scope, and the source of information differ. Surveys are used by companies to assess the level of their customer satisfaction, to find out what products their customers choose and to determine which section of the population is buying their products. The following are some examples of activities, which require collection and analysis of data in a systematic manner.

- Formulation of a theory such as "Tobacco Consumption Leads to Cancer"
- Framing of policies according to the existing nature of a population
- Finding the relationship between characteristics of units in the population

In other words, a search for knowledge by analysing numerical data is known as Statistical Survey or Statistical Investigation.

Objectives:

After studying this unit, you should be able to:

- recall the definition of statistical survey
- describe the activities involved in planning a statistical survey
- recall the definition of terms used in statistics
- differentiate between sample and population
- differentiate between quantitative and qualitative characteristics
- describe various methods of data collection
- distinguish between primary and secondary data
- explain various measurement scales

2.1.1 Relevance

The relevance, timeliness and accuracy of data are the standard tools for any statistical study. The quality of information and conclusion derived from a data depends on these characteristics. The absence of these reflects in

the popular way “Garbage in, garbage out” abbreviated as GIGO, mostly used in the field of computer science. Therefore, it is equally important in the context of statistical data and utmost care has to be taken while collecting the right data by the right process and from the right source.

2.1.2 Statistics in practice

Recent CASs (Country Assistance Strategy) for Kenya and Armenia provide good examples of assessing statistical capacity and proposing appropriate action. The Kenya CAS takes a comprehensive approach towards statistical capacity building, based on the implementation of a national statistical development strategy supported by IDA (International Development Association) and a number of other development partners. The CAS for Armenia finds that, Armenia’s capacity for poverty monitoring and analysis is reasonably good, as the National Statistical Service (NSS) has conducted regular household surveys for a number of years. The CAS identifies steps to further improve capacity, including ‘strengthening the linkages between different household surveys’ and ‘improving questionnaires to reflect current policies (for example, on social assistance) and to provide better information (for example, on employment and earnings).’

(Source: <http://siteresources.worldbank.org>)

2.1.3 Definition of Statistical Survey

A Statistical Survey is a scientific process of collection and analysis of numerical data. Statistical surveys are used to collect information about units in a population and it involves asking questions to individuals. Surveys of human populations are common in government, health, social science and marketing sectors.

2.2 Stages of Statistical Survey

Statistical surveys involve two stages namely – Planning and Execution. Figure 2.1 shows the two broad stages of Statistical Survey.

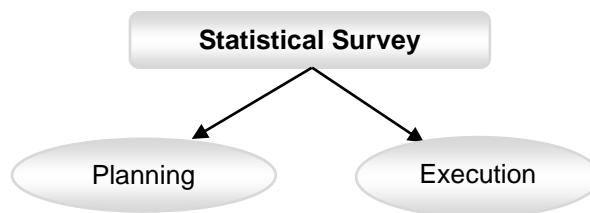


Fig. 2.1: Stages of Statistical Survey

2.2.1 Planning a Statistical Survey

The relevance and accuracy of data obtained in a survey depends upon the care taken in planning. A properly planned investigation can lead to the best results with least cost and time. Figure 2.2 gives the explanation of steps involved in the planning stage.

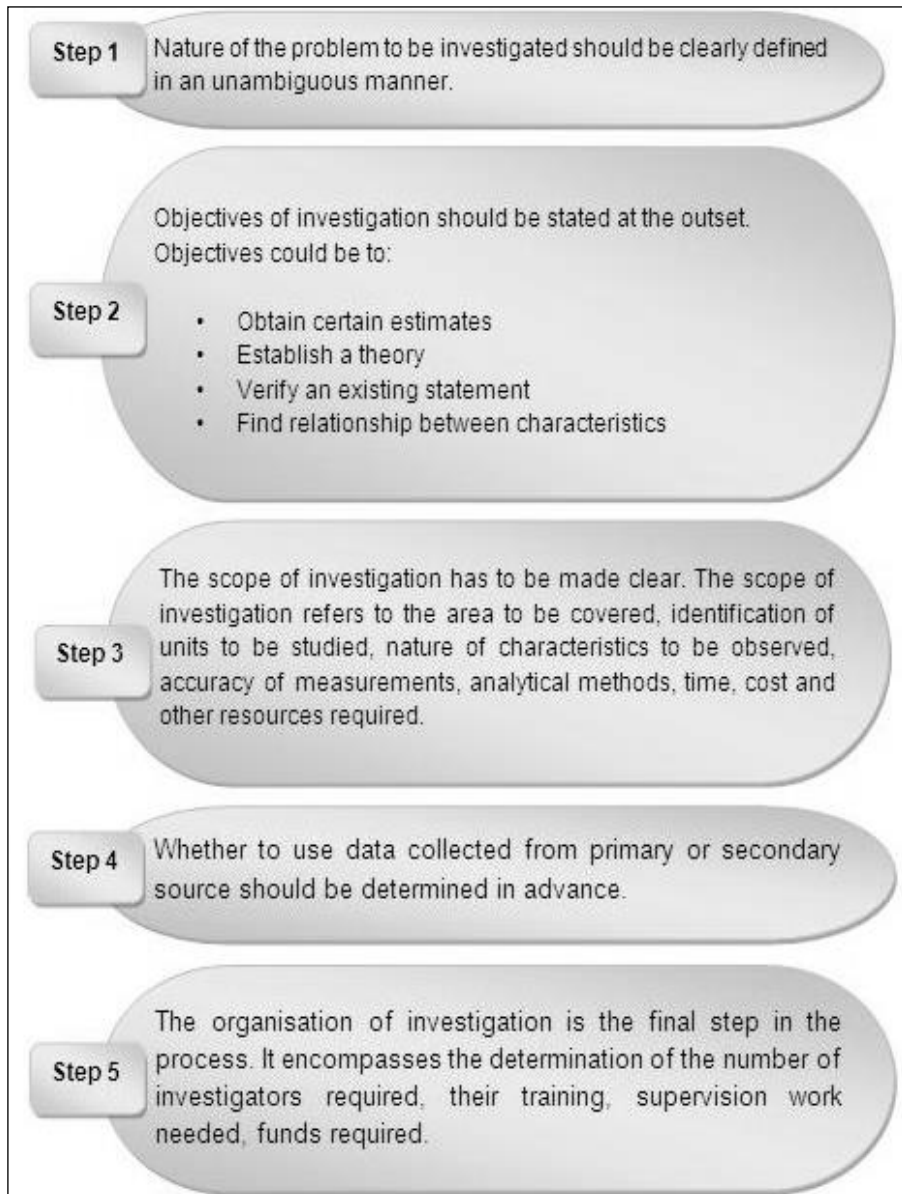


Fig. 2.2: Steps Involved in Planning of a Statistical Survey

2.2.2 Execution of statistical survey

Controlled methods should be adopted at every stage of carrying out the investigation to check the accuracy, coverage, methods of measurements, analysis and interpretation.

The collected data should be edited, classified, tabulated and presented in the form of diagrams and graphs. The data should be carefully and systematically analysed and interpreted.

Self Assessment Questions

1. What are the main stages in a survey?
2. Training of investigators belongs to which stage?
3. Analysis of data is a part of the execution of survey. Is this correct?

2.3 Basic Terms Used in Statistical Survey

Statistics, being a specialised subject, has a number of terms which have to be used. Knowledge and understanding of these terms is necessary to do any statistical work. Stated here are some of the basic terms used in Statistical Survey.

2.3.1 Units or Individuals

In a statistical survey, the objects on which the characteristics are measured are referred to as Units or Individuals.

2.3.2 Population or Universe

The totality of all units or individuals in a survey is called Population or Universe. If the number of objects in a population is finite, it is called finite population otherwise, it is known as infinite population.

The measure describing the characteristics of the population is known as parameter. In figure 2.3, the total number of eight consumers constitutes the entire population.

Key Statistics

A parameter is a measure of the characteristic of the population. Population can have many parameters.

A Statistic is a measure of characteristic corresponding to the sample. Sample can have many statistics.

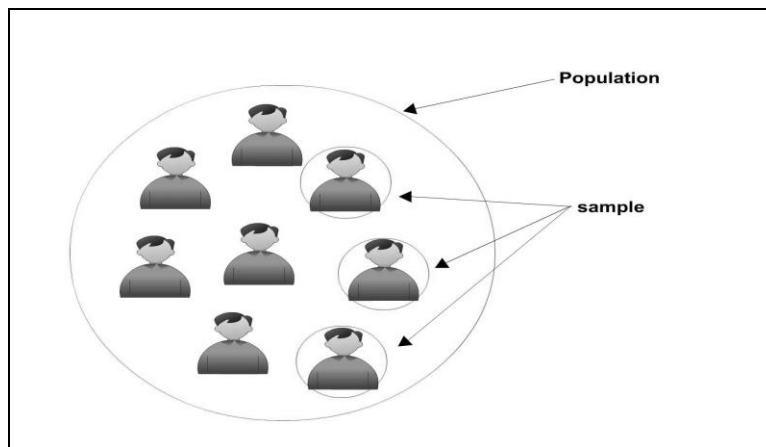


Fig. 2.3: Population versus Sample

2.3.3 Sample

A sample is a part or a subset of the population. By studying the sample, you can predict/comment on the characteristics of the entire population from where the sample is taken. The measure that describes the characteristics of a sample is known as statistics.

If the population is large, it is hard to collect data corresponding to the entire population. Hence, a part of the population is chosen to study the characteristics of the entire population. The size of the sample can never be as large as the size of the population. Proper care must be taken while choosing the samples. In the figure 2.3, a sample of three consumers is drawn from the entire population of eight consumers.

2.3.4 Quantitative Characteristic

A characteristic which is numerically measurable is called a Quantitative Characteristic. Quantitative data is data expressing a certain quantity, amount or range. Usually, there is measurement units associated with the data, for example, the height of a person in centimetres.

2.3.5 Qualitative Characteristic

A characteristic which is not numerically measurable is called a Qualitative Characteristic. Qualitative data is data describing the attributes or properties that an object possesses.

Let us understand the basic terminologies of statistical survey with the help of a Caselet.

Caselet

Consider the survey of the average number of children below 16 years in a ward of a municipality. The number of houses in the ward is finite and therefore, the population is finite. The objects are households. The characteristic measured is number of children below 16 years in a household. It is numerically measurable and hence quantitative. On the other hand, in a survey to find the total number of blind people in a locality, the characteristic 'blindness' is qualitative.

2.3.6 Variable

In a population, some characteristics remain the same for all units and some others vary from unit to unit. The quantitative characteristic that varies from unit to unit is called a variable. It is a measurable characteristic for example, age, height, income. The qualitative characteristic that varies from unit to unit is called an attribute. It is a non-measurable characteristic for example, religion, nationality and occupation.

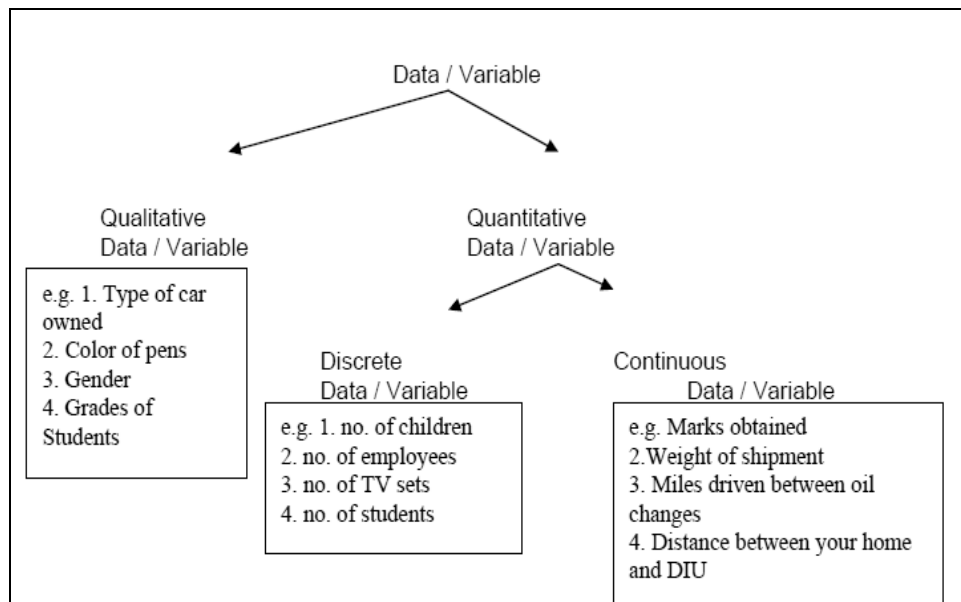


Fig. 2.4: Various Types of Variables

A variable that assumes only some specified values in a given range is known as 'discrete variable'. A variable that assumes all the values in the range is known as 'continuous variable'. For example, the number of children per family and the number of petals in a flower are examples of discrete variables. The height and weight of a person are examples of continuous variables. Figure 2.4 shows the difference between Qualitative and Quantitative Variable and also the difference between Discrete and Continuous Variable.

Self Assessment Questions

4. Classify the following as finite or infinite population.
 - i) Production of a product in a factory for a day
 - ii) The set of rational numbers
 - iii) The weight of new born babies measured up to first decimal place in a state during the first week of February 2008
5. Classify the following as an attribute or a variable.
 - i) Eye colour of human beings
 - ii) Number of pages in a book of various subjects
6. Classify the following as discrete or continuous variable
 - i) Number of shares sold each day in a stock market.
 - ii) Temperatures recorded every half hour at a regional meteorological centre.
7. Statistics can best be considered as
 - i) both Art and Science
 - ii) Art
 - iii) Science
 - iv) neither Art nor Science
8. Data that possess numerical properties are known as
 - i) Quantitative data
 - ii) Qualitative data
 - iii) Primary data
 - iv) Parametric data

9. A tool of all science in research and making an intelligent judgement is
- i) Statistics
 - ii) Collection
 - iii) Data
 - iv) Judgement

2.4 Collection of Data

Collection of data is the first and most important stage in any statistical survey. The method for collection of data depends upon various factors such as objective, scope, nature of investigation and availability of resources. Direct personal interviews, third party agencies, and questionnaires are some ways through which data is collected.

2.4.1 Primary data

Primary data is the one, which is collected by the investigator for the purpose of a specific inquiry or study. Such data is original in character and is generated by a survey conducted by individuals or a research institution or any organisation.

For Example:

If a researcher is interested to know the impact of a non-meal scheme for school children, he/she has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected is called primary data.

Data collected for the first time keeping in view the objective of the survey is known as primary data. Interview, questionnaire and telephone/mail are all examples of primary data.

They are likely to be more reliable. However, cost of collection of such data is much higher. Primary data is collected by either a census method or a sampling method.

Key statistic

A census is the procedure of systematically acquiring and recording information about the members of a given population

Collection of primary data is done by a suitable method as per the following:

1. Direct personal observation
2. Indirect oral interview
3. Information through agencies
4. Information through mailed questionnaires
5. Information through a schedule filled by investigators

Let us study each of them in detail which are as follows:

1. Direct Personal Observation – In the direct personal observation method, as illustrated in figure 2.5, the investigator collects data by having direct contact with the units of investigation. The accuracy of data depends upon the ability, training and attitude of the investigator.



Fig. 2.5: Direct personal observation

The direct personal observation method is suitable where,

- The scope of investigation is narrow
- Investigation is confidential and requires personal attention of the investigator
- Accuracy of data is important

Table 2.1 shows the merits and demerits of a direct personal observation method.

Table 2.1: Merits and Demerits of a Direct Personal Observation Method

Merits	Demerits
1. We get the original data which is more accurate and reliable.	1. This method is not cost efficient.
2. Satisfactory information can be extracted by the investigator through indirect questions.	2. This method consumes more time.

3. Data is homogeneous and comparable.	3. This method cannot be used when the scope of investigation is wide.
4. Additional information can be gathered.	4. Most of the data collected through this method is confidential.

2. Indirect oral interview – Indirect oral interview is used when the area to be covered is large. The investigator collects the data from a third party or a witness or the head of an institution. This method is generally used by the police department in cases related to enquiries on the cause of fires, thefts or murders.

In this method, the investigator contacts witnesses or neighbors or friends or some other third parties who are capable of supplying the necessary information. Enquiry committees appointed by governments use this method to get people's views and every possible detail regarding the enquiry. This method suits best when direct sources do not exist or cannot be relied upon or would be unwilling to take part in the survey. Table 2.2 shows the merits and demerits of indirect oral interview.

Table 2.2: Merits and Demerits of Indirect Oral Interview Method

Merits	Demerits
1. Economical in terms of time, cost and man power	1. The degree of accuracy of the information is less.
2. Confidential information can be collected	
3. Information is likely to be unbiased and reliable	

3. Collecting information through agencies – Methods of collecting information through local agencies or correspondents is generally adopted by newspapers and television channels. Local agents are appointed in different parts of the area under investigation. This method is illustrated in figure 2.6. They send the desired information at regular intervals.

This method is used where the area to be covered is very large and periodic information is required. However, one disadvantage of this method is that the information is likely to be biased.

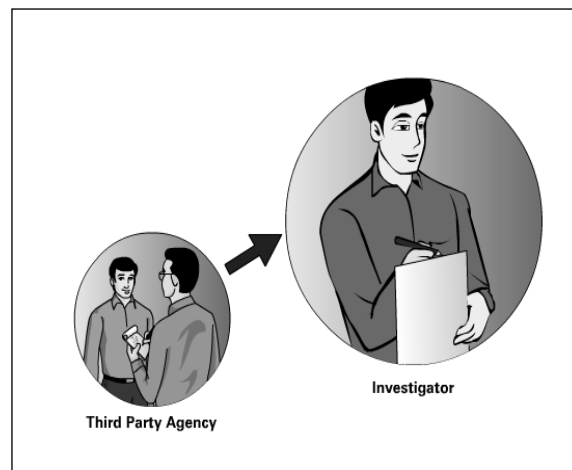


Fig. 2.6: Collecting Information through Agencies

4. Information collected through mailed questionnaires – Often, information is collected through questionnaires. The questionnaires are filled with questions pertaining to the investigation. They are sent to the respondents with a covering letter soliciting cooperation from the respondents (respondents are the people who respond to questions in the questionnaire). The respondents are asked to give correct information and to mail the questionnaire back. The objectives of the investigation are explained in the covering letter along with the assurance to keep the information confidential.

Questionnaire design

- **Initial considerations**

- Type of information required

- Type/nature of respondents

- Type and method by which survey is to be undertaken

- **Question content**

- Relevance of a question

- Clarity of a question

- Avoid ambiguous, leading, double-barrelled questions

- Ability and willingness of a respondent to answer the questions

- **Question phrasing**
 - Style appropriate to target population
 - Short, Clear and unambiguous questions
 - Avoid biased words and leading questions
 - Avoid negative questions
 - Discourage guessing
 - Do not assume anything for granted from the part of the respondents
- **Types of questions**
 - Closed ended questions**
 - Dichotomous
 - Multiple choice (4 to 5 options; neutral point)
 - Likert scale (Agree or disagree)
 - Semantic differential (scale connecting bipolar words)
 - Importance scale (importance of some attribute)
 - Rating scale (Excellent to poor)
 - Open ended questions**
 - Completely unstructured
 - Word association (first word that comes to mind ...)
 - Sentence completion
 - Story completion
 - Picture completion (filling balloons)
 - Thematic Apperception Test (relate story to picture)
- **Question sequence**
 - Logical order
 - Avoid questions which suggest answers to later questions (bias)
- **Questionnaire layout**
 - Good quality paper
 - As short as possible (20-30 questions)
 - Use lines, boxes, pictures, etc.
 - Instructions kept to a minimum but user-friendly
 - Purpose of survey explained at the beginning and guarantee of confidentiality
 - What is to be done with the completed questionnaire?

- **Pre-test, revision and final version of questionnaire**

- Uncover faults

- Misprints

- Grammatical mistakes

- Relevance of questions

- Expected range of answers

Construction of a good questionnaire is an important contributing factor to the success of a survey. When questionnaires are properly framed and constructed, they become important tools by which statements can be made about specific people or entire population.

This method is generally adopted by researchers and other official and non-official agencies. This method is used to cover large areas of investigation. It is more economical and free from investigator's bias. However, it results in many "non-response" situations. The respondent may be illiterate. The respondent may also provide wrong information due to wrong interpretation of questions.

If the questionnaire consists of invalid questions, or questions in incorrect order, or questions in inappropriate format, or questions that are biased, then the survey would be useless. An important method for checking and making sure whether a questionnaire is accurately capturing the intended information, is to pre-test it among a smaller subset of target respondents.

Success of this method of collection of data depends mainly on proper drafting of the questionnaire. You have to keep the following points in mind while preparing a questionnaire:

- The respondent should not take much time in completing the questionnaire. It should be small and not lengthy.
- The questions asked should be well structured and unambiguous.
- The questions asked should be in a proper logical sequence.
- Questions should be unbiased. The questions in the questionnaire should not disturb the privacy of the respondents.
- The questionnaire should not have much writing work.
- Necessary instructions and glossary should be given in covering letter.
- Questions involving technological jargons and mathematical calculations should be avoided.

- The completed questionnaire should be kept confidential and used only for the purpose of the survey as mentioned in the investigation.
- There should not be any scope for misinterpretation in the questions.

There are different types of questions that can be used in the questionnaire. A questionnaire can have contingency questions, matrix questions, close ended questions and open ended questions. Let's have a look at each one in detail.

- **Contingency questions** are questions that are answered only if the respondent gives a particular response to a previous question. This avoids asking people questions that do not apply to them.
- **Matrix questions** are questions which are placed one under the other, forming a matrix. The response categories are placed on the top and a list of questions are placed by the side. This is used to efficiently occupy space and respondents' time.
- **Closed ended questions** are those where the respondents' answers are limited to a fixed set of responses. Usually scales are closed ended.

There are various types of closed ended questions.

Dichotomous Question: A question that has two possible responses. The responses could be Yes/no, True/False, Agree/Disagree.

For example:

Example 1

- | | |
|-------------------------------------|--|
| • Are you a science graduate? | Yes [<input type="checkbox"/>] No [<input type="checkbox"/>] |
| • Did you watch a movie last night? | Yes [<input type="checkbox"/>] No [<input type="checkbox"/>] |

Multiple choices – here the respondents have several options from which to choose. For example:

Example 2

The sun rises in which direction?

- | | |
|------------------------------------|------------------------------------|
| East [<input type="checkbox"/>] | West [<input type="checkbox"/>] |
| North [<input type="checkbox"/>] | South [<input type="checkbox"/>] |

Scaled questions: here the responses are graded on a continuum (For example, rating the appearance of a product on a scale from 1 to 10, with 10 implying the most preferred appearance and 1 implying the least preferred appearance). Scaled questions are mostly questions related to attitudes. A Likert scale provides a number of attitude statements. The respondent has to say how much they agree or disagree with each one.

Example 3

Read the following statement and then indicate by a tick whether you Strongly Agree, Agree, Disagree or Strongly Disagree with the statement.

“Organised and prioritised tasks take less time to complete.”

- | | |
|----------------------|-----|
| 1. Strongly Agree | [] |
| 2. Agree | [] |
| 3. Disagree | [] |
| 4. Strongly Disagree | [] |

- **Open ended questions** are those questions for which the respondent provides their own answer without any fixed set of possible responses. Examples of the types of open ended questions are:

Sentence completion: In these, respondents complete an incomplete sentence.

Example 4

Complete the sentence below.

“I like the management courses offered by Manipal University Jaipur because ...”

Story completion: In these, respondents complete an incomplete story.

Picture completion: In these, respondents fill in an empty conversation balloon.

Thematic Apperception Test: In these, respondents explain a picture or make up a story about what they think is happening in the picture.

Activity

Design a questionnaire for consumer response in Facebook Vs Twitter in the Internet.

5. Information through schedule filled by investigator – Information can be collected through schedules filled by investigators through personal contact. In order to get reliable information, the investigator should be well trained, tactful, unbiased and hard working.

A schedule is suitable for an extensive area of investigation through investigator's personal contact. The problem of non-response is minimised.

There is a difference between a schedule and a questionnaire. A schedule is a form that the investigator fills personally, while surveying the units or individuals from the sample (respondent). A questionnaire is a form sent (usually mailed) by an investigator to respondents. The respondent has to fill it and then send it back to the investigator.

Table 2.3: Advantages and Disadvantages of Information through schedule filled by investigator method

Advantages	Disadvantages
1. The cost per respondent is likely to be less, i.e, more people can be sampled	1. Low response rate.
2. More questions can be asked since the respondent is answering them at his/her own convenience.	2. Questionnaires may only be partly answered.
3. The interviewer cannot influence any respondent (all questionnaires are the same).	3. Misunderstandings may not be clarified.
4. Questions on sensitive issues may be asked and answered easily.	4. There is no encouragement to think more deeply on the questions before answering them.
6. A large amount of response is obtained in a short period of time.	

Example of a customer opinion questionnaire used by a local restaurant in Bangalore:

Customer opinion questionnaire used by a local restaurant in Bangalore:				
We are happy you stopped by the local restaurant and want to make sure you will come back. So, if you have a little time, we will really appreciate it if you could fill out this form. Your comments and suggestions are extremely important to us. Thank you				
Server Name _____				
	Excellent	Good	Satisfactory	Unsatisfactory
Food Quality				
Friendly Service				
Prompt Service				
Cleanliness				
Management				
Comments _____				

What prompted your visit here? _____				

Please drop it in the suggestion box at the entrance. Thank you				

2.4.2 Secondary data

Any information, that is used for the current investigation but is obtained from some data, which has been collected and used by some other agency or person in a separate investigation, or survey, is known as secondary data. They are available in a published or unpublished form.

In published form, secondary data is available in research papers, newspapers, magazines, government publication, international publication, and websites. Secondary data is collected for different purposes. Therefore, care should be exercised while using it.

The accuracy, reliability, objectives and scope of secondary data should be examined thoroughly before use. Secondary data may be collected either by census or by sampling methods.

Published sources

The various sources of published data are:

- Reports and official publications of international and national organisations as well as central and state governments
- Publications of several local bodies such as municipal corporations and district boards
- Financial and economic journals
- Annual reports of various companies
- Publications brought out by research agencies and research scholars

Some of the journals (both academic and non-academic) are published at regular intervals like yearly, monthly, weekly whereas, other publications are more ad hoc. Internet is a powerful source of secondary data, which can be accessed at any time for any further analysis of the study.

Unpublished sources

It is not necessary that all statistical contents have to be published. Unpublished data such as records maintained by various government and private offices, studies made by research institutions and scholars can also be used where necessary.

Though, use of secondary data is economic in terms of expense, time and manpower requirement, researcher must be careful in choosing such secondary data. Secondary data must possess the following characteristic:

- 1) Reliability of data: The reliability related to secondary data can be tested by investigating
 - a) Who collected the data?
 - b) What were the sources of data?
 - c) Whether they are collected by a proper method?
 - d) At what time were they collected?
 - f) What level of accuracy was desired? Was it achieved?
- 2) Suitability of the data: The data that is suitable for one enquiry may not necessarily be suitable for another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher.

In this context, the researcher must carefully scrutinise the definition of various terms and units of collection used, at the time of collecting the data from the primary source originally. Similarly, the object, scope and nature of original enquiry must also be studied. If the researcher finds that the differences in the data will remain unsuitable for the present enquiry, it should not be used.

- 3) Adequacy of data: If the level of accuracy achieved in the data is found inadequate for the purpose of present enquiry, they will be considered as inadequate and should not be used by the researcher. The data will also be considered inadequate, if they are related to an area which may be either narrower or wider than the area of present enquiry.

With secondary data, people have to compromise between what they want and what they are able to find.

The merits of secondary data are:

- Secondary data is cheaper to obtain. Many government publications are relatively cheap and libraries have stock quantities of secondary data produced by the government, by companies and by other organisations.
- Large quantities of secondary data can be accessed through the internet and online databases.
- Most of the available secondary data has been collected over a period of several years and therefore it can be used to plot trends.
- Secondary data is valuable to the government, business and research areas. For governments, it helps in making decisions and in planning future policies. In the business and industry areas such as marketing and sales, it is used to appreciate the general economic and social conditions and to provide information of competitors. To the research organisations, it provides social, economical and industrial information.

The demerits of secondary data are:

- It is difficult to judge whether the secondary data is sufficiently accurate and also reliable.
- It might be difficult to fit secondary data to the needs of the investigator.
- Secondary data might not be available for certain investigations. In such situations, primary data has to be collected.

The differences between primary and secondary data are listed in the table 2.4.

Table 2.4: Differences between Primary Data and Secondary Data

Primary Data	Secondary Data
1. Data is original and thus more accurate and reliable.	1. Data may not be reliable.
2. Gathering data is expensive.	2. Gathering data is cheap.
3. Data is not easily accessible.	3. Data is easily accessible through internet or other resources.
4. Most of the data is homogeneous.	4. Data is not homogeneous.
5. Collection of data requires more time.	5. Collection of data requires less time.
6. Extra precautionary measures need not be taken.	6. Data selection needs extra care.
7. Data gives detailed information.	7. Data may not be adequate.

Self Assessment Questions

10. State whether the following data are Primary or Secondary.
- An official of the Census Board of India is preparing a report on census of population based on the survey data that is collected by the Census Board.
 - An HR representative of a software company is deciding on the time taken to perform a particular job on a project on the basis of random observations collected by him.
 - A neurologist is examining the relationship between cigarette smoking and brain tumor based on the data published in a famous neurology journal.
11. When population under investigation is infinite, we should use
- sample method
 - census method
 - neither census nor sample method
 - both a & b

2.4.3 Pilot survey

Pilot survey is a small trial survey undertaken before the main survey. It gives a measure of efficiency of the questionnaire and reduces the inconveniences and loss of information. It helps in introducing necessary changes in the main survey.

Regarding the nature of population under study, when prior information, operational and cost aspects of data collection and analysis is not available from surveys, it is desirable to design and carry out a pilot survey.

Pilot survey is a preliminary research conducted before a complete survey to test the effectiveness of conducting the research. Pilot survey should be completed before the final survey begins. By conducting the pilot survey, the investigator will be able to know any difficulties that might arise that were not known at the survey proposal stage.

Pilot surveys have several other advantages:

- Pilot surveys provide the investigator with many ideas, approaches and clues that are not foreseen before conducting the pilot survey. Such ideas and clues increase the chances of getting accurate findings in the main survey.
- Pilot surveys help in making necessary alterations in the data collecting methods. Hence investigators can analyse data in the main survey more efficiently.
- Pilot surveys save a lot of time and provide enough data for the investigator to decide whether to go ahead with the main survey or not.

Apart from advantages, pilot survey also has certain limitations that are discussed here:

- Pilot surveys are not based on strong statistical foundation and are based on very small sample sizes.
- There is a possibility that the investigator might make wrong predictions or assumptions on the basis of pilot data.
- If data and results from pilot surveys are included in the main survey then it might lead to incorrect decisions.
- If the pilot participants are included in the main survey, then data obtained from these participants might result in corruption of main data.

2.5 Measurement Scales

Variables differ in how well they can be measured, that is, how much measurable information their measurement scale can provide. There is obviously some measurement error involved in every measurement, which determines the amount of information that we can obtain.

Another factor that determines the amount of information that can be provided by a variable is its type of measurement scale. Specifically variables are classified under two categories – qualitative and quantitative.

2.5.1 Qualitative (categorical) data

Qualitative, also known as categorical data, cannot be measured on a numerical scale (quantified). Examples of categorical variables are gender (male or female) and size of T-shirt (XXS, XS, S, M, L, XL and XXL); yet, these two variables differ in a sense; the first is said to be nominal or purely categorical whereas the second is known as ordinal.

Nominal (purely categorical) data

Nominal variables allow for only qualitative classification. They can be measured only in terms of whether the individual items belong to some distinctively different categories; however, we cannot quantify or even rank order these categories. For example, 2 individuals are different in terms of a certain variable (for example, they are of different race), we cannot say which one has more of the quality represented by the variable. Typical examples of nominal variables are gender, race, colour, city, marital status, etc.

Example 1

Marital status

- | | |
|---|--|
| 1. Never married <input type="checkbox"/> | 4. Married/Cohabiting <input type="checkbox"/> |
| 2. Divorced <input type="checkbox"/> | 5. Separated <input type="checkbox"/> |
| 3. Widowed <input type="checkbox"/> | |

Clearly, the numbers associated with the options above have no numerical significance. Comparison between values is impossible and also descriptive statistics like the mean and standard deviation would make no sense if calculated.

Ordinal data

Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, however they do not allow us to say how much more. A typical example of an ordinal variable is the socioeconomic status of families. For example, we know that upper-middle is higher than middle but we cannot say that it is, for example, 18% higher. Also, this very distinction between nominal, ordinal, and interval scales itself represents a good example of an ordinal variable. For example, we can say that nominal measurement provides less information than ordinal measurement, but we cannot say how much less or how this difference compares to the difference between ordinal and interval scales.

Example 2**Employee's performance**

- | | |
|---------------------------------------|---------------------------------------|
| 1. Excellent <input type="checkbox"/> | 4. Poor <input type="checkbox"/> |
| 2. Good <input type="checkbox"/> | 5. Very poor <input type="checkbox"/> |
| 3. Average <input type="checkbox"/> | |

It can be easily deduced that 'Excellent' is better than 'Poor', that is, there is a latent scale on which comparison can be made among the various values.

Ordinal data can sometimes be treated as interval for the sake of statistical analysis, provided the assumption is founded. In this case, the values of the variable are mathematically considered to be 'equidistant' on its scale. The numbers associated with each value starts to get some numerical significance so that the mean, though not very convincingly, maybe statistically interpreted.

The variable 'Employee's performance' in Example 2 can be regarded as interval if we assume that the 'distance' between any pair of successive values is equal (for example, the distance between 'Excellent' and 'Good' is the same as that between 'Average' and 'Poor'). In this case, if the average performance score of 100 employees is calculated and found to be, say, 3.2, we may, within some margin of security, conclude that the overall performance of employees is just above 'Average', the latter having been assigned a value of 3.

However, it would be statistically dangerous to assume an interval scale for the following example.

Example 3			
Educational level			
1. None	<input type="checkbox"/>	5. Diploma	<input type="checkbox"/>
2. Primary	<input type="checkbox"/>	6. Degree	<input type="checkbox"/>
3. Vocational	<input type="checkbox"/>	7. Postgraduate	<input type="checkbox"/>
4. Secondary	<input type="checkbox"/>	8. Professional	<input type="checkbox"/>

It is clear that the 'distance' between 'None' and 'Primary' is not equal to that between 'Diploma' and 'Degree'.

2.5.2 Quantitative (numerical) data

Quantitative data can be easily measured on a numerical scale; variables which can be quantified in terms of units are all quantitative. Examples of quantitative variables are number of students per class and height (measured in centimetres). Again, these two variables differ in their nature; the first is said to be discrete whereas the second is continuous.

Discrete data

Discrete data occur as definite and separate values; a discrete variable assumes values which are countable so that there are gaps between its successive values. For example, when counting the number of children in a class, we use numbers (0, 1, 2... n).

Continuous data

Continuous data occur as the whole set of real numbers or a subset of it. In other words, there are no gaps between successive values so that a continuous variable assumes all the values (including all the decimals) between given boundaries. Temperature is a good example of a continuous variable – though thermometer readings are recorded to the nearest tenth of a degree (Centigrade or Fahrenheit), temperature does not 'jump' from, for example, 17.1^o C to 17.2^o C. It passes through all the real numbers between these two values. Height, weight and speed are also continuous variables.

Continuous data can be measured on interval and ratio scales.

Interval scale

Interval variables allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale. We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees. However, interval scale variables do not have an absolute zero. If the temperatures in Singapore and London are 30° C and 15° C respectively, we cannot say that it is twice as hot in Singapore than in London. This is simply because it would not be the case if these temperatures were measured in degrees Fahrenheit: 86° F and 59° F respectively.

Ratio scale

Ratio variables are very similar to interval variables. In addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus they allow for statements such as x is two times more than y. Typical examples of ratio scales are measures of time or space. For example, as the Kelvin temperature scale is a ratio scale, a temperature of 200 degrees is higher than 100 degrees, and it is twice as high. Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales. Height is also a ratio scale variable since, if a person is twice as tall as another, he/she will remain so, irrespective of the units used (centimetres, inches, etc...).

Figure 2.7 depicts the various categories of Measurements Scales.

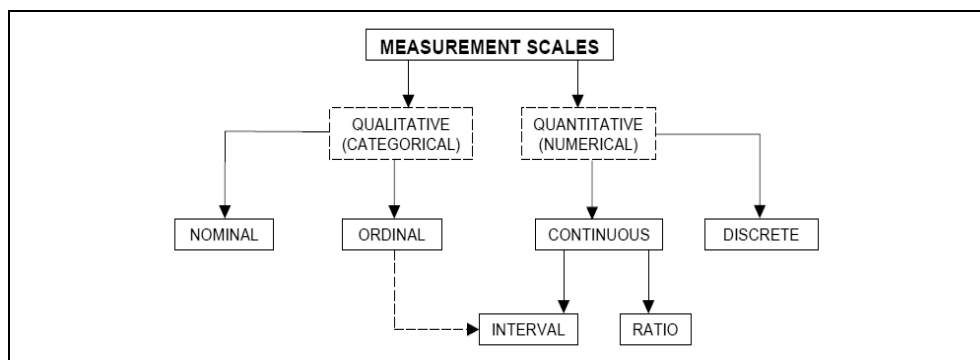


Fig. 2.7: Categories of Measurements Scales

2.6 Scrutiny and Editing of Data

Before using the collected data, it should be checked for its completeness, accuracy and reliability. By complete, we mean that all the required information should be available. Editing the data is important and it is a time consuming process.

The data collected through various sources will be highly unorganised and needs to be summarised and analysed for further studies. There is a possibility of missing the valuable data after summarisation. Hence, proper planning is required in the editing process of any collected data. While editing, it is important to have all the sources of collected data, and also the overall scope of survey.

There are different steps involved in editing the collected data. The data must be checked for:

Legibility: The data must be legible. If a response is not presented clearly, the investigator has to rewrite it.

Completeness: An unanswered response on a questionnaire implies that, either the respondent did not answer the entry or the investigator did not record the data. If the fault is from investigators in making an entry, then the investigator has to fill the missing entry. If an entry is missing as a result of omission of that entry by the respondent, then the investigator has to conduct the survey again to gather the missing entry.

Consistency: The investigator has to examine each questionnaire to check inconsistency or inaccuracy in any statement. For example, the numerical figures of attributes such as income, height, weight may be inconsistent. In such cases, it is the duty of the concerned investigators to make the necessary corrections. The investigators have to make sure that the collected data must be free from redundant responses or duplicate entries.

Self Assessment Questions

12. State True or False:

- i) Census conducted by Government of India is an example of primary data.
- ii) TV News Bulletins gather information on any event through their agents.

- iii) Schedules make respondents record their answers.
 - iv) A covering letter to the questionnaire brings confidence in respondents.
 - v) Questions in questionnaire should be lengthy.
13. State whether each of the following variables is qualitative or quantitative.
- i) Age
 - ii) Gender
 - iii) Class Rank
 - iv) Number of people favouring the death penalty
14. State whether each of the following variables is qualitative or quantitative and indicates the measurement scale that is appropriate for each.
- i) Annual sales
 - ii) Soft drink size (small, medium, large)
 - iii) Employee classification (GSI through GSIS)
 - iv) Earning per share
 - v) Methods of payments (cash, check, credit card)

2.7 Summary

Let us recapitulate the important concepts discussed in this unit:

- A statistical survey is a search for knowledge. There are two main stages in any statistical survey - Planning and Execution.
- Planning a statistical survey encompasses the following issues:
 - i) The nature of a problem
 - ii) The objectives
 - iii) The scope
 - iv) Statistical units
 - v) The degree of accuracy
 - vi) The time period
 - vii) The source of information and
 - viii) The organisation
- The collected data should be edited, analysed and interpreted for completeness, accuracy and consistency.

- Sample is a subset of the population. Sample can never be larger than the population from which the sample was taken.
- Quantitative characteristic is a characteristic which is numerically expressed; otherwise it is a qualitative characteristic.
- The quantitative characteristic that varies from unit to unit is called a variable. The qualitative characteristic that varies from unit to unit is called an attribute.
- There are two categories of data - Primary and Secondary data. Primary data is collected directly from the respondents.
- Any information, that is used for the current investigation but is obtained from some data, which has been collected and used by some other agency or person in a separate investigation, or survey, is known as secondary data. They are available in a published or unpublished form.
- The various methods of collecting Primary data are:
 1. Direct personal observation
 2. Indirect oral interview
 3. Information through agencies
 4. Information through mailed questionnaires
 5. Information through schedule filled by investigators
- Questionnaires must be structured well and must not be ambiguous. A covering letter must be included along with the questionnaire. Pilot survey is a beneficial method when prior information about the survey does not exist or when the results about the survey is needed quickly.

2.8 Glossary

Interval scale: An interval scale is a scale of measurement where the distance between any two adjacent units of measurement (or 'intervals') is the same but the zero point is arbitrary.

Nominal data: A set of data is said to be nominal if the values/observations belonging to it can be assigned a code in the form of a number where the numbers are simply labels.

Ordinal data: A set of data is said to be ordinal if the values/observations belonging to it can be ranked (put in order) or have a rating scale attached. You can count and order, but not measure, ordinal data.

Population: The set of all elements of interest in a particular study.

Primary data: Data collected for the first time keeping in view the objective of the survey.

Qualitative variables: A variable with qualitative data.

Quantitative variables: A variable with quantitative data.

Ratio scale: Ratio variables are very similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point.

Sample: A subset of the population.

Secondary data: Any information, which is used for the current investigation collected by some other agency or person in a separate investigation.

Statistical survey: A scientific process of collection and analysis of numerical data.

2.9 Terminal Questions

1. What is statistical survey?
2. Enumerate the factors which should be kept in mind for proper planning.
3. What do you understand by the unit of measurement? Explain with examples.
4. Distinguish between:
 - a) Primary and secondary data
 - b) Direct and indirect investigation
 - c) Questionnaire and schedule

2.10 Answers

Self Assessment Questions

1. Planning and execution
2. Planning
3. Yes
4. i) Finite ii) Infinite iii) Finite
5. i) Attribute ii) Variable
6. i) Discrete ii) Continuous
7. i) both Art & Science

8. i) Quantitative data
9. i) Statistics
10. i) Primary data, ii) Primary data, iii) Secondary data
11. i) sample method
12. i) True ii) True iii) False iv) True v) False
13. i) Quantitative ii) Qualitative iii) Qualitative iv) Quantitative
14. i) Quantitative, Ratio, ii) Qualitative, Nominal, iii) Qualitative, Ordinal, iv) Quantitative, Ratio, v) Qualitative, Nominal

Terminal Questions

1. Refer section 2.1.3.
2. Refer section 2.2.1.
3. It refers to the unit of the population on which measurements are made, for example, the height of employees in an office. Employees are individuals or units. Height is the measurement made on them.
4. a) Data collected for the first time by the investigator is primary data. Data collected by some other persons but used by the investigator for his/her study is known as secondary data.
b) Direct investigations are carried out directly by the investigator. Investigation conducted through mail questionnaire is called indirect investigation.
c) Questionnaires contain simple questions and are filled by respondents. Schedules also contain questions but responses are recorded directly by the investigator.

2.11 Case Study

Case Study 1

A firm is interested in testing the advertising effectiveness of a new television commercial. As part of the test, the commercial is shown on a 6.30 P.M local news program in Denver, Colorado. Two days later, a market research firm conducts a telephonic survey to obtain information on recall rates (percentage of viewers who recall seeing the commercial) and impression of the commercial.

Discussion Questions:

1. What is the population of the study?
2. What is the sample for this study?
3. Why would a sample be used in this situation? Explain.

Case Study 2

An AMC (Annual Maintenance Contract) company provides onsite IT support of hardware services clients. At one of the client's establishment, the hardware comprises 500 personal computers (PCS) and ten servers connected by local area network. The AMC covers, interalia, the maintenance of servers and PCs and the network on 24/7 basis.

The company has a team of 10 technical engineers and a coordinator posted at the client's establishment. The company is faced with the problem of too many complaints about the promptness and quality of service. The company wants to analyse the problem, for arriving at some appropriate solution.

Discussion Questions:

Design a questionnaire that would help the company in collecting relevant data and initiate remedial action. The questionnaire may cover the following aspects and also any other relevant issues.

- Technical competency
- Promptness
- Behavioural

Case Study 3

Telecom Company wanted to understand the perception of consumers about value added services of mobiles companies, with a view to add some new services in this segment. A consultant was hired, and a survey was planned. The following questionnaire was designed by the consultant.

Questionnaire for Consumers**1) Demographic Profile**

2) Name **Sex:** Male/ Female

3) Occupation: Employed / Self Employed/ Student/Housewife/Retired

4) Annual household Income (in Rs)

- i) ≤ 2 lakhs ii) 2 – 5 lakhs iii) 5–10 lakhs iv) ≥ 10 lakhs

5) Usage Patterns

How long have you owned a cell phone? Month

6) Cellular Operator being used:

- i) Vodafone
ii) Reliance
iii) Airtel
iv) BSNL
v) Others

7) What do you normally use the phone for? (You can choose more than one option)

- a) Keeping in touch with friends and family
b) Business
c) For emergencies
d) Others (please specify)

8) Your monthly expenditure on the mobile connection is around:

- i) \leq Rs 500
ii) Rs 500 – 1000
iii) Rs 1000 – 2000
iv) \geq Rs 2000

9) The mobile charges are paid for by: Self /Company/Spouse or Parent**10) Value Added Services**

Kindly tick(✓) those applicable

Services	Not Aware	Never Used	Used Occasionally	Used frequently
----------	-----------	------------	-------------------	-----------------

- a) SMS
b) Voice Mail
c) Messenger Services
d) Ringtones
e) GPRS
f) MMS
g) Roaming
h) Internet

11) Reasons for not using some services:

- i) Not Aware
- ii) Too Expensive
- iii) Complicated
- iv) No Utility/Time

Discussion Question:

Suggest improvements in the questionnaire.

(T.N. Srivastava & Shailaja Rejo (2008) *Statistics for Management* 5th ed. TMH)

References:

- Agarwal, B.L. (2006). *Basic Statistics*, 4TH Ed, New Age International Publishers.
- Bowerman, B. L & Connel, R.T.O. *Applied Statistics: Improving Business Processes*, Irwin 1996.
- Levin, R. I. & Rubin, D.S. (2008). *Statistics for Management*, 7th Ed, PHI Learning Private Limited.
- Lipschutz, S. & Schiller, J.J. *Schaum's Outline of Introduction to Probability and Statistics (Schaum's Outline Series)* (Sep 7, 2011).
- Pisani D.R. Freedman & R. Purves, *Statistics*, 3rd Ed, W.W Norton 1997.
- Schiller, J, Srinivasan, R. Alu, and Spiegel, Murray, *Schaum's Outline of Probability and Statistics*, 3rd Ed. (Schaum's Outline Series) (Aug 26, 2008).
- Spiegel, M. & Stephens, L. *Schaums Outline of Statistics, Fourth Edition (Schaum's Outline Series)* (Jan 31, 2011).
- Sternstein, M. *Barron's AP Statistics with CD-ROM (Barron's AP Statistics (W/CD))* (Feb 1, 2010).
- Tanur, J.M, *Statistics: A Guide to the unknown*, 4th Ed, Brooks /cole, 2002.
- Voelkar, D.H, Orton, P. Z. & Adams, S. *Statistics (Cliffs Quick Review)* (Jun 15, 2001).
- Wilcox, R.R. (2009). *Basic Statistics – Understanding Conventional Methods and Modern Insights*, Oxford University Press.

E-References:

- <http://www.textbooksonline.tn.nic.in/Books/11/Stat-EM/Chapter-3.pdf>.