SOCIAL RESEARCH

BA [Sociology]

Paper IV

RAJIV GANDHI UNIVERSITY

Arunachal Pradesh, INDIA - 791 112

Authors:

Dr. Harish Kumar, Units (1.2.1, 2.2, 2.3, 3.2.2, 4.3, 5) © Dr. Harish Kumar, 2018

Lokesh Kaul, Units (1.4, 3.2-3.2.1) © Lokesh Kaul, 2018

Deepak Chawla & Neena Sondhi, Units (2.4, 7) © Reserved, 2018

Dr. JS Chandan, Unit (6.5) © Dr. JS Chandan, 2018

Vikas Publishing House, Units (1.0-1.1, 1.2, 1.2.2, 1.3, 1.5-1.9, 2.0-2.1, 2.2.1, 2.5-2.9, 3.0-3.1, 3.3-3.10, 4.0-4.1, 4.2, 4.4-4.6, 4.7-4.11, 6.0-6.4, 6.6-6.10) © Reserved, 2018

About the University

Rajiv Gandhi University (formerly Arunachal University) is a premier institution for higher education in the state of Arunachal Pradesh and has completed twenty-five years of its existence. Late Smt. Indira Gandhi, the then Prime Minister of India, laid the foundation stone of the university on 4th February, 1984 at Rono Hills, where the present campus is located.

Ever since its inception, the university has been trying to achieve excellence and fulfill the objectives as envisaged in the University Act. The university received academic recognition under Section 2(f) from the University Grants Commission on 28th March, 1985 and started functioning from 1st April, 1985. It got financial recognition under section 12-B of the UGC on 25th March, 1994. Since then Rajiv Gandhi University, (then Arunachal University) has carved a niche for itself in the educational scenario of the country following its selection as a University with potential for excellence by a high-level expert committee of the University Grants Commission from among universities in India.

The University was converted into a Central University with effect from 9th April, 2007 as per notification of the Ministry of Human Resource Development, Government of India.

The University is located atop Rono Hills on a picturesque tableland of 302 acres overlooking the river Dikrong. It is 6.5 km from the National Highway 52-A and 25 km from Itanagar, the State capital. The campus is linked with the National Highway by the Dikrong bridge.

The teaching and research programmes of the University are designed with a view to play a positive role in the socio-economic and cultural development of the State. The University offers Undergraduate, Post-graduate, M.Phil and Ph.D. programmes. The Department of Education also offers the B.Ed. programme.

There are fifteen colleges affiliated to the University. The University has been extending educational facilities to students from the neighbouring states, particularly Assam. The strength of students in different departments of the University and in affiliated colleges has been steadily increasing.

The faculty members have been actively engaged in research activities with financial support from UGC and other funding agencies. Since inception, a number of proposals on research projects have been sanctioned by various funding agencies to the University. Various departments have organized numerous seminars, workshops and conferences. Many faculty members have participated in national and international conferences and seminars held within the country and abroad. Eminent scholars and distinguished personalities have visited the University and delivered lectures on various disciplines.

The academic year 2000-2001 was a year of consolidation for the University. The switch over from the annual to the semester system took off smoothly and the performance of the students registered a marked improvement. Various syllabi designed by Boards of Post-graduate Studies (BPGS) have been implemented. VSAT facility installed by the ERNET India, New Delhi under the UGC-Infonet program, provides Internet access.

In spite of infrastructural constraints, the University has been maintaining its academic excellence. The University has strictly adhered to the academic calendar, conducted the examinations and declared the results on time. The students from the University have found placements not only in State and Central Government Services, but also in various institutions, industries and organizations. Many students have emerged successful in the National Eligibility Test (NET).

Since inception, the University has made significant progress in teaching, research, innovations in curriculum development and developing infrastructure.

SYLLABI-BOOK MAPPING TABLE

Social Research

Syllabi	Mapping in Book
Unit 1 Understanding Social Research Meaning, Scope and Significance of Social Research; Major Steps in Social Research; Qualitative and Quantitative Research.	**Unit 1:** Understanding Social Research **(Pages 3-22)**
Unit 2 Hypothesis Conceptualisation and Formulation of Hypothesis; Importance of Hypothesis in Social Research and Source of Hypothesis.	**Unit 2:** Hypothesis **(Pages 23-51)**
Unit 3 Scientific Study of Social Phenomena The Scientific Method, Objectivity and Subjectivity, Debate in Social Research; Positivism in Sociology.	**Unit 3:** Scientific Study of Social Phenomena **(Pages 53-71)**
Unit 4 Types of Research Basic, Applied; Historical, Empirical; Descriptive, Exploratory, and Experimental.	**Unit 4:** Types of Research **(Pages 73-94)**
Unit 5 Techniques of Data Collection Questionnaire, Schedule, Interview Case Study, Observation and Content Analysis.	**Unit 5:** Techniques of Data Collection **(Pages 95-135)**
Unit 6 Analysis and Use of Statistics Analysis of Data, Coding, Tables, Graphs and Diagram; Use of Statistics–Mean, Median, Mode and Standard Deviation.	**Unit 6:** Analysis and Use of Statistics **(Pages 137-198)**
Unit 7 Report Writing Importance of Report Writing; Components of Report–Preliminary Pages, Main Text and End Text.	**Unit 7:** Report Writing **(Pages 199-222)**

CONTENTS

INTRODUCTION

Research is the search for knowledge or a systematic investigation in order to establish facts. The basic aim of research is to discover, interpret and develop methods and systems to advance human knowledge on diverse scientific matters. Social research refers to the conduction of research on various groups of a society by social scientists. Research methodology refers to the way research can be conducted. It is also known as the process of collecting data for various research projects.

Social research pertains to research carried out by social scientists on various facets of society. Research plays a very significant role in the field of social science. In order to study the importance and relationship between social science and research, social research is conducted or undertaken. The research that attempts to measure, describe, explain and predict the social and economic phenomena or social behaviour of human beings is known as 'social research'.

The methodology of social research is the science of studying how research is conducted scientifically. It helps to understand both the products as well as the process of scientific enquiry. A research process involves selection and formulation of a research problem, research design, sample strategy or sample design, as well as the interpretation and preparation of research report. Research can be undertaken in the form of descriptive/ survey research, applied or fundamental research, quantitative or qualitative research, conceptual or empirical research, and other types of research.

This book, *Social Research*, is written in a self-instructional format and is divided into seven units. Each unit begins with an *Introduction* to the topic followed by an outline of the *Unit Objectives*. The content is then presented in a simple and easy-to-understand manner, and is interspersed with *Check Your Progress* questions to test the reader's understanding of the topic. A list of *Questions and Exercises* is also provided at the end of each unit, and includes short-answer as well as long-answer questions. The *Summary* and *Key Terms* section are useful tools for students and are meant for effective recapitulation of the text.

UNIT 1 UNDERSTANDING SOCIAL RESEARCH

Structure

1.0 INTRODUCTION

Simply defined, research is a search for knowledge. One can also define research as a scientific and systematic pursuit of information on a specific topic. Scientifically, research can also be termed as scientific investigation. Thus, research and scientific enquiry can be considered synonymous. The only difference between the two is that while it is possible to employ scientific method without research, it is not possible to conduct any research without employing scientific methods. Thus, research is a more specialized form of scientific enquiry which in turn is the result of gathering of data, information and facts for the specific purpose.

Social research pertains to research carried out by social scientists on various facets of society. Research plays a very significant role in the field of social science. In order to study the importance and relationship between social science and research, social research is conducted or undertaken. The research that attempts to measure, describe, explain and predict the social and economic phenomena or social behaviour of human beings is known as 'social research'. In this unit, you will get acquainted with the meaning, characteristics and objective of scientific research, aims and types of social research, steps in social research and the concept of qualitative and quantitative research.

1.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Explain the meaning of the term social research

- Analyse the scope and significance of social research

- Assess the types of social research

- Describe the major steps involved in social research
- Assess the concept of qualitative and quantitative research

1.2 MEANING OF SOCIAL RESEARCH

Society is an amalgamation of individuals with different needs, aspirations and goals in life. However, social individuals are also associated with each other through shared interests, familial bonds and common objectives. Social research is basically research conducted by social scientists in order to analyse a vast breadth of social phenomena. The methods used in social research find their roots in classical sociology and statistics. Social research methods may be divided into two broad divisions. These are: Qualitative and Quantitative methods. While the former approaches social phenomena through quantifiable evidence, the latter approaches social phenomena through observation, communication with partners and analysis of text. However, the choice of method depends largely on what the scientist wishes to investigate. Prof. Bent Flyvbjerg of Oxford University maintains that the divide between the quality and quantity oriented camps in social research is clearly unfortunate as good research methods require a combination of both.

Definitions

While C.A. Moser defines social research as: 'A systematized investigation to gain new knowledge about social phenomenon and problems', P.V. Young maintains: 'Social research is a scientific undertaking which by means of logical methods, aim to discover new facts or old facts and to analyse their sequences, interrelationships, casual explanations and natural laws which govern them.'

Objectivity in Social Research

Social scientists are often influenced by their biases, passions, likes and dislikes and preconceived notions. These are seen to interfere with the scientific objectivity that they would need while researching on social sciences. Objectivity is the capacity to represent truthfully and without prejudice, the results of one's research. A social researcher needs to be aware of his personal biases and prejudices and take adequate care that these do not affect the objectivity of the research. Max Weber, an exponent in social research argued, that actually, the thoughts and beliefs of the researchers *should* affect their topics of study. However, the social scientist needs to be value-neutral once the research question has been framed. Objectivity can be attained by sharing the results of research with experts who then may be asked to critically examine them. In his *Logic of Scientific Discovery* (1959), Karl Popper maintained that *confirmation* and *refutation* are the essence of scientific discovery. Social researchers publish their work so that their work can be scrutinized by others. Journals have dedicated teams to decide whether the research material lives up to the standard of the journal and should, therefore, be published. Once a research material is published, other scholars look at it critically, especially when they do not agree with the findings.

Some others may wish to replicate the study by changing the strategies and settings to check if the conclusion would remain the same.

Characteristics of Social Research

Social research possesses certain unique characteristics. These are as follows:

- Social research is directed towards finding solutions for social problems.
- It emphasizes the development of generalizations, theories and principles that help in predicting future occurrences.
- It is primarily based on empirical/observable experience.
- It requires meticulous observation.
- Though it may appear to be unsystematic, social research most often involves carefully designed procedure.
- It requires an expert researcher who is already acquainted with the previous nuances of the problem.
- It is characterized by patient and unhurried activity.

1.2.1 Scope and Significance of Social Research

The subject matter of sociology is society. Sociologists study man's social behaviour in a variety of contexts. They use a number of methods in social research including 'comparative method', 'participant observer method', 'community studies', etc. Descriptive and explanatory research aims only at describing, in detail, a situation or set of circumstances. On the other hand, action research refers to 'that is done when some reform or change has been introduced. Its purpose is to monitor the effect of the change and to decide whether it has achieved what it was supposed to achieve.'

Most disciplines undertake research. Research is more of a way of thinking than a set of skills. Research entails critically examining aspects of the study; making guiding principles for testing particular procedure; developing testing theories, etc.

For any study undertaken to be called a 'research', it should adhere to the following three criteria:

- A set of philosophies guide the research
- Methods, techniques and procedures which have proven reliability and validity are used
- Research has to be objective as well as unbiased

The philosophical orientation of research may stem from one of the two paradigms in research—*positivism* and *interpretivism*. Validity ensures that in a research study correct procedures have been applied to find answers to a question. Reliability refers to quality of a measurement procedure. 'Unbiased and objective' means that researchers take each step and draws each conclusion to the best of their ability without introducing their own biases and prejudices. (Ranjit Kumar, 1999).

Research plays a very significant role in the field of social science. In order to study the importance and relationship between social science and research, social research is conducted or undertaken. Research that attempts to measure, describe, explain and predict the social and economic phenomena or social behaviour of human beings is known as 'social research'.

One of the main objectives of conducting social research is to find information about the behaviour of an individual and solutions to the problems related to human relations. The outcome of social research provides the following benefits:

- It helps professionals in earning their livelihood.
- It helps students in knowing how to write a report for various findings.
- It helps philosophers to think on wider new perspectives.
- It helps in developing new styles for creative work.

In order to conduct social research and examine the social life of human beings, social scientists use different methods. Quantitative and qualitative research are the two methods of research that are generally used by social scientists to conduct a research. In quantitative method, numerical data is collected and then analysed in order to measure the social phenomena. Qualitative method is basically the study of data, such as words, pictures and objects. However, the data collected with the help of this method is not very effective and cannot be generalized very easily.

Social research is very helpful for a country as it helps the government to explore the following things:

- Social and economic structures
- Social attitudes
- Social values and behaviours
- Factors motivating individuals and groups of a society

Researchers share a close relationship with government analysts, such as economists, statisticians and operational researchers. The relationship between researchers and government analysts is essential in order to find out high quality research data. Social research also informs about development, implementation and evaluation of a wide range of government policies.

Social research also helps to examine the consequences of government policies and economic changes in an organization, and the effects of globalization and its impact on small-scale and cottage industries.

1.2.2 Types of Social Research

Sociologists employ a variety of methods to learn about the social world. These methods are not mutually exclusive. Since each research method has strengths and weaknesses, a good research strategy may use several of them. Appelbaum and Chambliss (1997:40) hold that the principal methods of social research include survey and fieldwork.

Survey

A survey entails administering a precisely worded questionnaire to a group of people in order to determine their characteristics, opinions and behaviours. First, the researcher has to define a *population universe* to which the study applies: this is the group of people about whom generalization is to be made. Once the population universe is identified, a *sample*—a subset of cases selected to represent the larger population—must be selected, since it is seldom economically feasible or desirable to interview everyone in a chosen population universe.

Two principal type of sampling are used: *probability* and *non-probability sampling.* In the most common type of probability sampling, termed *random sampling,*

everyone in the population universe has an equal chance of being in the sample. In *non-probability sampling*, subjects are deliberately chosen because of their specific characteristics. Once the sample is constructed or drawn, the questionnaire is administered. Questionnaires may contain *open- or close-ended questions.* In *close-ended questions*, the respondents are required to choose only from predetermined alternative responses. In *open-ended questions,* there are no fixed responses to choose from. The respondents are provided with a wide range of opportunities to express a wide range of feelings and opinions.

One of the strengths of survey method is that it permits the researcher to draw conclusions about a large number of people on the basis of a much smaller number of interviews. This is a major advantage in terms of time and money. Surveys also have some weaknesses. Sometimes, surveys can be superficial since in order to be feasible economically, they usually call for brief responses to close-ended questions. Many-a-times, responses are self-serving, just intended to make the interviewee look good in the eyes of the researcher.

Fieldwork

Fieldwork consists of many methods. The most common fieldwork is that of *participant observation*. The researchers become a part of the community under study; immerse themselves completely in the daily life of the community and participate in the activities of the members of the community but with a sense of detachment. They then attempt to report all their findings on every aspect of their lives with a sense of impartiality and disinterestedness. Classic examples of fieldworks are William Whyte's (1915) *Street Corner Society (1943)*, A Study of Italian-American Working-Class Men and B. Malinowski's *Study of the Tribes of Trobriand Island*.

Sometimes the research strategy requires that the researchers stay away from the people they are studying, and simply observe what is going on. A sociologist studying crowd behaviour at a rally or student participation in a seminar would be an example. The researcher in such a study tries to be a 'fly on the wall', invisible and unobtrusive, yet constantly records what is going on. This technique is called *detached observation.*

Interview is another method of fieldwork. It is a detailed conversation designed to obtain in-depth information about a person. In a *structured interview*, researchers have a detailed list of specific questions to ask. In the *semi-structured interview*, the researchers have a list of topics to cover depending on the interview situation, to determine the course of questioning and the details of the question.

Participatory research is another method under fieldwork. It is designed to involve the subjects of the research in the research process itself, with an eye to empowering them to overcome some difficulty or problem. This research is usually tied with community action. It is conducted when a group or community wants to engage in some form of social change but lacks the expertise to do so. The researcher is invited to become a fully engaged member of the social change process, helping the members of the group to conduct the necessary research and training them in the techniques for doing so.

Another method is *experiment.* In it, two groups are chosen—*the experimental group* and *the control group*. An experimental group is one which is exposed to the independent variable. The control group is kept constant—no experiment is carried out on it. In the end, both the groups are compared to find out the resultant effects of the experiment.

Working with *available information* is another strategy. This involves working with data collected by other people. Often such data are the only information available. Examples include statistical data, documentary analysis or comparative-historical research (study of several different countries as well as examination of changing historical patterns in a single country).

APPROACH	WHEN APPROPRIATE
Survey	Basic information about a large population is required and sampling is a feasible strategy.
Interview	In-depth information is desired and direct access to informants is possible.
Detached observation	Information should be gathered but the data gathering should be as unobtrusive as possible.
Participant observation	First hand knowledge of the direct experience of subjects is required.
Participatory research	Primary goal is empowerment: training people to acquire the necessary skill to do research themselves.
Experiments	To determine specific causal relationships.
Using the available information	Direct acquisition of data is either not feasible or not desirable.

1.3 MAJOR STEPS IN SOCIAL RESEARCH

Research process includes steps or a series of actions and logical sequence of those steps to carry out research effectively. The various steps in a research process are not mutually separate, exclusive or discrete, but they at the same time need not always follow each other. The researcher, at each step, anticipates subsequent steps and requirements. The tentative order of the steps and the procedural guidelines of the research process are as given below:

(i) **Formulating the research problem:** At the very beginning of research, the researcher must clearly define the research problem, i.e., the area of interest, the matter to be inquired into, etc. The problem, before being solved, is initially stated in a broader perspective and then the researcher arrives at the specific question by gradually reducing the ambiguities, if any. Then, immediately after formulating the problem, the feasibility of different solutions is studied before choosing the right solution.

(ii) **Extensive literature survey:** After formulating the research problem, a brief summary of it should be prepared—this is an essential step. While writing a Ph.D. thesis a researcher has to prepare a synopsis of the topic and submit it to the appropriate committee or research board for approval. Synopsis preparation needs extensive survey of the literature connected with the problem.

(iii) **Development of a working hypothesis:** After surveying the literature, the researcher should clearly state the working hypothesis, which is a tentative assumption made before testing it in logical or empirical sequences. Hypothesis must be as specific as possible and should be limited to the intended research. This helps to choose the right process.

(iv) **Preparing the research design:** The next step, after clearly defining the research problem, is preparing the suitable research design. The research design includes the conceptual framework within which the research would be carried out. A

good and planned research design helps to carry out the study in an efficient manner saving time and resources. It helps to gather the most useful information and assists in arriving at the accurate results. Simply put, a good research design facilitates the collection of relevant evidence with minimal expenditure of money, effort, time and other resources.

(v) **Determining sample design:** A universe or population includes all the items under inquiry. If all the items in the population are inquired then such an inquiry is called census inquiry. In a census survey, all the items are covered and so the highest accuracy is obtained. But this may not be practicable in surveys involving a big population. Census surveys need huge amounts of time, money and energy. Hence, quite often it is wise to select only a few items from the universe for the purpose of study. Technically, such a small and convenient number of items selected, is called a sample. Specified plan of the size and method of collecting the sample is technically known as sample design.

(vi) **Collecting the data:** In most cases, the data in hand is insufficient and there is always a need of fresh data. There are different ways of collecting the appropriate data which differ considerably in terms of relevance, expenditure, time and other resources. Therefore, the researcher must select the most appropriate method of collecting the data after considering the objective of the research, the nature of investigation, time and financial resources available, scope of the inquiry, and the desired degree of accuracy.

(vii) **Execution of the project:** This is an important step in the research process because if the execution proceeds on the correct lines, the collected data would be dependable, adequate and accurate. Therefore, systematic and timely execution of a project plays a crucial role in ensuring right results at the end.

(viii) **Analysis of data:** After collecting the data, the next step is analysing the data. The data analysis includes a number of closely-related operations like specifying different categories of data, differentiating and tabulating the data into different categories, applying the statistical techniques and formulae to the data, doing the right calculations and then drawing statistical inferences. Various tests, such as chi-square test, t-test, F-test, etc., help in data analysis.

(ix) **Hypothesis-testing:** After analysing the data, the researcher should test the working hypothesis against the statistical inferences obtained after analysing the data. The question that should be answered now is: Do the findings support the working hypothesis or do they contradict it?

(x) **Generalizations and interpretation:** If a hypothesis is tested and upheld sufficient number of times, the researcher can arrive at a generalization. The degree of success of a research is calculated on the basis of how close the arrived generalizations are to the acceptability. If the researcher starts with no hypothesis, the researcher will interpret his findings on the basis of some existing theory and this is known as **interpretation**. The process of interpretation often triggers new questions which lead to further researches.

(xi) **Preparation of the report or the thesis:** Finally, the researcher has to prepare the report of what has been studied. Report must be written with great care keeping the following layout in mind:

- **Preliminary pages:** These pages of the report should contain the title, the date, acknowledgments, foreword, table of contents, list of tables, list of graphs and charts (if any).

- **Main text:** The main text of the report should have introduction, summary of findings, main report, conclusion and suggestions for future research.
- **Closure:** At the end of the report, appendices should be listed in respect of all technical data, followed by bibliography. Index terms should also be given specially in a published research report. All references should be cited as per the research writing formats.

Flow Chart: Research Process

RESEARCH PROCESS IN FLOW CHART

Where (F) = feed back (Helps in controlling the sub-system to which it is transmitted)

(F) = feed forward (Serves the vital function of providing criteria for evaluation)
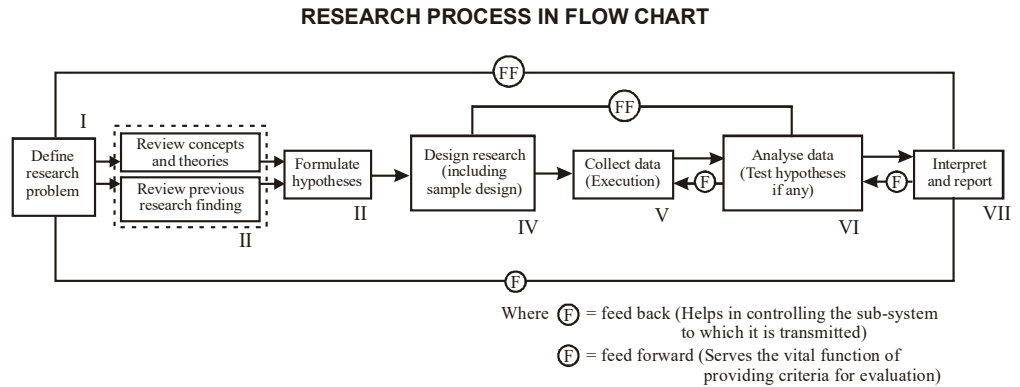
Fig. 1.1 Research Process

In Figure 1.1, the flow chart indicates the sequential steps to be followed in the research process, as studied in this section. We can recollect that the research process starts with defining the research problem along with reviewing the relevant literature in the field to become familiar with the concepts and theories relevant to the issue to be investigated. The next step is the formulation of the hypothesis, which is followed by the research design and sample selection. Then the collection of data and its analysis is to be attempted. After that the interpretation and the report writing stages complete the research report. These have to be written step by step and then edited and refined several times before preparing the final report.

Criteria of Good Research

Whatever be the type of research one undertakes, certain common criteria of good scientific methods have to be followed. A good research follows logical methods, is systematic, and structured in accordance with well-defined sets of rules and practices to enable the researcher in arriving at dependable conclusions. Both, deductive reasoning and inductive reasoning, should be followed for meaningful research.

Good research also implies obtaining reliable data which provides sound validity to the research findings.

The following principles underlie a good research criteria:

- The aim and objective of the research being conducted should be clearly specified.
- The research procedure should be replicable so that if the research needs to be continued or repeated, it can be done easily.
- The research design should be so chosen that the results are as objective as possible.

- Interpretation of any research should be done keeping in mind the flaws in the procedural design and the extent to which it has an effect on the results.
- Research should be carried out systematically. It should progress in pre-defined stages, and researchers should avoid using their intuition or guesswork to arrive at conclusions.
- Research should be logical so that it is meaningful, and help in decision-making.
- Research should be empirical as far as possible.
- The results of the research should only be used and generalized for the population for which the data provides an adequate basis.
- The validity and reliability of the data used in research should be double checked.

A good research produces results that are examinable by peers, methodologies that can be replicated, and knowledge that can be applied to real-world situations.

Problems Encountered by Researchers in India

There are some common problems faced by researchers in developing countries and India is no exception. Essentially, there is a dearth of tools required for good research. Many of the universities and research institutions are now providing computers with Internet connection to researchers but the facilities provided are not adequate. Luckily, the costs of both hardware and Internet bandwidth have reduced over a period of time. While Indian researchers now have easy access to these tools, there is still the problem of low visibility of papers published by them. Indian researchers often become demotivated to continue further research. Other factors like lack of scientific training in the methodology of research and a non-existent code of conduct also serve as challenges for the Indian researcher. There is also insufficient interaction between the researchers and the end-users. End-users of research are the ones who stand to benefit from the research and if they are not made aware of the benefit they can derive, getting sponsors to provide funds for research would be difficult.

There is also a lack of safeguards against any violation of confidentiality in data collection. Research studies that overlap lead to unnecessary repetition. There is an absence of research culture in our country.

Other problems that Indian researchers face that are common to developing countries are:

- Limited or no access to international research journals
- Lack of infrastructure except in a few metropolitan cities
- Low investment in research due to financial constraints
- Inadequate library facilities and where such facilities exist, they are not easily accessible
- Poor encouragement to do research

These problems need to be surmounted effectively in order to promote research as a professional activity.

Check Your Progress

4. What does a researcher need to consider before selecting the appropriate method of colleting the data?
5. List the constituents of the preliminary pages of a report.

1.4 QUALITATIVE AND QUANTITATIVE RESEARCH

In this section, we will study qualitative and quantitave research.

1.4.1 Logical Positivism: Quantitave Research

Logical positivism uses quantitative approach in the verification of theoretical propositions. It is based on statements such as 'anything that exists in a certain quantity and can be measured.' Quantification is essentially to enhance precision in the description of variables and the discernment of the relationships among them. It is structuring an empirical event into a mathematical model which, when juxtaposed on a specific mathematical proposition already formulated, verifies the latter.

Quantitative approach uses 'measurement' as the most precise and universally accepted method for assigning quantitative values to the characteristics or properties of objects or events for the purpose of discovering relationships between variables under study.

Measurement is defined as the assignment of numbers to objects and events according to logically accepted rules. There are certain properties of numbers that must have parallels in the observed phenomena. These properties are described as under:

1. *The property of identity:* A number has identity. Every number is unique and no other number is exactly the same.

2. *The property of order:* In the number system, numbers have their order or rank, i.e., one number being greater than another.

3. *The property of additivity:* In the system of numbers, summing of a certain number with certain other number must yield a unique number. This property is the basis for almost all useful operations that can be performed with numbers, because if the numbers can be added, they can also be subtracted, multiplied and divided.

It is not necessary that the phenomena to which the numbers are applied must have all the three properties of identity, order, and additivity in order to measure those phenomena. However, the advantage of the numbers used in measurement depends upon how many of the three properties do apply. There are several levels of measurement, the use of which are dependent upon the number of those properties that do apply. A convenient way of classification of the measurement levels is made by Stevens (1951). According to this classification, there are four levels of measurement scales:

1. *Nominal scales***:** Nominal scales of measurement are used when a set of objects among two or more categories are to be differentiated on the basis of qualitative differences. Usually, a number of symbols or numerals are chosen to represent all objects in a given category, thus taking the advantage of the property of identity. We may assign individuals to such categories as sex (male and females), nationality (Indians and Americans), educational level (school students and college students), professional rank (lecturers, readers and professors), etc.

   Each individual can be a member of only one category and all the members of the category have the same defined characteristics. Nominal scales are non-orderable and the only arithmetical operation applicable to such scales is counting, the mere enumeration of individuals in each category, class or set. The nominal scale is

primitive form of measurement and the statistical techniques based on counting are permissible in this type of measurement.

2. *Ordinal scales:* The ordinal scales of measurement correspond to quantitative classification of a set of objects. The sets or classes of objects are ordered on some continuum in a series ranging from lowest to highest according to the characteristics we wish to measure. The ranking of students in class for height, weight or scholastic achievement are the examples of ordinal scale of measurement. It may be noted that the successive intervals between consecutive points on the ordinal scale may not be equal throughout the entire scale. Suppose we place three students Ram, Sham and Ali in the order of height with Ram being the tallest, and assign the numbers 3, 2 and 1 respectively. All we have is information about serial arrangement. We cannot say that Ram, is as much taller than Sham as Sham is taller than Ali, even though the three numbers assigned to them are equally spaced on the scale of measurement. The common arithmetical operations — addition, subtraction, multiplication, and division — cannot be legitimately used with ordinal scales, but statistical procedures based on ranks are appropriate.

3. *Interval scales:* Taking the previous example, we actually measure the height of the three students Ram, Sham and Ali by using a metre scale and find their heights to be 185 cm, 172 cm and 159 cm., respectively, then we have measurements on a scale of equal units. This scale of measurement is called an interval scale. By this type of measurement we can make certain exact and meaningful decisions. We can say that Ram is 13 cm taller than Sham and 26 cm taller than Ali. We can also infer that the difference in heights between Ram and Ali is twice than that between Ram and Sham.

   In the interval scale, the differences between consecutive points on the scale are equal over the entire scale but there is no true zero point on it. The zero point of the scale is chosen conventionally or arbitrarily. Most psychological tests and inventories are based on interval scales. They have no real zero point. It is true that a student may occasionally get a score of zero on a test of mathematics or general science. However, this does not mean that the student has no knowledge of mathematics or general science.

   We can perform the operations of addition and subtraction on interval scales. The statistical techniques based on these arithmetical operations are permissible. The operations of multiplication and division cannot be legitimately used with interval scales, since these operations presuppose the existence of an exact zero point.

4. *Ratio scales:* The fourth and highest level of measurement is the ratio scale. All the four operations of addition, subtraction, multiplication and division can be used with ratio scales. All statistical techniques are permissible with such scales. These scales have all the characteristics of interval scales, with the additional advantage of a true zero point. It is possible to indicate the complete absence of an attribute. For example, the zero point on a centimetre scale indicates the complete absence of length or height.

   The numerals of the ratio scale can be expressed in ratio relationships. For example, the weight of 4 grams is one-half of 8 grams; the height of 10 cm is one-third of 30 cm, and so on.

   Ratio scales are almost non-existent in psychological measurement, except in the area of psycho-physical judgement. For example, when we measure reaction time, we use the customary time units, seconds and fractions of a second.

Ratio scales permit us to perform operations of (1) equality of magnitudes, (2) equality of differences, and (3) equality of ratios.

It may be noted that each higher scale can be transformed into a lower category scale but no lower category scale can be transformed into a higher category scale. The statistics permissible under higher category scale would also include the statistics that is permitted under the lower category scale. The higher we go in the hierarchy of the scales, the more relevant information is provided. If the nature of the variables permits, the scale that provides the most precise description should be used.

Quantitative research emphases: (*i*) concepts and their measurement in numbers (phenomenalism), (*ii*) establishment of casual relationships between independent and dependent variables, (*iii*) law-like generalizability of phenomena, (*iv*) replicability, (*v*) methodological individualism, and (*vi*) objectivity in the collection, analysis and interpretation of data, and reporting of results. It is dominated by the largely unquestioned, natural science paradigm of hypothetico-deductive methodology. Quantitative research actually goes beyond just analysing numbers. Using a deductive approach, it seeks to establish facts, make predictions, and test hypotheses that have already been stated.

Characteristics of Quantitative Research

The chief characteristics of quantitative research are mentioned as under:

1. Quantitative research uses deductive or 'top-down' approach. The researcher formulates and uses hypotheses and theory with data.

2. It is based on logical-positive paradigm which utilises experimental methods and quantitative measures to test specific hypothetical generalizations with narrow-angle lens.

3. Quantitative research uses scientific method with 'hard science' trappings. The researchers treat their subjects of study as having an existence independent of themselves and without any intrinsic meaning.

4. Behaviour of the subjects under study is assumed to be regular and predictable.

5. Most of the common research objectives in quantitative approach aim at description, explanation and prediction of social phenomenon. The emphasis is not on the deep understanding of the phenomenon or content.

6. Quantitative research attempts to study behaviour under controlled conditions. The nature of observation is objective, *i.e.*, different observers agree on what is observed.

7. Closed ended structured questionnaires, tests, attitudes scales, rating scales, etc. are used to collect quantitative data based on precise measurement.

8. The dominant sampling strategy in quantitative research is probability sampling, which depends on the selection of a random and representative sample from a larger population. The purpose of probability sampling is subsequent generalization of the research findings to the population from which the sample was selected. Generally, large samples are used in quantitative survey studies.

9. Quantitative research is deductive in that it tests theories which have already been proposed. It aims at analysis of representative and validated quantitative data, through the use of sophisticated statistical methods and software packages.

10. The findings are based on identified statistical relationships and generalizable findings. Using the principles of probability, quantitative research makes predictions representative of a large population.

11. The form of final report is statistical with details about the use of various types of statistics, e.g., correlations, comparison of means, percentages, etc. and their statistical significance.

In conclusion, quantitative research involves successive phases of hypothesis formulation, data collection, analysis and interpretation. Using deductive approach, it seeks to establish facts, make predictions, and test hypotheses that have already been stated.

Major Types of Quantitative Research

Keeping in view the distinguishing characteristics of quantitative research, following are the main types of approaches to quantitative research:

1. *Descriptive survey research:* This type of research attempts to answer questions about the current status of a phenomenon under study. Usually, it involves studying the preferences, attitudes, practices, concerns, or interests of some group of people.

2. *Correlation research:* These studies are conducted to determine whether, and to what degree, a relationship exists between two or more variables.

3. *Causal-comparative research:* This type of research seeks to discover a cause-effect relationship between two or more different programmes, methods, or groups. It is also called *ex-post-facto* research because in this type, the researcher usually does not have control over the causal factor or independent variable as it is studied after the occurrence of the fact.

4. *Experimental research:* The experimental research also looks for a cause-effect relationship between two or more variables. But this relationship is studied under the controlled condition which is not the case in causal-comparative research. Various types of experimental designs are used in conducting experimental research. The selection and use of a particular experimental design depends upon the nature of problem and its objectives.

Advantages of Quantitative Research

Quantitative research has played a significant role in conducting educational studies because of the following advantages:

1. The results are statistically reliable. Quantitative research mostly uses statistical methods in drawing comparison between concepts, ideas, products, packages, etc.

2. Quantitative research involves quantifications based on numbers. Thus, it is well-suited to addressing the 'who', 'what', 'when' and 'where' of individual (consumer) behaviour.

3. The results of quantitative research can be generalized. The findings can be projected to the whole population.

4. The use of multivariate methods and analysis is helpful in measuring and controlling the variable or variables which intervene between the independent and dependent variables.

Limitations of Quantitative Research

Despite a number of advantages of the quantitative research, there are also many disadvantages. Some of those are mentioned as under:

1. Quantitative research uses quantitative data. These data are closed-ended and hence do not provide depth and detail.

2. Advance formulation of specific hypotheses is an important requirement of quantitative research. In certain research contexts, especially in the field of education, it is not always possible to formulate specific hypothetical generalizations.

3. The occurrence of an event is both quantitative and qualitative. Hence, its measurement on the basis of selective observation and selective recording of information may involve researcher 'bias'.

4. The primary disadvantage of quantitative research is that issues are only measured if they are known prior to the beginning of the study especially in the survey research.

5. Quantitative research is neither appropriate nor cost effective for studying why people act or think as they do. In such cases it is difficult to quantify the action or thinking of people. The selection and use of large samples for drawing reliable statistics also involve lot of efforts in terms of money and manpower.

1.4.2 Phenomenological Inquiry: Qualitative Research

Phenomenological inquiry uses qualitative approach to the verification of proposition, which takes into consideration the totality of a phenomenon and does not attempt at analysing it into quantifiable (measurable) components. It employs a naturalistic approach based on phenomenological paradigm, which uses a variety of interpretive research methodologies, that seeks to understand phenomenon in context-specific settings. Qualitative research, in contrast to quantitative approach, is by some regarded as less 'scientific' and 'softer'. It describes social phenomena as they occur naturally. No attempt is made to manipulate the situation under study. Qualitative research emphasises: (*i*) an 'emic' perspective viewing events, actions, values, beliefs, etc. from the point of view of people who are being studied (phenomenological), (*ii*) detailed perspectives of the participants in the 'naturalistic' setting, (*iii*) contexualising behaviour, events, etc. within a holistic frame, (*iv*) an inductive, open and flexible approach, and (*v*) a definite preference for 'theory generation' rather than theory testing during the research process.

Characteristics of Qualitative Research

Qualitative research has some characteristics which distinguish it from quantitative research. Best and Kahn (2002, pp. 184-185) has quoted ten themes proposed by Patton (1990, pp. 40-41) which highlight the following main characteristics of qualitative research:

1. Qualitative research makes use of naturalistic inquiry. It aims at studying real world situations as they unfold naturally without any manipulation and pre-determined constraints on outcomes.

2. It employs inductive or 'bottom-up' approach. The researcher generates new hypotheses and grounded theory from data collected during field work. It aims to discover important categories, dimensions, and interrelationships. In the process of induction the researcher begins by exploring genuinely open questions rather

than testing theoretically derived (deductive) hypotheses. The data are used to develop concepts and theories that help the researcher to understand the phenomenon.

3. Most of the common research objectives in qualitative research aim at description, exploration, and discovery using 'wide-angle' and 'deep-angle' lens approach so as to examine the breadth and depth of phenomenon and to learn more about it. The whole phenomenon under study is 'understood as a complex system that is more than the sum of its parts; focus on complex interdependencies not meaningfully reduced to a few discrete variables and linear, cause-effect relationships'.

4. The behaviour of the subjects under study is assumed to be fluid, dynamic, situational, social, contextual and personal. The behaviour is studied in the natural environments not under the controlled conditions.

5. Qualitative research makes use of qualitative data which are gathered from natural settings. These include 'detailed, thick description; inquiry in depth; direct quotations capturing people's personal perspectives and experiences'. The researcher attempts to observe and describe the settings as they are, maintaining what Patton calls 'emphatic neutrality'. The total emphasis is on understanding of the situation in all its complexities by not proving something, not advocating, not advancing personal opinions and views, but researcher includes his personal experiences and emphatic insight as part of the relevant data while taking a 'neutral nonjudgemental stance toward whatever content may emerge'.

6. Purposive sampling is the dominant strategy in qualitative research. The researcher uses small samples. He has direct contact with and gets close to the people, situation, and phenomenon under study. He collects qualitative data using in-depth interviews, participant observation, field notes, and open-ended questions. The researcher is the primary data collection instrument, 'researcher's personal experiences and insights are an important part of inquiry and critical to understanding the phenomenon'. The data are in the form of words, images, and categories.

7. Qualitative research emphasises 'unique case orientation'. It assumes each case is special and unique. Cross-case analysis follows from and depends on the quality of individual case studies.

8. The analysis of qualitative data requires organizing raw data into logical, meaningful categories, and examining them in holistic fashion for interpretation to others. The reports are narrative with contextual description and direct quotations from research participants.

The characteristics described above indicate that qualitative research is not one single method or strategy for research but a wide range of discrete strategies and methods. These strategies normally have one thing in common that they analyse complex and unique data through exploration. Qualitative research is concerned with the opinions, experiences and feelings of individuals producing subjective data. It is a 'kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification' (Strauss and Corbin, 1990, p. 17). Whereas quantitative research seeks causal determination, prediction, and generalization of findings, qualitative research seeks understanding, exptrapolation and explanation to similar situations.

Major Types of Qualitative Research

The variety in types of qualitative research is directly based on the varieties in the methods as well as varieties in the sources of data. For example, a researcher may focus on different sources of data like:

1. One's own immediate experience;
2. Others' experiences, which one might seek to understand through:
   (*i*) their speaking or writing,
   (*ii*) their other behaviours,
   (*iii*) their other products: technology, artwork, etc.
3. Collecting data which concern past, present or future:
   (*i*) collecting things that are the results of past, like artifacts or literature,
   (*ii*) observing or introspecting what is happening now,
   (*iii*) eliciting data, making things to happen, as in interview or a project.

The questions and problems which most often come from real-world observations, contents or situations also provide a number of different ways to view the theoretical perspectives of various types of qualitative research. Patton (1990, p. 88) has suggested a list of ten theoretical perspectives of qualitative research alongwith their disciplinary roots and the questions they pose. These include: (1) Ethnography; (2) Phenomenology; (3) Heuristics; (4) Ethnomethodology; (5) Symbolic interactionism; (6) Ecological psychology; (7) Systems theory; (8) Chaos theory: nonlinear dynamics; (9) Hermeneutics; and (10) Orientational, qualitative.

Advantages of Qualitative Research

Qualitative research based upon the phenomenological paradigm has the following advantages:

1. Qualitative research is not a unitary approach. In reality, it is a variety of alternative approaches to the traditional, positivistic research. Thus, this approach has the advantage of studying a phenomenon in a holistic perspective.
2. It utilises qualitative data which are detailed and descriptive. These data indicate what people have said in their own words about their experiences and interactions in natural setting, and after careful analysis, the data provide useful and depth answers to the research questions of decision makers and information users.
3. Qualitative research is most suitable in the study of human behaviour which is fluid, dynamic, situational, social, contextual and personal.
4. It does not start with the advance formulation of specific deductive hypotheses as is the case with quantitative research. The researcher uses inductive analysis for generating new hypotheses from the details and specifics of the data during field work. He/she begins by exploring genuinely open questions rather than testing theoretically derived deductive hypotheses.
5. Qualitative research utilises flexible design and 'avoids getting locked into rigid designs.' Design flexibility permits the researcher to adjust the direction of the research process to the selection and use of tools as well as samples.
6. The researcher has direct and close contact to the people, situation, and phenomenon under study which are helpful in understanding a phenomenon in depth and detail.

7. Qualitative research is cost effective. It uses case study, small and purposive sampling strategies in collecting detailed information about the phenomenon.

8. The final reports of qualitative research studies are detailed and interesting narrations about the phenomenon.

Limitations of Qualitative Research

Qualitative research has some limitations also which are mentioned as under:

1. Subjective bias is a constant threat to objective data gathering tools and analysis techniques. For example, an individual may intentionally attempt to exhibit an artificial behaviour when he/she knows that he is being observed during observation. Similarly during an interview, the interviewees may not respond freely, frankly and accurately. There is a constant danger of subjectivity on the part of an observer/ interviewer during observation/interview.

2. The findings of qualitative research lack generalizations because of the nature and size of samples used for data collection. The samples are small in size and mostly purposive. Pure subjectivity in the selection of such samples undermines their credibility.

3. Qualitative research utilizes a variety of methodologies in studying a phenomenon in holistic perspective. In certain cases, it is difficult to focus on complex interdependencies of its parts and understand the meaning of the phenomenon as a whole.

In view of the above discussion on the nature, advantages and limitations of the quantitative (logical positivism) and qualitative (phenomenological inquiry) research paradigms, no particular paradigm can claim to be the sole and appropriate approach which may be used for conducting educational research. Education as a discipline has a wide base with diverse concerns, thus many of its problems can certainly be meaningfully investigated by means of different approaches. Moreover, with increased diversification in the context of education, there is a need for adopting multi-method approach, involving both qualitative and quantitative paradigms, to the methodology of educational studies.

1.5 SUMMARY

- Research that attempts to measure, describe, explain and predict the social and economic phenomena or social behaviour of human beings is known as 'social research'.

- The outcomes of social research provide the following benefits: helps professionals in earning their livelihood, helps students in knowing how to write a report for various findings, helps philosophers to think on wider new perspectives and helps in developing new styles for creative work.

- Social research helps the government to explore the following things: social and economic structures, social attitudes, social values and behaviours and factors motivating individuals and groups of a society.

- The subject matter of sociology is society. Sociologists study man's social behaviour in a variety of contexts. They use a number of methods in social research including comparative method, participant observer method, community studies, etc.

Check Your Progress

6. Which arithmetical operation is applicable to nominal scales?

7. Why can't we legitimately use the operations of multiplication and division with interval scales?

8. Why is a causal-comparative research called ex-post-facto research?

- For any study undertaken to be called a 'research', it should adhere to the following three criteria: a set of philosophies guide the research, methods, techniques and procedures which proven reliability and validity are used, research has to be objective as well as unbiased.

- The application of scientific methods practised in natural sciences like physics and chemistry in researching various areas in social sciences is known as the positivist approach.

- The approach in which in order to explain human behaviour, social researchers need to be conversant with people's interpretations of social phenomena is known as interpetivism.

- Objectivity is the capacity to represent truthfully and without prejudice, the result of one's research. A social research needs to be aware of his/her personal biases and prejudices and take adequate care that these do not affect the objectivity of the research.

- The following are the different types of social research: survey and fieldwork. Fieldwork further has different classifications including interview, participatory research, experiment and working with available information.

- Major steps in social research include: formulating the research problem, extensive literature survey, development of a working hypothesis, preparing the research design, determining sample design, collecting data, execution, analysis, hypothesis testing, generalization and interpretation and preparation of the report or the thesis.

- A good research follows logical methods, is systematic, and structured in accordance with well-defined sets of rules and practices to enable the researcher in arriving at dependable conclusions.

- Problems encountered by researchers in India include dearth of tools, infrastructural problems, lack of scientific training, insufficient interaction and difficulty in getting sponsors, etc.

- Logical positivism uses quantitative approach in the verification of theoretical propositions. Quantitative approach uses 'measurement' as the most precise and universally accepted method for assigning quantitative values to the characteristics or properties of objects or events for the purpose of discovering relationships between variables under study.

- Major types of quantitative research include: descriptive survey research, correlation research, causal-comparative research and experimental research.

- Phenomenological inquiry uses qualitative approach to the verification of proposition which takes into consideration the totality of a phenomenon and does not attempt at analysing it into quantifiable components.

1.6 KEY TERMS

- **Social research:** It refers to that research which attempts to measure, describe, explain and predict the social and economic phenomena or social behaviour of human beings.

- **Quantitative approach:** It is an approach in social research where measurement is used the most precise and universally accepted method for assigning quantitative

values to the characteristics or properties of objects or events for the purpose of discovering relationships between variables under study.

- **Ex-post facto:** It is a term used to define an action taken to change the effect given to a set of circumstances.

- **Positivist approach:** The application of scientific methods practiced in natural sciences like physics and chemistry in researching various areas in social sciences is known as the positivist approach.

- **Interpretivism:** In order to explain human behaviour, social researchers need to be conversant with people's interpretations of social phenomena. This approach is known as interpretivism.

1.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Social scientists are often influenced by their biases, passions, likes and dislikes and preconceived notions. These are seen to interfere with the scientific objectivity that they would need while researching on social sciences.

2. Social research is very helpful for a country as it helps the government to explore the following things:
   - Social and economic structures
   - Social attitudes
   - Social values and behaviours
   - Factors motivating individuals and groups of a society

3. A population universe is the group of people about whom generalization is to be made in a survey.

4. A researcher considers the objective of the research, the nature of the investigation, time and financial resources available, scope of the inquiry, and the desired degree of accuracy for selecting the most appropriate method of collecting the data.

5. The preliminary pages of the report contain the title, the date, acknowledgements, foreword, table of contents, list of tables, list of graphs and charts (if any).

6. The only arithmetical operation applicable to nominal scales is counting, the mere enumeration of individuals in each category, class or set.

7. We cannot use the operations of multiplication and division with interval scales legitimately since these operations presuppose the existence of an exact zero point.

8. A causal-comparative research is also called ex-post-facto research because in this type of research, the researcher usually does not have control over the casual factor or independent variable as it is studied after the occurrence of the fact.

1.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is social research?
2. What are the qualities of good research?

3. How can the qualitative approach to research be classified?

4. What is the similarity between experimental research and ex-post facto research?

5. List the nature and scope of a good and effective research.

6. Describe the positivist and interpretivist approach to research.

7. What is a fieldwork? Describe the various methods of fieldwork.

8. State the problems faced by researchers in India.

9. Briefly list the characteristics of qualitative research.

Long-Answer Questions

1. Evaluate the characteristics of good research.

2. Critically analyse the aims of social research.

3. 'Appelbaum and Chambliss (1997:40) hold that the principal methods of social research include survey and fieldwork.' With regard to this statement, assess the two types of social research.

4. Describe the major steps involved in social research.

5. Discuss the characteristics of quantitative research with its major types.

6. What are the advantages and limitations of quantitative research?

7. Explain the advantages and disadvantages of qualitative research.

1.9 FURTHER READING

Chawla, D. and N. Sondhi. 2011. *Research Methodology*. New Delhi: Vikas Publishing House.

Kothari, C. R. 2008. *Research Methodology*. New Delhi: New Age International Publishers.

Kumar, B. 2006. *Research Methodology*. New Delhi: Excel Books.

Paneerselvam, R. 2009. *Research Methodology*. New Delhi: Prentice Hall of India.

Payne, Geoff and Judy Payne. 2004. *Key Concepts in Social Research*. London: SAGE Publications.

Kumar, C. Rajender. 2008. *Research Methodology*. Delhi: APH Publishing Corporation.

Kumar, Ranjit. 2011. *Research Methodology: A Step-by-Step Guide for Beginners*. Australia: SAGE.

Gupta, D. 2011. *Research Methodology*. New Delhi: PHI Learning Private Limited.

UNIT 2 HYPOTHESIS

Structure

2.0 INTRODUCTION

A hypothesis is an assumption or a statement that may or may not be true. The hypothesis is tested on the basis of information obtained from a sample. A well formulated or good hypothesis helps the researchers to focus/concentrate on the key points of investigation. A hypothesis is significant because it guides the research. The researchers or investigators refer to the hypothesis in order to direct their thought processes toward the result of the research problem or sub-problems. There are several reasons why hypothesis is significant and these will be dealt in the unit.

It is important that the hypothesis formulated guides the researcher in the right direction. For this, it is important that one knowns the characteristics of valid research. It is only after this that the steps in the formulation of a hypothesis can be understood well. Further, knowing the source of hypothesis is also crucial for the researcher. You will also learn about the hypothesis testing. First step is to establish the hypothesis to be tested. The next step in the testing of hypotheses exercise is to choose a suitable level of significance. The level of significance denoted by a is chosen before drawing any sample.

2.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Explain the meaning of hypothesis
- Discuss the characteristics of valid hypothesis
- Describe the steps in the formulation of a hypothesis
- Identify the importance of hypothesis in social research
- Discuss the sources of hypothesis
- Explain the steps in the hypothesis testing exercise

2.2 CONCEPTUALISATION AND FORMULATION OF HYPOTHESIS

A hypothesis is an approximate assumption that a researcher wants to test for its logical or empirical consequences. It can contain either a suggested explanation for a phenomenon or a proposal having deductive reasoning to suggest a possible interrelation between multiple phenomena. A deductive reasoning can be defined as a type of reasoning that can be derived from previously known facts.

Some definitions of hypothesis are:

- According to Townsend, 'Hypothesis is defined as suggested answer to a problem.'

- According to McGuigan, 'A hypothesis is a testable statement of a potential relationship between two or more variables.'

- According to Uma Sekaran, 'A hypothesis is defined as a logically conjectured relationship between two or more variables in the form of testable statement. These relationships are based on theoretical framework formulated for the research problem. The hypotheses are often statements about population parameters like expected value and variance, for example a hypothesis might be that the expected value of the height of 10-year-old boys in the Scottish population is not different from that of 10-year-old girls.'

- According to Kerlinger, 'A good hypothesis is one which satisfies the following criteria:

  (i) Hypothesis should state the relationship between variables.

  (ii) They must carry clear implications for testing the stated relations.'

This means that: (a) statements contain two or more variables which can be measured, (b) they must state clearly how the two or more variables are related, and (c) it is important to note that facts and variables are not tested but relations between variables exist.

Characteristics of Valid Hypothesis

There are several characteristics of hypothesis, which are as follows:

- **Conceptually clear and accurate:** The hypothesis must be conceptually clear. The concepts and variables should be clearly defined operationally. The definition should use terms which are commonly accepted and it should ensure that communication is not hindered. Hypothesis should be clear and accurate so as to draw a consistent conclusion.

- **Statement of relationship between variables:** If a hypothesis is relational, it should state the relationship between the different variables.

- **Testability:** A hypothesis should have empirical referents which means that it should be testable through the empirical data. Hypothesis involving mystical or supernatural arenas are impossible to test. For example, the hypothesis 'education brings all-round development' is difficult to test because it is not easy to operationally isolate the other factors that might contribute towards all-round development. Since a hypothesis predicts the outcome of a study and it must relate variables

that are capable of being measured. The hypothesis such as 'there is a positive relationship between the learning style and academic achievement of 8th grade students' can be tested since the variables in the hypothesis are operationally defined, and therefore can be measured.

- **Specific with limited scope:** A hypothesis, which is specific with limited scope, is easily testable than a hypothesis with limitless scope. Therefore, a researcher should give more time to conduct research on such a kind of hypothesis.

- **Simplicity:** A hypothesis should be stated in simple and clear terms to make it understandable.

- **Consistency:** A hypothesis should be reliable and consistent with established and known facts.

- **Time limit:** A hypothesis should be capable of being tested within a reasonable amount of time. In other words, the excellence of a hypothesis is judged by the time taken to collect the data needed for the test.

- **Empirical reference:** A hypothesis should explain or support all the sufficient facts needed to understand what the problem is all about.

A few more characteristics of a good hypothesis are as follows:

- It ensures that the sample is readily approachable.

- It maintains a very apparent distinction with what is called theory, law, facts, assumptions and postulates.

- It ideally has logical simplicity, large number of consequences and is expressed in quantified form.

- It displays equal chances of confirmation and rejection.

- It permits the application of deduction reasoning.

- Its tools and data are easily available and effectively used.

- It is based on the study of previous literature and an existing theory, and verifiable.

As soon as a research question is formulated, it makes the hypothesis formulation imperative since a *hypothesis* is a tentative solution or an intelligent guess about a research question under study. It is an assumption or proposition whose tenability is to be tested on the basis of its implications with empirical evidence and previous knowledge. Modern investigators agree that, whenever possible, research should proceed from a hypothesis. In the words of Van Dalen (1973), 'a hypothesis serves as a powerful beacon that lights the way for the research worker'.

2.2.1 Formulating a Hypothesis

As per the Concise Oxford Dictionary (1990) hypothesis is 'A proposition made as a basis for reasoning, without the assumption of its truth, a supposition made as a starting point for further investigation from known facts'.

Leedy and Ormrod (2001) have defined the term hypothesis as 'A hypothesis is a logical supposition, a reasonable guess, an educated conjecture. It provides a tentative explanation for a phenomenon under investigation'.

There is no certainty that the hypothesis formulated for a problem is true or correct. Formulated hypothesis is the initial point, a statement that the researcher has to

prove true after further research and investigations. It is also possible that after further research the researcher might find that this hypothesis is not valid for the problem and that it needs modifications.

Why is a Hypothesis Required?

A well formulated or good hypothesis helps the researchers to focus/concentrate on the key points of investigation. Also a hypothesis is significant because it guides the research. The researchers or investigators refer to the hypothesis in order to direct their thought processes toward the result of the research problem or sub-problems. The hypothesis also helps the researcher or investigator to collect the right, precise and accurate data required for the research or investigation. As per Leedy and Ormrod (2001), 'Hypotheses are exceptionally essential and significant because they help an investigator or researcher to locate information needed to resolve the research problem or sub-problems'.

Accepting or Rejecting Hypothesis

'A hypothesis is never proved or disproved. In fact, an investigator or researcher who sets out to prove a hypothesis would lose the impartiality of the research investigation' (Leedy and Ormrod, 2001).

In research, an investigator or researcher is proficient to either accept (support) or reject a hypothesis. If a hypothesis is rejected, it will lead an investigator or researcher to develop new hypothesis to explain the phenomenon in question. If a hypothesis is continually supported or accepted, then it may evolve into a '**Theory**' (Leedy and Ormrod, 2001).

Therefore, when a hypothesis is continually accepted or supported over time by a growing body of data, then it becomes a **theory**. As per Leedy and Ormrod (2001), 'A theory is an organized body of concepts and principles intended to explain a particular phenomenon'. A theory is similar to a hypothesis in that it presents a tentative explanation for a phenomenon that new data will either support or not support. Both are supported or rejected based on the testing performed by various investigators or researchers under different conditions. An example of a well known theory is 'Einstein's Theory of Relativity' (Leedy and Ormrod, 2001).

Further, a theory that is continually validated over time by a growing body of data becomes a '**Law**'. An example of a well known law is the 'Law of Gravity' (Leedy and Ormrod, 2001).

Steps in Hypothesis Generation

Often times, an investigator will formulate a hypothesis based on the problem or sub-problems of the research. Typically, the hypothesis is driven by the research question (Leedy and Ormrod, 2001). The following steps helps in generating or writing an effective hypothesis:

- **Step One: Preliminary Research:** The researcher or investigator must review the information collected up to now and then decide which information is significant for the research and how it will help to develop the hypothesis.
- **Step Two: Write Your Hypothesis:** The hypothesis is a statement that the researcher or investigator intend to prove through the research. It should state or

affirm the focus of research. When the final hypothesis is written, verify it to be certain that it has the following criteria:

1. It is written in the form of a concise statement.
2. It reflects a situation specified by the researcher or investigator.
3. It is arguable and a contrary situation can be taken.
4. It requires research to determine whether or not it is true.
5. It is a significant theme to social scientists.
6. It is a complex notion, dealing with a number of variables.
7. It is not written in the first person.
8. It can be tested.

- **Step Three: Test It Against the Criteria in Step Two:** Take the hypothesis and verify to perceive if it has the criteria listed in Step Two. If it is not so, then the researcher or investigator has to again check the formulated hypothesis for the research problem. It should be reworked such that it fits well with the research assumptions. Also the researcher or investigator has to be certain that they are diverting from the focus of research.

Formulating the Hypothesis and Research Question

Formulating a hypothesis helps by defining an initial explanation to be tested in the research process. The following are the essential key points that must be defined at the time of hypothesis formulation:

- Hypotheses are testable explanations of a problem, phenomenon or observation.
- Both quantitative and qualitative research involves formulating a hypothesis to address the research problem.
- Hypotheses that suggest a causal relationship involve at least one independent variable and at least one dependent variable; in other words, one variable which is presumed to affect the other.
- An independent variable is one whose value is manipulated by the researcher or investigator.
- A dependent variable is a variable whose values are presumed to change as a result of changes in the independent variable.

In an equation, a dependent variable is the variable whose value depends on one or more variables in the equation. An independent variable in an equation is any variable whose value is not dependent on any other variable in the equation. Hypothesis is a tentative assumption explaining an observation, phenomenon or scientific problem that can be tested by further observation, investigation or experimentation. Characteristically, the research is a process of investigation of a particular/specific topic of study with the aim of studying a problem or question. The research topic for study is established by the researcher o investigator according to the specific assignment that needs to be explored.

After the final section of topic, the researcher or investigator has to develop a question for research and hypothesis that relates to the research being conducted. The formulation of a research question must be made before the researcher initiates conducting research on specified topic. This will be a question developed from the purpose statement and will be the specification that the researcher intents to find out by conducting the research. The question selected will guide the researcher or investigator through their

research process and will also concentrate on the objective of the research. As already discussed, the hypothesis is a prediction regarding the outcome of the research being conducted. The key objective for the researcher or investigator in developing a hypothesis statement is to test and ultimately accept or reject it when the assessment of the research is performed.

A well researched and planned research question will help and ensure the researcher or investigator that they are collecting the appropriate data. This is a critical and most significant step in the research process. The research question determines *what, where, when* and *how* the data are collected as it is an important association between the abstract, theoretical, conceptual and logistic aspects of the research plan.

2.3 IMPORTANCE AND SOURCE OF HYPOTHESIS IN SOCIAL RESEARCH

In this section, we will have a look at the significance of hypothesis in social research along with the sources of hypothesis.

2.3.1 Importance of Hypothesis in Social Research

The reasons for formulating a hypothesis are as follows:

- A hypothesis directs, monitors and controls the research efforts. It provides tentative explanations of facts and phenomena and can be tested and validated. Such explanations, if held valid, lead to generalizations, which help significantly in understanding a problem. They, thereby, extend the existing knowledge in the area to which they pertain and thus help in theory building and facilitate the extension of knowledge in an area.

- The hypothesis not only indicates what to look for in an investigation but also how to select a sample, choose the design of research, how to collect data and how to interpret the results to draw valid conclusions.

- The hypothesis orients the researcher to be more sensitive to certain relevant aspects of the problem so as to focus on specific issues and pertinent facts. It helps researchers to delimit their study in scope so that it does not become broad and unwieldy.

- The hypothesis provides rational statements to the researcher, consisting of elements expressed in a logical order of relationships, which seek to describe or explain conditions or events that have not been confirmed by facts. Some relationships between elements or variables in a hypotheses are known facts, and others transcend the known facts to give reasonable explanations for known conditions. Hypothesis help researchers to relate logically known facts to intelligent guesses about unknown conditions (Ary, *et al*., 1972, pp. 73–74).

- Hypothesis formulation and its testing add a scientific rigour to all types of researches. A well thought set of hypothesis places a clear and specific goal before the researcher and equips him/her with understanding. It provides the basis for reporting the conclusions of the study on the basis of these conclusions. Researchers can make their research report interesting and meaningful to the reader. The importance of a hypothesis is generally recognized more in the studies which aim to make predictions about some outcome. In an experimental study, the researcher is interested in making

predictions about the expected outcomes and hence the hypothesis takes on a critical role. In the case of historical or descriptive studies, however, the researcher investigates the history of an event, or the life of a man, or seeks facts in order to determine the *status quo* of a situation and hence may not have a basis for making a prediction of the results. In studies of this nature, where finding facts is the objective of the study, a hypothesis may not be required.

Most historical or descriptive studies involve fact finding as well as the interpretation of facts in order to draw generalizations. For all such major studies, including social research a hypothesis is recommended so as to explain observed facts, conditions or behaviour and to serve as a guide in the research process. If a hypothesis is not formulated, researchers may waste time and energy in gathering extensive empirical data, and then find that they cannot state facts clearly and detect relevant relationships between variables as there is no hypothesis to guide them.

2.3.2 Sources of Hypothesis

Since the mind is fed by innumerable streams and sources, it is difficult to pinpoint how a particular good idea comes to a researcher. The following are some of the popularly known sources of research hypothesis:

- **Scientific theories:** A systematic review and analysis of theories developed in the field of psychology, sociology, economics, political science and biological science may provide the researcher with potential clues for constructing a good and testable hypothesis.

- **Expert opinions:** Discussion with the experts in the field of research may further help the researcher obtain necessary insight and skill into the problem and in the formulation of a hypothesis.

- **Method of related difference:** When we find that two phenomena differ constantly and the other circumstances remain the same, we suspect a causal connection. For example, when we find uncontrolled traffic in a locality, resulting in a greater number of road accidents, we suspect a causal connection between uncontrolled traffic and road accidents. This method also suggests hypothesis.

- **Intellectual equipment of researcher:** Intellectual abilities of a researcher like creative thinking and problem solving techniques are very helpful in the formulation of a good hypothesis.

- **Related literature:** Related literature is the most important source of hypothesis formulation. A review of this literature may reveal to the researcher the variables that have been considered important in relation to his/her problem, which aspects have already been studied and which are left to be studied, which theories have supported the relationships and which theories present a contradictory relationship. Familiarity with related literature may give the researcher a tremendous advantage in the construction of hypothesis.

- **Experience:** One's own experience may be a rich source of hypothesis generation. Personal experiences of an individual which has been gained through reading biographies, autobiographies, newspaper readings or through informal talks among friends, etc., can be a potential source of generation of a hypothesis. For example, a researcher who is working on the effectiveness of guidance in teaching, can think of factors such as the teacher's polite behaviour, techniques of counselling, mastery over the subject, effective use of teaching skills, decision-

making capability, perception of his/her competence, perception of student's capacity for better interaction, use of communication skills, etc.

- **Analogies:** Several hypotheses in a branch of knowledge may be made by using analogies from other sciences. Models and theories developed in a discipline may help, through extrapolation, in the formulation of hypothesis in another discipline. By comparing the two situations, analysing their similarities and differences, some rationale may emerge in the mind of the researcher which may take the form of a hypothesis for testing. For example, in a research problem like studying the factors of unrest among college level students, the researcher insightfully thinks: 'Why was unrest found among school students?' and 'What has changed them: quality of teaching or quality of leadership?'

  Arguing analogically in this way may lead the investigator to some conclusions which may be used for identifying variables and relationships, which form the basis of hypothesis construction. If a researcher knows from previous experience that the old situation is related to other factors Y and Z as well as to X, he/she may reason out that the new situation may also be related to Y and Z.

- **Methods of residues:** When the greater part of a complex phenomenon is explained by some causes already known, we try to explain the residual part of phenomenon according to the known law of operation. It also provides possible hypothesis.

- **Induction by simple enumeration:** Sometimes scientists take common experience as a starting point of their investigation. For example, after observing a large number of scarlet flowers that are devoid of fragrance, we frame a hypothesis that all scarlet flowers are devoid of fragrance. Thus, induction by simple enumeration is a source of discovery.

- **Formulation of hypothesis:** It may also originate from the need and practice of present times.

- **Existing empirical uniformities:** In terms of common sense proposition, the existing empirical uniformities may form the basis for scientific examination.

- **A study of general culture:** It is also a good source of hypothesis.

- **Suggestions:** When given by other researchers in their reports, suggestions are quite helpful in the establishment of hypothesis for future studies.

2.4 TESTING OF HYPOTHESES

Till now, we have learnt about the formulation and sources of hypothesis, let us now more towards the testing of hypothesis. The testing of hypothesis requires an understanding of some important concepts.

Below are discussed some concepts on testing of hypotheses:

- **Null hypothesis:** The hypotheses that are proposed with the intent of receiving a rejection for them are called null hypotheses. This requires that we hypothesize the opposite of what is desired to be proved. For example, if we want to show that sales and advertisement expenditure are related, we formulate the null hypothesis that they are not related. If we want to prove that the average wages of skilled workers in town 1 is greater than that of town 2, we formulate the null hypotheses that there is no difference in the average wages of the skilled workers in both the towns. A null hypothesis is denoted by H_0.

- **Alternative hypotheses:** Rejection of null hypotheses leads to the acceptance of alternative hypotheses. The rejection of null hypothesis indicates that the relationship between variables (e.g., sales and advertisement expenditure) or the difference between means (e.g., wages of skilled workers in town 1 and town 2) or the difference between proportions have statistical significance and the acceptance of the null hypotheses indicates that these differences are due to chance. The alternative hypotheses are denoted by H_1.

- **One-tailed and two-tailed tests:** A test is called one-sided (or one-tailed) only if the null hypothesis gets rejected when a value of the test statistic falls in one specified tail of the distribution. Further, the test is called two-sided (or two-tailed) if null hypothesis gets rejected when a value of the test statistic falls in either one or the other of the two tails of its sampling distribution. For example, consider a soft drink bottling plant which dispenses soft drinks in bottles of 300 ml capacity. The bottling is done through an automatic plant. An overfilling of bottle (liquid content more than 300 ml) means a huge loss to the company given the large volume of sales. An underfilling means the customers are getting less than 300 ml of the drink when they are paying for 300 ml. This could bring bad reputation to the company. The company wants to avoid both overfilling and underfilling. Therefore, it would prefer to test the hypothesis whether the mean content of the bottles is different from 300 ml. This hypothesis could be written as:

$$H_0 \quad : \quad \mu = 300 \text{ ml.}$$
$$H_1 \quad : \quad \mu \neq 300 \text{ ml}$$

The hypotheses stated above are called two-tailed or two-sided hypotheses.

However, if the concern is the overfilling of bottles, it could be stated as:

$$H_0 \quad : \quad \mu = 300 \text{ ml.}$$
$$H_1 \quad : \quad \mu > 300 \text{ ml.}$$

Such hypotheses are called one-tailed or one-sided hypotheses and the researcher would be interested in the upper tail (right hand tail) of the distribution. If however, the concern is loss of reputation of the company (underfilling of the bottles), the hypothesis may be stated as:

$$H_0 \quad : \quad \mu = 300 \text{ ml.}$$
$$H_1 \quad : \quad \mu < 300 \text{ ml.}$$

The hypothesis stated above is also called one-tailed test and the researcher would be interested in the lower tail (left hand tail) of the distribution.

Type I and type II error: The acceptance or rejection of a hypothesis is based upon sample results and there is always a possibility of sample not being representative of the population. This could result in errors, as a consequence of which inferences drawn could be wrong. The situation could be depicted as given in Figure 2.1.

	Accept H_0	Reject H_0
H_0 True	Correct decision	Type I Error
H_0 False	Type II Error	Correct decision

Fig. 2.1 *Type I and Type II Errors*

If null hypothesis H_0 is true and is accepted or H_0 when false is rejected, the decision is correct in either case. However, if the hypothesis H_0 is rejected when it is actually true, the researcher is committing what is called a Type I error. The probability of committing a Type I error is denoted by alpha (α). This is termed as the level of significance. Similarly, if the null hypothesis H_0 when false is accepted, the researcher is committing an error called Type II error. The probability of committing a Type II error is denoted by beta (β). The expression $1 - \beta$ is called power of test. To decrease the risk of committing both types of errors, you may increase the sample size.

Steps in Testing of Hypothesis Exercise

The following steps are followed in the testing of a hypothesis:

Setting up of a hypothesis: The first step is to establish the hypothesis to be tested. As it is known, these statistical hypotheses are generally assumptions about the value of the population parameter; the hypothesis specifies a single value or a range of values for two different hypotheses rather than constructing a single hypothesis. These two hypotheses are generally referred to as (1) the null hypotheses denoted by H_0 and (2) alternative hypothesis denoted by H_1.

The null hypothesis is the hypothesis of the population parameter taking a specified value. In case of two populations, the null hypothesis is of no difference or the difference taking a specified value. The hypothesis that is different from the null hypothesis is the alternative hypothesis. If the null hypothesis H_0 is rejected based upon the sample information, the alternative hypothesis H_1 is accepted. Therefore, the two hypotheses are constructed in such a way that if one is true, the other one is false and vice versa.

Setting up of a suitable significance level: The next step is to choose a suitable level of significance. The level of significance denoted by α is chosen before drawing any sample. The level of significance denotes the probability of rejecting the null hypothesis when it is true. The value of α varies from problem to problem, but usually it is taken as either 5 per cent or 1 per cent. A 5 per cent level of significance means that there are 5 chances out of hundred that a null hypothesis will get rejected when it should be accepted. When the null hypothesis is rejected at any level of significance, the test result is said to be significant. Further, if a hypothesis is rejected at 1 per cent level, it must also be rejected at 5 per cent significance level.

Determination of a test statistic: The next step is to determine a suitable test statistic and its distribution. As would be seen later, the test statistic could be t, Z, χ^2 or F, depending upon various assumptions to be discussed later in the book.

Determination of critical region: Before a sample is drawn from the population, it is very important to specify the values of test statistic that will lead to rejection or acceptance of the null hypothesis. The one that leads to the rejection of null hypothesis is called the critical region. Given a level of significance, α, the optimal critical region for a two-tailed test consists of that $\alpha/2$ per cent area in the right hand tail of the distribution plus that $\alpha/2$ per cent in the left hand tail of the distribution where that null hypothesis is rejected.

Computing the value of test-statistic: The next step is to compute the value of the test statistic based upon a random sample of size n. Once the value of test statistic is computed, one needs to examine whether the sample results fall in the critical region or in the acceptance region.

Making decision: The hypothesis may be rejected or accepted depending upon whether the value of the test statistic falls in the rejection or the acceptance region. Management decisions are based upon the statistical decision of either rejecting or accepting the null hypothesis.

In case a hypothesis is rejected, the difference between the sample statistic and the hypothesized population parameter is considered to be significant. On the other hand, if the hypothesis is accepted, the difference between the sample statistic and the hypothesized population parameter is not regarded as significant and can be attributed to chance.

Test Statistic for Testing Hypothesis about Population Mean

In this section, we will take up the test of hypothesis about population mean in a case of single population.

One of the important things that have to be kept in mind is the use of an appropriate test statistic. In case the sample size is large ($n > 30$), Z statistic would be used. For a small sample size ($n \leq 30$), a further question regarding the knowledge of population standard deviation (σ) is asked. If the population standard deviation σ is known, a Z statistic can be used. However, if σ is unknown and is estimated using sample data, a t test with appropriate degrees of freedom is used under the assumption that the sample is drawn from a normal population. It is assumed that you have the knowledge of Z and t distribution from the course on statistics. However, these would be briefly reviewed at the appropriate place. Table 2.1 summarizes the appropriateness of the test statistic for conducting a test of hypothesis regarding the population mean.

Table 2.1 Appropriateness of Test Statistic in Testing Hypotheses about Means

Sample Size	Knowledge of Population Standard Deviation (σ)	
	Known	Not Known
Large ($n > 30$)	Z	Z
Small ($n \leq 30$)	Z	t

Tests Concerning Means-the Case of Single Population

In this section, a number of illustrations will be taken up to explain the test of hypothesis concerning mean. Two cases of large sample and small samples will be taken up.

Case of large sample

As mentioned earlier, in case the sample size n is large or small but the value of the population standard deviation is known, a Z test is appropriate. There can be alternate cases of two- tailed and one-tailed tests of hypotheses.

Corresponding to the null hypothesis $H_0 : \mu = \mu_0$, the following criteria could be used as shown in Table 2.2.

The test statistic is given by,

$$Z = \frac{\bar{X} - \mu_{H0}}{\frac{\sigma}{\sqrt{n}}}$$

Where,

$\bar{X}$ = Sample mean

σ = Population standard deviation

μ_{H0} = The value of μ under the assumption that the null hypothesis is true.

n = Size of sample.

Table 2.2 *Criteria for Accepting or Rejecting Null Hypothesis under Different Cases of Alternative Hypotheses*

S. No.	Alternative Hypothesis	Reject the Null Hypothesis if	Accept the Null Hypothesis if
1.	$\mu < \mu_0$	$Z < -Z_\alpha$	$Z \geq -Z_\alpha$
2.	$\mu > \mu_0$	$Z > Z_\alpha$	$Z \leq Z_\alpha$
3.	$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$	$-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$

If the population standard deviation σ is unknown, the sample standard

deviation $s = \sqrt{\dfrac{1}{n-1}\sum\left(X - \bar{X}\right)^2}$

is used as an estimate of σ. It may be noted that Z_α and $Z_{\alpha/2}$ are Z values such that the area to the right under the standard normal distribution is α and $\alpha/2$ respectively. Below are solved examples using the above concepts.

Example 2.1: A sample of 200 bulbs made by a company give a lifetime mean of 1540 hours with a standard deviation of 42 hours. Is it likely that the sample has been drawn from a population with a mean lifetime of 1500 hours? You may use 5 per cent level of significance.

Solution:

In the above example, the sample size is large ($n = 200$), sample mean ($\bar{X}$) equals 1540 hours and the sample standard deviation (s) is equal to 42 hours. The null and alternative hypotheses can be written as:

H_0 : $\mu = 1500$ hrs

H_1 : $\mu \neq 1500$ hrs

It is a two-tailed test with level of significance (α) to be equal to 0.05. Since n is large ($n > 30$), though population standard deviation σ is unknown, one can use Z test. The test statistics are given by:

$$Z = \frac{\bar{X} - \mu_{H0}}{\dfrac{\sigma}{X}}$$

Where, μ_{H0} = Value of μ under the assumption that the null hypothesis is true $\dfrac{\sigma}{X}$

= Estimated standard error of mean

Here $\mu_{H0} = 1,500, \dfrac{\sigma}{X} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{s}{\sqrt{n}} = \dfrac{42}{\sqrt{200}} = 2.97$

(Note that $\hat{\sigma}$ is estimated value of σ.)

$$Z = \frac{\bar{X} - \mu_{H0}}{\dfrac{s}{\sqrt{n}}} = \frac{1,540 - 1,500}{2.97} = \frac{40}{2.97} = 13.47$$

The value of $\alpha = 0.05$ and since it is a two-tailed test, the critical value Z is given by $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ which could be obtained from the standard normal table given in Appendix 1 at the end of the book.
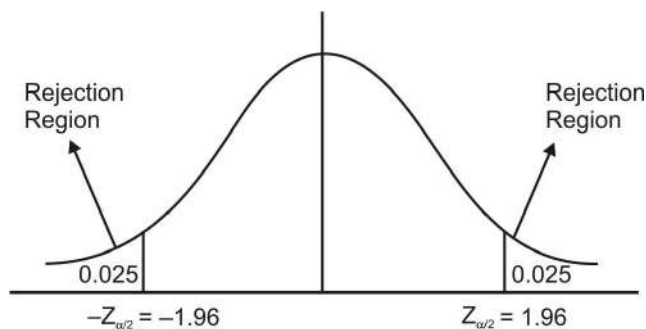
Rejection regions for Example 2.1

Since the computed value of $Z = 13.47$ lies in the rejection region, the null hypothesis is rejected. Therefore, it can be concluded that the average life of the bulb is significantly different from 1,500 hours.

Example 2.2: On a typing test, a random sample of 36 graduates of a secretarial school averaged 73.6 words with a standard deviation of 8.10 words per minute. Test an employer's claim that the school's graduates average less than 75.0 words per minute using the 5 per cent level of significance.

Solution:

$$H_0 \ : \ \mu = 75$$
$$H_1 \ : \ \mu < 75$$

$\bar{X} = 73.6$, $s = 8.10$, $n = 36$ and $\alpha = 0.05$. As the sample size is large $(n > 30)$, though population standard deviation σ is unknown, Z test is appropriate.

The test statistic is given by:

$$Z = \frac{\bar{X} - \mu_{H0}}{\hat{\sigma}_{\bar{x}}} = \frac{73.6 - 75}{1.35} = \frac{-1.4}{1.35} = -1.04$$

$$\left(\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{8.10}{\sqrt{36}} = \frac{8.10}{6} = 1.35 \right)$$

Since it is a one-tailed test and the interest is in the left hand tail of the distribution, the critical value of Z is given by $-Z_\alpha = -1.645$. Now, the computed value of Z lies in the acceptance region, and the null hypothesis is accepted as shown below:
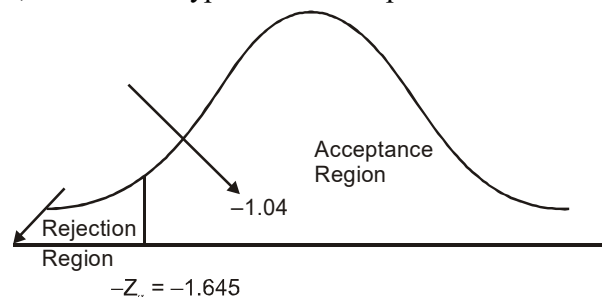
Rejection region for Example 2.2

Case of small sample

In case the sample size is small ($n \leq 30$) and is drawn from a population having a normal population with unknown standard deviation σ, a t test is used to conduct the hypothesis for the test of mean. The t distribution is a symmetrical distribution just like the normal one. However, t distribution is higher at the tail and lower at the peak. The t distribution is flatter than the normal distribution. With an increase in the sample size (and hence degrees of freedom), t distribution loses its flatness and approaches the normal distribution whenever $n > 30$. A comparative shape of t and normal distribution is given in Figure 2.2.
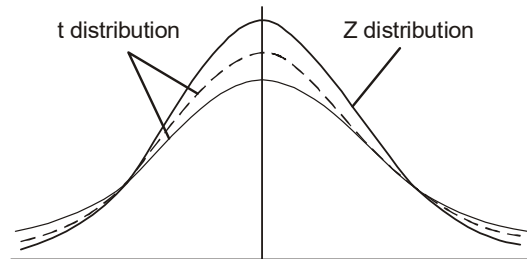
Fig. 2.2 *Shape of t and Normal Distribution*

The procedure for testing the hypothesis of a mean is similar to what is explained in the case of large sample. The test statistic used in this case is:

$$t_{n-1} = \frac{\bar{X} - \mu_{H0}}{\hat{\sigma}_{\bar{x}}}$$

Where, $\dfrac{\hat{\sigma}}{\bar{x}} = \dfrac{s}{\sqrt{n}}$ (where s = Sample standard deviation)

$n - 1$ = degrees of freedom

A few examples pertaining to 't' test are worked out for testing the hypothesis of mean in case of a small sample.

Example 2.3: Prices of share (in ₹) of a company on the different days in a month were found to be 66, 65, 69, 70, 69, 71, 70, 63, 64 and 68. Examine whether the mean price of shares in the month is different from 65. You may use 10 per cent level of significance.

Solution:

$H_0 : \mu = 65$

$H_1 : \mu \neq 65$

Since the sample size is $n = 10$, which is small, and the sample standard deviation is unknown, the appropriate test in this case would be t. First of all, we need to estimate the value of sample mean ($\bar{X}$) and the sample standard deviation (s). It is known that the sample mean and the standard deviation are given by the following formula.

$$\bar{X} = \frac{\sum X}{n} \qquad s = \sqrt{\frac{1}{n-1} \sum (X - \bar{X})^2}$$

The computation of $\bar{X}$ and s is shown in Table 10.3.

$$\sum X = 675, \quad \bar{X} = \frac{\sum X}{n} = \frac{675}{10} = 67.5$$

$$\sum\left(X - \bar{X}\right)^2 = 70.5$$

$$s^2 = \frac{1}{n-1}\sum\left(X - \bar{X}\right)^2 \frac{70.5}{9} = 7.83$$

$$s = \sqrt{7.83} = 2.80$$

Table 2.3 *Computation of Sample Mean and Standard Deviation*

S. No.	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	66	− 1.5	2.25
2	65	− 2.5	6.25
3	69	1.5	2.25
4	70	2.5	6.25
5	69	1.5	2.25
6	71	3.5	12.25
7	70	2.5	6.25
8	63	− 4.5	20.25
9	64	− 3.5	12.25
10	68	0.5	0.25
Total	675	0	70.5

The test statistic is given by:

$$t_{n-1} = \frac{\bar{X} - \mu_{H0}}{\dfrac{\hat{\sigma}}{x}} = \frac{\bar{X} - \mu_{H0}}{\dfrac{s}{\sqrt{n}}} = \frac{67.5 - 65}{\dfrac{2.8}{\sqrt{10}}} = \frac{2.5 \times \sqrt{10}}{2.8}$$

$$= 2.5 \times 3.16/2.8 = 7.91/2.8 = 2.82$$

The critical values of t with 9 degrees of freedom for a two-tailed test are given by –1.833 and 1.833. Since the computed value of *t* lies in the rejection region (see figure below), the null hypotheses is rejected.
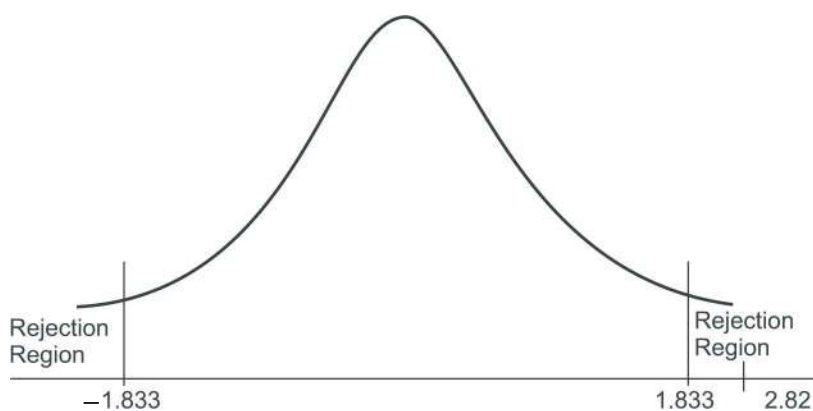
Rejection regions for Example 2.3

Therefore, the average price of the share of the company is different from 65.

Example 2.4: Past records indicate that a golfer has averaged 82 on a certain course. With a new set of clubs, he averages 7 over five rounds with a standard deviation of 2.65. Can we conclude that at 0.025 level of significance, the new club has an adverse effect on the performance?

Solution:

$$H_0 \; : \; \mu = 82$$
$$H_1 \; : \; \mu < 82$$

$\overline{X} = 7.9$, n = 5, s = 2.65, $\alpha = 0.025$. As the population standard deviation is unknown and the sample size is small ($n < 30$), a *t* test would be appropriate. The test statistic is given by:

$$t_{n-1} = \frac{\overline{X} - \mu_{H_0}}{\hat{\sigma}_{\overline{x}}} = \frac{\overline{X} - \mu_{H_0}}{s/\sqrt{n}} = \frac{7.9 - 8.2}{1.185} = \frac{-0.3}{1.185} = -0.25$$

$$\left(\hat{\sigma}_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{2.65}{\sqrt{5}} = 1.185 \right)$$

The critical value of *t* at 0.025 level of significance with four degrees of freedom is given by $-t_\alpha = -2.776$ (see Appendix 2). As the sample *t* value of –0.25 lies in the acceptance region, the null hypothesis is accepted (see figure below).
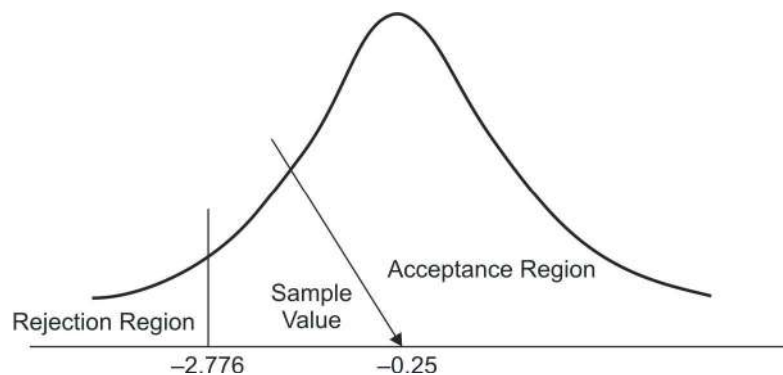
Rejection region for Example 2.4

Therefore, there is no adverse effect on the performance due to a change in the club and the performance can be attributed to chance.

Tests for Difference between two Population Means

So far, we have been concerned with the testing of means of a single population. We took up the cases of both large and small samples. It would be interesting to examine the difference between the two population means. Again, various cases would be examined as discussed below:

Case of large sample

In case both the sample sizes are greater than 30, a Z test is used. The hypothesis to be tested may be written as:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Where,

μ_1 = mean of population 1

μ_2 = mean of population 2

The above is a case of two-tailed test. The test statistic used is:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$\overline{X}_1$ = Mean of sample drawn from population 1

$\overline{X}_2$ = Mean of sample drawn from population 2

n_1 = size of sample drawn from population 1

n_2 = size of sample drawn from population 2

If σ_1 and σ_2 are unknown, their estimates given by $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are used.

$$\hat{\sigma}_1 = s_1 = \sqrt{\frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2}$$

$$\hat{\sigma}_2 = s_2 = \sqrt{\frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(X_{2i} - \overline{X}_2)^2}$$

The Z value for the problem can be computed using the above formula and compared with the table value to either accept or reject the hypothesis. Let us consider the following problem:

Example 2.5: A study is carried out to examine whether the mean hourly wages of the unskilled workers in the two cities—Ambala Cantt and Lucknow are the same. The random sample of hourly earnings in both the cities is taken and the results are presented in the Table 2.4.

Table 2.4 Survey Data on Hourly Earnings in Two Cities

City	Sample Mean Hourly Earnings	Standard Deviation of Sample	Sample Size
Ambala Cantt	₹ 8.95 ($\overline{X}_1$)	0.40 (s_1)	200 (n_1)
Lucknow	₹ 9.10 ($\overline{X}_2$)	0.60 (s_2)	175 (n_2)

Using a 5 per cent level of significance, test the hypothesis of no difference in the average wages of unskilled workers in the two cities.

Solution: We use subscripts 1 and 2 for Ambala Cantt and Lucknow respectively.

$$H_0 \;:\; \mu_1 = \mu_2 \;\to\; \mu_1 - \mu_2 = 0$$
$$H_1 \;:\; \mu_1 \neq \mu_2 \;\to\; \mu_1 - \mu_2 \neq 0$$

The following survey data is given:

$$\overline{X}_1 = 8.95,\ \overline{X}_2 = 9.10, s_1 = 0.40, s_2 = 0.60, n_1 = 200, n_2 = 175, \alpha = 0.05$$

Since both n_1, n_2 are greater than 30 and the sample standard deviations are given, a Z test would be appropriate.

The test statistic is given by:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

As σ_1, σ_2 are unknown, their estimates would be used.

$$s_1 = \hat{\sigma}_1, \qquad s_2 = \hat{\sigma}_2$$

$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{(0.4)^2}{200} + \frac{(0.6)^2}{175}} = \sqrt{0.0028} = 0.0053$$

$$Z = \frac{(8.95 - 9.10) - 0}{0.053} = -2.83$$

As the problem is of a two-tailed test, the critical values of Z at 5 per cent level of significance are given by $-Z_{\alpha/2} = -1.96$ and $Z_{\alpha/2} = 1.96$. The sample value of $Z = -2.83$ lies in the rejection region as shown in the figure below:
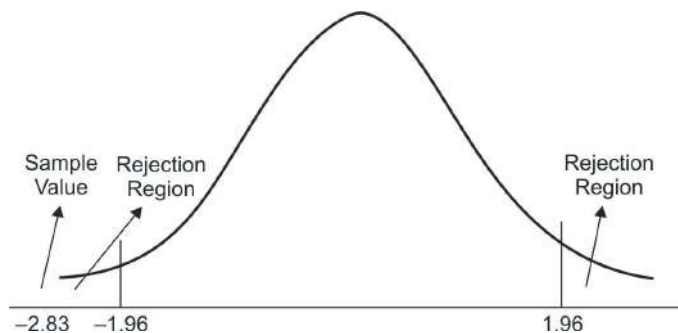
Rejection regions for Example 2.5

Case of small sample

If the size of both the samples is less than 30 and the population standard deviation is unknown, the procedure described above to discuss the equality of two population means is not applicable in the sense that a t test would be applicable under the assumptions:

(a) Two population variances are equal.

(b) Two population variances are not equal.

Population variances are equal

If the two population variances are equal, it implies that their respective unbiased estimates are also equal. In such a case, the expression becomes:

$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

(Assuming $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}^2$)

To get an estimate of $\hat{\sigma}^2$, a weighted average of s_1^2 and s_2^2 is used, where the weights are the number of degrees of freedom of each sample. The weighted average is called a 'pooled estimate' of $\hat{\sigma}^2$. This pooled estimate is given by the expression:

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

The testing procedure could be explained as under:

$H_0 : \mu_1 = \mu_2 \quad \rightarrow \quad \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 \neq \mu_2 \quad \rightarrow \quad \mu_1 - \mu_2 \neq 0$

In this case, the test statistic t is given by the expression:

$$t_{n_1 + n_2 - 2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where,

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (\mu_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Once the value of *t* statistic is computed from the sample data, it is compared with the tabulated value at a level of significance α to arrive at a decision regarding the acceptance or rejection of hypothesis. Let us work out a problem illustrating the concepts defined above.

Example 2.6: Two drugs meant to provide relief to arthritis sufferers were produced in two different laboratories. The first drug was administered to a group of 12 patients and produced an average of 8.5 hours of relief with a standard deviation of 1.8 hours. The second drug was tested on a sample of 8 patients and produced an average of 7.9 hours of relief with a standard deviation of 2.1 hours. Test the hypothesis that the first drug provides a significantly higher period of relief. You may use 5 per cent level of significance.

Solution: Let the subscripts 1 and 2 refer to drug 1 and drug 2 respectively.

$H_0 : \mu_1 = \mu_2 \quad \rightarrow \quad \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 \neq \mu_2 \quad \rightarrow \quad \mu_1 - \mu_2 \neq 0$

The following survey data is given:

$\overline{X}_1 = 8.5, \overline{X}_2 = 7.9, s_1 = 1.8, s_2 = 2.1, n_1 = 12, n_2 = 8$

As both n_1, n_2 are small and the sample standard deviations are unknown, one may use a *t* test with the degrees of freedom $= n_1 + n_2 - 2 = 12 + 8 - 2 = 18$ d.f.

The test statistics is given by:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

Where,

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(12 - 1)(1.8)^2 + (8 - 1)(2.1)^2}{12 + 8 - 2}} = \sqrt{\frac{11 \times 3.24 + 7 \times (4.41)}{18}}$$

$$= \sqrt{\frac{35.64 + 30.87}{18}} = \sqrt{\frac{66.61}{18}} = \sqrt{3.698} = 1.92$$

$$t_{18} = \frac{(8.5 - 7.9) - (0)}{1.92\sqrt{\dfrac{1}{12} + \dfrac{1}{8}}} = \frac{0.6}{1.92\sqrt{0.2083}}$$

$$= \frac{0.6}{1.92 \times 0.456} = \frac{0.6}{0.8755} = 0.685$$

The critical value of *t* with 18 degrees of freedom at 5 per cent level of significance is given by 1.734. The sample value of $t = 0.685$ lies in the acceptance region as shown in figure below:
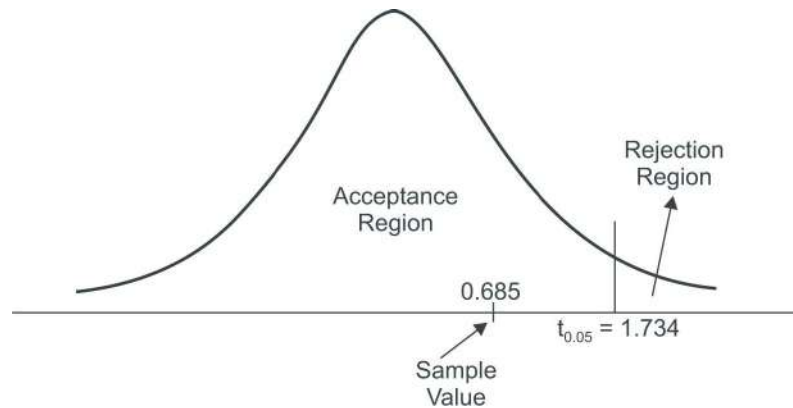
Rejection region for Example 2.6

Therefore, the null hypothesis is accepted as there is not enough evidence to reject it. Therefore, one may conclude that the first drug is not significantly more effective than the second drug.

When population variances are not equal

In case population variances are not equal, the test statistic for testing the equality of two population means when the size of samples are small is given by:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\dfrac{\hat{\sigma}_1^2}{n_1} + \dfrac{\hat{\sigma}_2^2}{n_2}}}$$

The degrees of freedom in such a case is given by the expression:

$$d.f = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)^2}$$

The procedure for testing of hypothesis remains the same as was discussed when the variances of two populations were assumed to be same. Let us consider an example to illustrate the same.

Example 2.7: There were two types of drugs (1 and 2) that were tried on some patients for reducing weight. There were 8 adults who were subjected to drug 1 and seven adults who were administered drug 2. The decrease in weight (in pounds) is given below:

Drug 1	10	8	12	14	7	15	13	11
Drug 2	12	10	7	6	12	11	12	

Do the drugs differ significantly in their effect on decreasing weight? You may use 5 per cent level of significance. Assume that the variances of two populations are not same.

Solution:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Let us compute the sample means and standard deviations of the two samples as shown in Table 2.5.

Table 2.5 Intermediate computations for sample means and standard deviations

S.No.	X_1	X_2	$(X_1 - \overline{X}_1)$	$(X_2 - \overline{X}_2)$	$(X_1 - \overline{X}_1)^2$	$(X_2 - \overline{X}_2)^2$
1	10	12	−1.25	2	1.5625	4
2	8	10	−3.25	0	10.5625	0
3	12	7	0.75	-3	0.5625	9
4	14	6	2.75	-4	7.5625	16
5	7	12	−4.25	2	18.0625	4
6	15	11	3.75	1	14.0625	1
7	13	12	1.75	2	3.0625	4
8	11		−0.25		0.0625	
Total	90	70	0	0	55.5	38
Mean	11.25	10				

$$n_1 = 8, \qquad\qquad\qquad n_2 = 7,$$

$$\overline{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{90}{8} = 11.25 \qquad\qquad \overline{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{70}{7} = 10$$

$$s_1^2 = \frac{\Sigma(X_1 - \overline{X}_1)^2}{n_1 - 1} = \frac{55.5}{7} = 7.93$$

$$s_2^2 = \frac{\Sigma(X_2 - \overline{X}_2)^2}{n_2 - 1} = \frac{38}{6} = 6.33$$

$$\hat{\sigma}_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{7.93}{8} + \frac{6.33}{7}} = \sqrt{0.99 + 0.90} = \sqrt{1.89} = 1.37$$

$$d.f. = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right) + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)} = \frac{\left(\dfrac{7.33}{8} + \dfrac{6.33}{7}\right)^2}{\dfrac{1}{7}\left(\dfrac{7.33}{8}\right)^2 + \dfrac{1}{6}\left(\dfrac{6.33}{7}\right)^2}$$

$$= \frac{3.314}{0.12 + 0.136} = \frac{3.314}{0.12 + 0.136} = 12.996 = 13 \text{ (approx.)}$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)H_0}{\sqrt{\dfrac{\hat{\sigma}_1^2}{n_1} + \dfrac{\hat{\sigma}_2^2}{n_2}}}$$

$$t = \frac{11.25 - 10}{1.37} = \frac{1.25}{1.37} = 0.912$$

The table value (critical value) of t with 13 degrees of freedom at 5 per cent level of significance is given by 2.16. As computed t is less than tabulated t, there is not enough evidence to reject H_o.

Tests Concerning Population Proportion-the Case of Single Population

We have already discussed the tests concerning population means. In the tests about proportion, one is interested in examining whether the respondents possess a particular attribute or not.

The random variable in such a case is a binary one in the sense it takes only two values—yes or no. As we know that either a student is a smoker or not, a consumer either uses a particular brand of product or not and lastly, a skilled worker may be either satisfied or not with the present job. At this stage it may be recalled that the binomial distribution is a theoretically correct distribution to use while dealing with proportions. Further, as the sample size increases, the binomial distribution approaches the normal distribution in characteristic. To be specific, whenever both np and nq (where n = number of trials, p = probability of success and q = probability of failure) are at least 5, one can use the normal distribution as a substitute for the binomial distribution.

The case of single population proportion

Suppose we want to test the hypotheses,

H_0 : $p = p_0$

H_1 : $p \neq p_0$

For large sample, the appropriate test statistic would be:

$$Z = \frac{\bar{p} - p_{H_0}}{\sigma_{\bar{p}}}$$

Where,

$\bar{p}$ = sample proportion

p_{H_0} = the value of p under the assumption that null hypothesis is true

$\sigma_{\bar{p}}$ = Standard error of sample proportion

The value of $\sigma_{\bar{p}}$ is computed by using the following formula:

$$\sigma_{\bar{p}} = \frac{\sqrt{p_{H_0} q_{H_0}}}{n}$$

Where, q_{H_0} = $1 - p_{H_0}$

n = Sample size

For a given level of significance α, the computed value of Z is compared with the corresponding critical values, i.e. $Z_{\alpha/2}$ or $-Z_{\alpha/2}$ to accept or reject the null hypothesis. We will consider a few examples to explain the testing procedure for a single population proportion.

Example 2.8: An officer of the health department claims that 60 per cent of the male population of a village comprises smokers. A random sample of 50 males showed that 35 of them were smokers. Are these sample results consistent with the claim of the health officer? Use a level of significance of 0.05.

Solution:

Sample size (n) = 50

Sample proportion = $\bar{p} = \frac{x}{n} = \frac{35}{50} = 0.70$

H_0 : $p = 0.60$

H_1 : $p > 0.60$

The test statistic is given by:

$$Z = \frac{\bar{p} - p_{H_0}}{\sigma_{\bar{p}}} = \frac{0.70 - 0.60}{0.069} = \frac{0.10}{0.069} = 1.44$$

$$\left(\sigma_{\bar{p}} = \sqrt{\frac{p_{H_0} q_{H_0}}{n}} = \sqrt{\frac{0.6 \times 0.4}{50}} = \sqrt{\frac{0.24}{50}} = 0.069 \right)$$

It is a one-tailed test. For a given level of significance $\alpha = 0.05$, the critical value of Z is given by $Z_\alpha = Z_{0.05} = 1.645$. It is seen that the sample value of $Z = 1.44$ lies in the acceptance region as shown below (see figure).
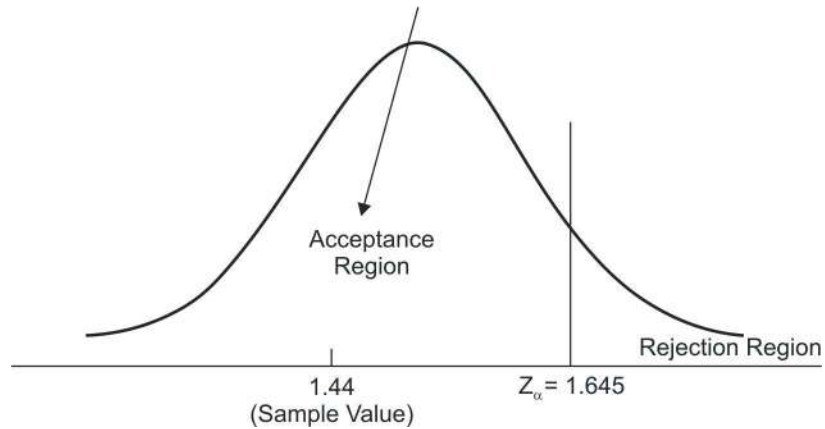
Rejection region for Example 2.8

Therefore, there is not enough evidence to reject the null hypothesis. So it can be concluded that the proportion of male smokers is not statistically different from 0.60.

Tests for Difference between two Population Proportions

Here, the interest is to test whether the two population proportions are equal or not. The hypothesis under investigation is:

$$H_0 : p_1 = p_2 \quad \rightarrow p_1 - p_2 = 0$$
$$H_1 : p_1 \neq p_2 \quad \rightarrow p_1 - p_2 \neq 0$$

The alternative hypothesis assumed is two sided. It could as well have been one sided. The test statistic is given by:

$$Z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)H_0}{\sigma_{\bar{P}_1 - \bar{P}_2}}$$

Where,

$\bar{p}_1$ = Sample proportion possessing a particular attribute from population 1

$\bar{p}_2$ = Sample proportion possessing a particular attribute from population 2

$\sigma_{\bar{P}_1 - \bar{P}_2}$ = Standard error of difference between proportions.

$(p_1 - p_2)_{H0}$ = Value of difference between population proportion under the assumption that the null hypothesis is true.

The formula for $\sigma_{\bar{P}_1 - \bar{P}_2}$ is given by:

$$\sigma_{\bar{P}_1 - \bar{P}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

We do not know the value of p_1, p_2, etc., but under the null hypothesis $p_1 = p_2 = p$.

$$\sigma_{\bar{P}_1 - \bar{P}_2} = \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

The best estimate of p is given by:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Where,

x_1 = Number of successes in sample 1

x_2 = Number of successes in sample 2

n_1 = Size of sample taken from population 1

n_2 = Size of sample taken from population 2

It is known that $\bar{p}_1 = \dfrac{x_1}{n_1}$ and $\bar{p}_2 = \dfrac{x_2}{n_2}$.

Therefore, $x_1 = n_1 \bar{p}_1$ and $x_2 = n_2 \bar{p}_2$

Therefore, $\hat{p} = \dfrac{n_1 \bar{p}_1 + n_2 \bar{p}_2}{n_1 + n_2}$

Therefore, the estimate of standard error of difference between the two proportions is given by:

$$\hat{\sigma}_{\bar{P}_1 - \bar{P}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where $\hat{p}$ is as defined above and $\hat{q} = 1 - \hat{p}$. Now, the test statistic may be rewritten as:

$$Z = \frac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)H_0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Now, for a given level of significance α, the sample Z value is compared with the critical Z value to accept or reject the null hypothesis. We consider below a few examples to illustrate the testing procedure described above.

Example 2.9: A company is interested in considering two different television advertisements for the promotion of a new product. The management believes that advertisement A is more effective than advertisement B. Two test market areas with virtually identical consumer characteristics are selected. Advertisement A is used in one area and advertisement B in the other area. In a random sample of 60 consumers who saw advertisement A, 18 tried the product. In a random sample of 100 customers who saw advertisement B, 22 tried the product. Does this indicate that advertisement A is more effective than advertisement B, if a 5 per cent level of significance is used?

Solution:

$H_0 : p_a = p_b$

$H_1 : p_a > p_b$

$n_A = 60,$ $x_A = 18,$ $n_B = 100,$ $x_B = 22$

$\left(\bar{p}_A = \dfrac{x_A}{n_A} = \dfrac{18}{60} = 0.3 \right)$ $\left(\bar{p}_B = \dfrac{x_B}{n_B} = \dfrac{22}{100} = 0.22 \right)$

$Z = \dfrac{\bar{P}_A - \bar{P}_B - (p_A - p_B) H_0}{\sigma_{\bar{P}_A - \bar{P}_B}} = \dfrac{0.3 - 0.22 - 0}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_A} + \dfrac{1}{n_B} \right)}}$

$= \dfrac{0.08}{\sqrt{0.25 \times 0.75 \left(\dfrac{1}{60} + \dfrac{1}{100} \right)}} = \dfrac{0.08}{\sqrt{0.25 \times 0.75 (0.0267)}} = \dfrac{0.08}{0.071} = 1.3$

$\left(\hat{p} = \dfrac{x_A + x_B}{n_A + n_B} = \dfrac{18 + 22}{60 + 100} = \dfrac{40}{160} = 0.25 \right)$

The critical value of Z at 5 per cent level of significance is 1.645. The sample value of $Z = 1.13$ lies in the acceptance region as shown in the figure below:
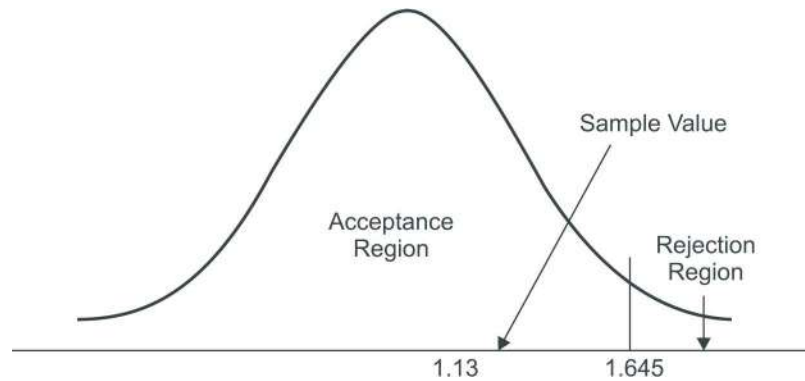
Rejection region for Example 2.9

Case Study

M L Steel Works Ltd

Mr. Mohan Lal is the proprietor of M L Steel Works Ltd., a company that manufactures and sells stainless steel utensils. Mr. Mohan Lal had set up the business in 2001. It was growing at an annual growth rate of 7 per cent and in 2008 its sales turnover was ₹75 lakh. Mr. Mohal Lal was happy with the growth of the company. However, after 2008 its sales got stagnant at ₹75 lakh. This was a matter of concern to Mr. Lal since the cost of production was going up resulting in reduced profitability.

Mr. Kapoor, the friend of Mr. Lal who was working for a consulting organization advised him to send his sales people for training. Mr. Lal had chosen 36 salesmen and sent them for a one-week training programme. After the training programme, it was noticed that the average sales for their salesmen has increased to ₹80 lakh with a standard deviation of ₹3 lakh. Mr. Lal was wondering whether it was due to chance or was it due to the effectiveness of the training programmes.

Discussion Question

Formulate a suitable hypothesis to test that training programme is effective. Test it using 5% level of significance.

(Hint: You need to test the following hypothesis.)

H_0 : $\mu = 75$

H_1 : $\mu > 75$

2.5   SUMMARY

- A hypothesis is an approximate assumption that a researcher wants to test for its logical or empirical consequences. It can contain either a suggested explanation for a phenomenon or a proposal having deductive reasoning to suggest a possible interrelation between multiple phenomena.

- The following are the characteristics of a valid hypothesis: conceptually clear and accurate, statement of relationship between variables, testability, specific with limited scope, simple, consistent, has a time limit or frame, provides empirical evidence, etc.

- There is no certainty that the hypothesis formulated for a problem is true or correct. Formulated hypothesis is the initial point, a statement that the researcher has to prove true after further research and investigations. It is also possible that after further research the researcher might find that this hypothesis is not valid for the problem and that it needs modifications.

- A well formulated or good hypothesis helps the researchers to focus/concentrate on the key points of investigation. Also a hypothesis is significant because it guides the research. The researchers or investigators refer to the hypothesis in order to direct their thought processes toward the result of the research problem or sub-problems.

- In research, an investigator or researcher is proficient to either accept (support) or reject a hypothesis. If a hypothesis is rejected, it will lead an investigator or researcher to develop new hypothesis to explain the phenomenon in question. If a hypothesis is continually supported or accepted, then it may evolve into a theory.

- Steps in hypothesis generation are: preliminary research, writing your hypothesis, and testing it against the criteria in set while writing the hypothesis.

- Hypothesis is a tentative assumption explaining an observation, phenomenon or scientific problem that can be tested by further observation, investigation or experimentation. Characteristically, the research is a process of investigation of a particular/specific topic of study with the aim of studying a problem or question.

- The key objective for the researcher or investigator in developing a hypothesis statement is to test and ultimately accept or reject it when the assessment of the research is performed.

- The importance of hypothesis in social research is that: is directs, monitors and controls research efforts; it indicates the sample selection procedure, design of research, data collection, etc.; it orients the researcher towards relevant aspects of the problem; provides rational statements to the researcher, adds a scientific rigour to the research.

Check Your Progress

7. State the condition under which a test is called one-sided.

8. Define Type-II error.

9. What is the meaning of level of significance?

- The sources of hypothesis include: scientific theories, expert opinions, method of related difference, intellectual equipment of researcher, experience, analogies, formulation of hypothesis, induction by simple enumeration, a study of general culture, etc.

- The important concepts in the testing of hypothesis includes an understanding of the topics: null hypothesis, alternate hypothesis, one-tailed and two-tailed tests, type I and type II errors.

- The following are the steps in testing of hypothesis exercise: setting up of a hypothesis, setting up of a suitable significance level, determination of a test statistic, determination of critical region, computing the value of test-statistic and making the decision.

2.6   KEY TERMS

- **Hypothesis:** It is an approximate assumption that a researcher wants to test for its logical or empirical consequences.

- **Null hypothesis:** It refers to the hypotheses that is proposed with the intent of receiving a rejection for them.

- **Critical region:** It refers to the region that leads to rejection of null hypothesis.

- **Level of significance:** It is the probability of committing a Type 1 error.

2.7   ANSWERS TO 'CHECK YOUR PROGRESS'

1. The characteristics of hypothesis are:
   - Simplicity: A hypothesis should be stated in the most simple and clear terms to make it understandable.
   - Consistency: A hypothesis should be reliable and consistent with established and known facts.

2. As soon as a research question is formulated, it makes the hypothesis formulation imperative since a hypothesis is a tentative solution or an intelligent guess about a research question under study.

3. The key objective for the researcher or investigator in developing a hypothesis statement is to test and ultimately accept or reject it when the assessment of the research is performed.

4. The hypothesis not only indicates what to look for in an investigation but also how to select a sample, choose the design of research, how to collect data and how to interpret the results to draw valid conclusions.

5. A hypothesis may not be required in historical or descriptive studies where finding facts is the objective.

6. A review of related literature helps in the formulation of hypothesis by revealing to the researcher the variables that have been considered important in relation to his/her problem, which aspects have already been studied and which are left to be studied, which theories have supported the relationships and which theories present a contradictory relationship.

7. A test is called one-sided (or one-tailed) only if the null hypothesis gets rejected when a value of the test statistic falls in one specified tail of the distribution.

8. If the null hypothesis H_0 when false is accepted, the researcher is committing an error which is called Type II error.

9. The level of significance is a step in the testing of hypothesis exercise which denotes the probability of rejecting the null hypothesis when it is true. It is denoted by Q.

2.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the essential key points that must be defined at the time of hypothesis formulation.

2. How does a hypothesis add scientific rigour to the research process?

3. Write the difference between null hypothesis and alternative hypothesis.

4. Explain the difference between one-tailed and two-tailed tests.

5. What are Type I and Type II errors?

Long-Answer Questions

1. Discuss the characteristics of valid hypothesis.

2. Describe the steps in the formulation of hypothesis with the help of example.

3. Examine the importance of hypothesis.

4. Discuss the sources of hypothesis.

5. Explain the steps in testing of hypothesis.

6. Discuss what happens if the two population variances are equal.

2.9 FURTHER READING

Chawla D and Sondhi, N. 2011. *Research Methodology: Concepts and Cases*, New Delhi: Vikas Publishing House.

Cooper, Donald R. 2006. *Business Research Methods.* New Delhi: Tata McGraw-Hill Publishing Company Ltd.

Malhotra, N K. 2002. *Marketing Research – An Applied Orientation*. 3rd edn. New Delhi: Pearson Education.

Krishnaswami, O R. 2013. *Methodology of Research in Social Sciences*. New Delhi: Himalaya Publishing.

David, Matthew and Carole Sutton. 2004. *Social Research: The Basics*. United States (California): SAGE Publications Ltd.

UNIT 3 SCIENTIFIC STUDY OF SOCIAL PHENOMENA

Structure

3.0 INTRODUCTION

A very important part of understanding the nuances of social research is the scientific study of the social phenomenon. How is science related to society and its element? Can science and its method be used for analysing the society? Social research, too involves scientific methods of study. The basic elements that define scientific methods are known as concepts, yet all concepts are by and large only abstractions. Facts are logical constructions of concepts. A close relationship exists between a theory and fact where a theory is only a speculation and it has to be proved before it can be called a theory definitively. When a theory is proved, it becomes a fact.

There are various methods of scientific enquiry. The two most important ones are—deductive and inductive reasoning. Deductive reasoning is one where the researcher moves from the general to the specific part of the research. Inductive reasoning is also called the bottom-up approach. It tends to move from particular observations to wide generalizations.

There is an ongoing debate on the subject of objectivity or subjectivity of research carried out by researchers where the researchers are supposed to be unbiased and objective rather than subjective. Herein, value neutrality is the duty of sociologists to strive to be impartial and overcome their biases as they conduct their research. There is also a very common debate in sociology pertaining to the approach that is used in the social research: quantitative or qualitative.

But understanding the method of scientific study will be incomplete without a discussion on its background. Positivism is a very dominant philosophy to sociology. As a philosophical ideology and movement, positivism first assumed its distinctive features in the work of the French philosopher Auguste Comte, who named the systematized science of sociology. It then developed through several stages known by various names, such as Empiriocriticism, Logical Positivism and Logical Empiricism and finally in the mid-20th century flowed into the movement known as Analytic and Linguistic philosophy.

In its basic ideological posture, positivism is worldly, secular, anti-theological and anti-metaphysical. In this unit, you will learn about the scientific method, objectivity and subjectivity, debate in social research and positivism in sociology.

3.1   UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss the nature of science
- Describe the scientific method
- Explain the concept of objectivity and subjectivity
- Examine the debate in social research
- Explain the theory of positivism in sociology

3.2   THE SCIENTIFIC METHOD

Science refers to organized knowledge, but this knowledge and these facts are seldom conclusive. New experiences and additional information constantly alter the previous findings and replace them with generalizations that confirm the latest findings.

The same is the case with social sciences. The scientific method can also be applied to subjects in social sciences.

3.2.1 The Nature of Science

The method which integrates the most important aspects of the deductive and inductive methods of research is recognized as *scientific method*. It is generally attributed to Charles Darwin.

The scientific method is a back-and-forth movement of thought in which man first operates inductively from partially known or sometimes confused information learned from experience, previous knowledge, reflective thinking, observation and so on, towards a meaningful whole or hypothesis, and then deductively from suggested whole or hypothesis to the particular parts in order to connect these with one another in a meaningful pattern to find valid relationships. In the words of (Dewey 1933, p. 87):

While induction moves from fragmentary details (or particulars) to a connected view of situation (universal), deduction begins with the latter and works back again to particulars, connecting them and binding them together.

Although, in practice, scientific method involves a double movement of reasoning from induction to deduction, in its simplest form, it consists of working inductively from observations to hypotheses and then deductively from the hypotheses to the logical implications of the hypotheses in relation to what is already known.

Scientific method differs from other methods of acquiring knowledge on the basis of generalizations from authority, tradition, experience, and syllogism. It also differs from the methods of chance, of trial-and-error, and of intuition. When using the scientific method, one engages himself in a thinking process called *reflective thinking*. The five stages of reflective thinking furnished by Dewey (1911) may be summarised as under (Van Dalen 1973, p. 13; Whitney 1964, p. 3):

1. ***The occurrence of a felt difficulty****. Man comes across some obstacle, experience, or problem that puzzles him.

   (*a*) He lacks the means to achieve the end desired.

   (*b*) He feels difficulty in identifying the character of an object.

   (*c*) He is unable to explain an unexpected event.

2. ***Identification and definition of the difficulty in terms of a problem statement.*** Man makes observations and gathers facts so that he is able to define his difficulty more precisely.

3. ***Suggested solutions of the problem—hypotheses.*** Man makes intelligent guesses about possible solutions of the problem from the preliminary study of the facts. Such guesses that he makes to explain the facts about the cause of difficulty are called *hypotheses.*

4. ***Deriving consequences of the suggested solutions with the help of deductive reasoning.*** With the help of deductive arguments, man reasons that if each hypothesis is true, certain consequences should follow.

5. ***Experimental verification of the hypotheses.*** Man verifies each hypothesis by searching for observable evidence that will confirm whether or not the consequences that follow actually occur. This process enables him to test which hypothesis is in conformity with observable facts.

The stages involved in reflective thinking presented above suggest a pattern that is employed in the scientific method. It will be seen that the pattern describing this method runs parallel to the stages involved in reflective thinking.

From the earlier times, man has been curious about anything he could not understand. Slowly and gradually he developed the scientific method of thinking and of investigating his problems which, today is producing astonishing results. It is an orderly system of searching for truth which, by basing conclusions upon factual evidence, and by using reasoning as a means of showing relationship between ideas, has given him better and more accurate answers to his many problems, not only in physical and biological sciences, but also in behavioural and social sciences. By attempting to apply this method of inquiry to behavioural and social sciences, the fields of psychology, economics, political science, sociology, anthropology, and education have become recognized as sciences. The term *science*, therefore, is now thought of as a method or attitude rather than a field of subject matter. It is described as a method of inquiry that permits man to examine the phenomena of interest to him.

Science is based on certain beliefs and assumptions which are briefly described as under:

1. All events in nature are, at least to a degree, lawful or ordered, predictable and regular. This order, predictability and regularity of nature can be discovered through the activities of the scientific method.

2. Truth can ultimately be derived only from observation. Scientist does not depend upon authority as a source of truth, but relies upon empirical observation. Thus, the phenomena that can actually be observed to exist are within the domain of scientific method.

3. The scientist maintains a doubtful attitude towards data. He regards findings as tentative unless they are verified. Verification of the findings requires that

other scientists must be able to repeat the observations and get the same results.

4. The scientist is objective, impartial and logical in collecting and interpreting data or making observations. His personal bias does not in any way influence the truth and facts even when they are not in conformity with his own opinions.

5. Scientist does not bother about the moral implications of his findings. He always deals with facts and does not consider what finding is good or what is bad for us.

6. The ultimate goal of science is to integrate and systematise findings into a meaningful pattern or *theory.* The theory, however, is regarded as tentative and not the ultimate truth. It is subject to revision or modification as new evidence is found.

Theories are statements that explain a particular segment of phenomena by specifying certain relationships among variables. According to Kerlinger (1978, p. 9):

A theory is a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena.

This definition emphasises three things:

- First, a theory is a set of statements comprising of interrelated constructs, definitions and propositions.

- Secondly, a theory sets out the relations among the set of constructs, definitions or propositions in order to present a systematic view of the phenomena.

- Thirdly, a theory explains a phenomenon by specifying what variables are related to what variables and how they are related, thus, enabling the scientist to predict from certain variables to certain other variables.

A teacher, for example, has a theory of low academic achievement. His variables might be home environment, intelligence, verbal and numerical aptitudes, anxiety, economic factors and achievement motivation. The phenomenon to be explained is low academic achievement. It is explained by specified relations between each of the seven variables and low academic achievement, or by combinations of the seven variables and low academic achievement. The teacher seeks to understand low academic achievement with the help of this set of relations or constructs. He is then able to 'explain' and to some extent at least 'predict' it. Moreover, he can also to some extent 'control' the low academic achievement by making changes in the environment or manipulating some of the variables.

Scientific theories serve as means and ends in the development of science. As means, they provide a framework which guides scientist in making observation and discovery. Theories summarize knowledge and put in order the knowledge within a given field. They also clarify and provide meaning to this summarized knowledge comprising of isolated empirical findings. As ends, theories provide scientists explanations for observed events and relationships for specific phenomena with maximum objectivity. They do so by showing what variables are related and how they are related. On the basis of such relationships, scientists make deductions and predict about what will happen in certain situations under specified conditions. In this way, theories help in the development of new knowledge.

A sound theory must meet the following criteria if it is to serve its purpose in the development of science.

1. A theory must be stated in simple and precise terms. A theory that explains the most in a simple form is preferred to one that has more complexities and assumptions. This is the *Law of Parsimony.*

2. A theory must be in conformity both with the observed and with the previously established body of knowledge or validated theories.

3. A theory must provide means for its own interpretation and verification. In other words, it must provide deductions which can be tested empirically.

4. A theory must guide new discoveries and identify areas which are in need of investigation. It can do so if it is based on empirical facts and relationships.

Theories often offer a crude and general explanation of phenomena. They are refined and modified as knowledge in the form of facts accumulates. The discovery of pertinent facts is essential in order to determine whether a theory can be confirmed or should be rejected or reformulated. For example, if the facts found do not substantiate the theory, a scientist must reject or reformulate the theory to fit the new facts.

3.2.2 Steps, Process and Objecives of Scientific Method

In this section, we will study, in detail, the steps, process, objectives and characteristics of the scientific method.

Steps in Scientific Method

The steps involved in the scientific method are as follows:

- Collection of data as per the problem in hand, according to some adequate plan and their systematic observation.

- Observations are made with a well-defined purpose and they are recorded in definite terms.

- Classification and organization of data on the basis of similarities, variations, activities, causes and results.

- Generalization of data for the purpose of formulating principles and theories. The principles and theories must be specifically defined so that it can solve the problems in the related field.

- Verification of generalizations through controlled experiments by tested prediction of results and by repetition of experiments. Correlation coefficient of original as well as verification of results is also calculated and probable errors are estimated. It is also determined whether the error lies in procedure or apparatus.

- Assumptions and limitations are noted down on the basis of verification of results.

- Reporting the research in detail.

- Announcement of the results before the general public for practical use.

Steps in Scientific Process

The steps involved in a scientific process are as follows:

(i) **Purposeful observation:** Observation should be accurate and extensive, and it must be done under various controlled conditions.

(ii) **Analysis-synthesis:** This includes the following:

- The essential elements in a problematic situation must be selected by analysis.
- Similarities as well as dissimilarities must be isolated.
- Exceptions are to be given special attention.

(iii) **Selective recall:** A wide range of experiences is essential.

(iv) **Hypothesis:** It is a tentative solution to the problem. There may be more than one solution depending on the nature of the problem.

(v) **Verification by inference and experiment:** Here, only one variable is manipulated and judgment is made on the adequacy and accuracy of data.

Objectives of Scientific Research

The terms research and scientific method are sometimes used inter changeably.

Research is a systematic process for developing a theory by applying scientific methods. It is an impartial, objective, empirical and logical analysis and recording of controlled observation that finally leads to the development of a theory, principles, laws, etc., and helps us to predict about the phenomenon in future.

A research is said to begin with a question or a problem. The purpose of a research is to find out solutions through the application of systematic and scientific methods. Thus, research is a systematic approach to a purposeful investigation.

The main aim of research is to uncover answers to questions by applying scientific procedures. Research aims to discover hidden truths. While each research initiative has a particular purpose, the objectives of research can be broadly characterized as follows:

- **Exploratory/formulative research:** It attempts to get familiar with a concept or to develop new insights into it.
- **Descriptive research**: It seeks to accurately portray the key characteristics of an individual, a situation or a group.
- **Diagnostic research:** It establishes the frequency with which an event occurs, or the frequency with which it is associated with something else.
- **Hypothesis-testing research:** This type of research tests the hypotheses of a causal relationship between variables.

Characteristics of Good Research

The process of research helps to increase the creative ability of a decision-maker. The various characteristics of research are as follows:

- **Interdisciplinary team approach:** This approach is based on the principle of using expertise and experience of different personnel working in different disciplines within an organization. An individual cannot be an expert in all the areas of operation. So, researchers take help from other experts, who are specialists in their respective fields. Under interdisciplinary team approach, an expert may use old solutions, which were used in the past as research material for finding the most appropriate solution to a problem.
- **Methodological process:** The researcher uses scientific methods and techniques to provide optimum solution to problems. The scientific methods include observing and defining a problem and formulating hypothesis related to the results of the

scientific methods and techniques. If the hypothesis is accepted, its results should be executed in an organization; but if the hypothesis is not accepted, another hypothesis is formulated.

- **Objectivistic approach:** The aim of an organization is to have optimal solutions to various problems. It is essential to measure the desirability of a solution for achieving the organizational objective. This measured desirability helps in comparing the alternative courses of action with respect to their outcomes.

- **Economical in nature:** In an uncertain and complex situation, research helps in reducing the costs of inventory, thereby improving profits. For example, in inventory control, research can provide scientific rules for reducing acquisition costs and inventory-carrying costs.

The qualities of good research are as follows:

- **Good research is systematic:** This means that the research lays out clear steps in a specified sequence in compliance with well-defined rules. Being systematic does not mean that the research cannot be based on creative thinking. On the other hand, it dramatically reduces guesswork-based and intuitive conclusions.

- **Good research is logical:** This implies that the use of sound logic provides a foundation for reasoning, induction and deduction, which are of great significance for carrying out high quality research. Induction entails reasoning from a part to the whole, while deduction is the process of reasoning, wherein a premise is driven to a conclusion which is based on that very premise. In fact, logical reasoning leads to more meaningful research and better eventual decision-making.

- **Good research is empirical:** This means that research is related to one or several aspects of a real situation and uses concrete data which provides a basis for external validity to the research results.

- **Good research is replicable:** Good research allows for research results to be verified by replicating the study, thereby building a sound basis for decisions.

Further, Best and Kahn (1992) have summarized the main characteristics of research as follows:

- Research seeks to find a solution to a problem. In this objective, it could answer a question or even determine the relationship between several variables.

- Research creates generalizations, principles and theories that enable the prediction or anticipation of future occurrences. Research studies specific objects, groups or situations and then applies these characteristics observed to a larger population than the sample observed. Research goes beyond just simply retrieving or gathering information. There are many schools where the research teams gather and tabulate statistical information. This information can be used for decision-making, but it is not necessary to do so.

- Research is based on observations or empirical evidence. There are many questions which are interesting or relevant but, since they cannot be observed, they do not become research procedures. Research does not accept revelation or dogma as a basis for establishing knowledge. Research only accepts that which can be verified by observation.

- Research requires accuracy of observation and description. Researchers rely on quantitative or numerical measuring devices which are accepted as precise means

of description. They identify or create appropriate data gathering instruments or procedures and employ effective mechanical, electronic, or psychometric techniques to improve human observation, recording, computation and analysis of data.

- Research entails obtaining new data from first-hand sources, or uses existing data towards a new purpose. Teachers often guide their students to undertake a project which requires them to write a paper detailing the life of a prominent person. The students consult encyclopaedias, books, or periodicals and summarize the information in writing. This is not research, the information is not new. Simply rewriting or representing what is already known may be a valuable learning experience, but it is not research. It does not provide any new information.

- Research may sometimes appear to be random or unsystematic. However, it is actually always based on carefully designed procedures and rigorous analysis. Although researchers may sometimes employ trial and error methodologies, research is not a blind, random investigation, where the researcher is just experimenting to see what happens.

- A good researcher requires significant expertise. He/She is already aware of what is known about the problem including the investigations carried out by others. The researcher familiar with the related literature and also understands the terminology, concepts and technical skills necessary to thoroughly assess the data that he/she has gathered.

- Researchers must apply objectivity and logic and must also remove all their personal biases. They must employ all possible tests in order to comprehensively validate the procedure followed, the data sourced, and the results or conclusions that have been arrived at. Researchers should not make any effort to be additionally persuasive in order to prove an emotionally held conviction by them. Their focus must be on testing, and not on proving the hypothesis. Total objectivity is just as rare as absolute righteousness, and therefore, researchers must not allow bias or emotion to affect their analysis.

- Research involves the quest for answers to unsolved problems. Pushing back the frontiers of ignorance is its goal, and originality is frequently the quality of a good research project. However, previous important studies are deliberately repeated, using identical or similar procedures, with different subjects, different settings, and at different times. This process is a replication, a fusion of the words, repetition and duplication. Replication is always desirable to confirm or to raise questions about the conclusions of a previous study.

- Research must be carried out patiently and not in a rushed manner. It's outcome and result are mundane rather than spectacular, and the research team must be prepared to face disappointment in the pursuit of answers to their unanswered questions.

- The process and outcomes of research are meticulously recorded. Every key term is defined, restrictive factors are acknowledged, procedures are carefully described, all references are recorded, results are objectively documented, and the final outcomes are presented with caution and restraint. The final research reports and supporting data are made available for associates and other scholars to study, analyse, evaluate and even replicate.

3.3 OBJECTIVITY-SUBJECTIVITY DEBATE

It is important for a study to be objective in nature, no matter which discipline it represents. Whether it is natural or social science research, a study which is subjective in nature does not produce results which can be depended upon. That is because being objective in social science means being able to check and examine the evidences which the study has produced. An objective study is free from personal biases and beliefs of the researcher. Objective studies are not speculative in nature but are dominated by actual observations.

When compared with natural sciences, it is much more difficult to obtain objectivity in social sciences. This is largely because social sciences deals with humans who live in a society and pursue their own interests. Both of these are highly dynamic and their attitudes and approaches to life can change at any given time and also evolve with time. Therefore, while objectivity is important, it is difficult to obtain the same in social sciences as personal views and biases often enter the researcher's work. There are hindrances like emotional values, complex subjects, uniformity problems, self-interest, lack of time for research, and bias and prejudices. Among all these, the most severe in terms of research are prejudices and biases which can greatly affect the nature of the study. These biases can enter different stages of research, like during the collection of the sample, in the choice of the participants, during data collection, analysis and finally in the findings.

Value Neutrality

Value neutrality is the duty of sociologists to strive to be impartial and overcome their biases as they conduct their research. A researcher can easily mix his past experiences and present them as part of the present study. This leads to inclusion of subjectivity in the study. Biases can also enter during the supply of information because many times information is collected by those who may carry their own prejudices. They may present information which is suited to their needs, thus keeping the wholesome information out of reach of the researcher. All these can increasingly influence the findings of the study and mar the objectives with which it was undertaken in the first place. There are several ways by which subjectivity can be removed from research. These are, first, depending less on the investigation. This means that while data is critical to social science research, the researcher can avoid hiring other investigators and have first-hand information of the field to avoid subjectivity. Second, use of statistical measurements and methods can be made wherein data can be correlated and checked to remove subjectivity. Third, concepts and terms could be standardized to avoid multiple usage and inculcate clear and precise use to suit the views of the researcher. Fourth, the introduction of a questionnaire through which information is collected can lead to clarification and also provision of matter which can be cross-checked any time. Fifth, at various stages, the researcher can employ counter-checks and compare findings at several stages to avoid biases and prejudices from entering the study. Another method is that of random sampling with which the researcher can select participants from different shades to avoid his own prejudice that might enter the study during the collection of the sample.

Problems of Subjectivity and Objectivity

Qualitative methodology is familiar with the fact that the subjectivity of the researcher is a deep ingredient of his scientific research. Everything is guided by subjectivity, beginning

from the choice of topic that one studies, to devising hypotheses, to selection of methodology and interpretation of data. Qualitative methodology motivates researchers to mirror the values and objectives brought by them into their research and the manner in which they influence the research project. It also motivates other researchers to replicate the values used by a particular investigator.

The most important issue that comes up when subjectivity is recognized, is its impact on objectivity. Two types of arrangements have been expressed. Many qualitative researchers make subjectivity and objectivity counterbalance each other. It is believed that objectivity negates subjectivity. The reason for this is because it makes the observer an inert recipient of external information, without an agency. The subjectivity of the researcher is said to negate the possibility of objectivity, knowing a social psychological world. The investigator's values are said to define the world that is studied. It never happens that one really sees or talks about the world, per se. However, one only sees and talks about what one's values dictate.

Subjectivity is usually considered as the most essential and initiating part of qualitative methodology. Nevertheless, this is not true. Qualitative methodology also has traces of objectivity. Objectivity indicates that the subjectivity of a researcher can empower him to precisely understand the world in its true state. Beyond doubt, subjectivity can cause prejudice to affect the researcher and rule out objective understanding of a subject's psychological reality. However, this is not unavoidable. In the true sense, one of the benefits of getting to recognize subjectivity is to reflect on whether it makes objective comprehension easy or obstructs it. Distorting values can then be replaced by values that improve objectivity.

Objectivism puts subjectivity and objectivity together since it reasons that objective knowledge needs active, sophisticated and subjective processes like: perception, analytical reasoning, synthetic reasoning, logical deduction and the ability to differentiate essences from appearances. On the other hand, subjective processes can augment objective conception of the world.

Objectivism is the highest form of the subjects being studied as part of research methodology. It considers psychological reality as something meaningful and important which must be accurately comprehended. Subjectivism either denies a psychological reality to subjects, or else makes it unknowable. The psychology of other people is clouded by the subjectivity of the observer and is not recognized for what it (truly) is.

3.4 DEBATE IN SOCIAL RESEARCH

We have learnt about the two paradigms of research in Unit 1: quantitative and qualitative research. Both of them have different advantages and disadvantages. Their uses too are not similar.

Use of Quantitative Research

There are four chief categories of research questions that quantitative research can answer:

- The first type of research question that demands a quantitative answer is: For instance, How many students from below poverty line have opted for studying education? Quantitative research can be used for answering this type of question.

- Only quantitative methods can be used to study numerical changes accurately: Is the number of girls in our university falling or rising? Is there any improvement in performance or is it dipping? A quantitative study is needed to answer these questions.

- Generally, for explaining phenomena: What factors relate to change in the livelihood of socially backward classes over time?

- The ultimate activity which requires quantitative research is the testing of hypotheses: For example: Whether there is a relationship between students' achievements and their self-esteem and social background can be explained with the help of quantitative research. One could come up with the hypothesis that lower social class background leads to low self-esteem, which would in turn, be related to low achievement. Using quantitative research, we can try to test this model.

Use of Qualitative Research

Qualitative research is extensively employed where small sections of the population (or groups of people who have common traits) particularly interest the researcher. Some of the main reasons for carrying out qualitative researches are as follows:

- For assessing a market, product or consumer, in the absence of any other information.

- For identification and investigation of concepts.

- Apprise the researchers about a group of consumers, if they are not aware of them.

- For identification of behavioural patterns, values, approaches, views and purpose.

- For setting up precedence among types of behaviour, beliefs, opinions and attitudes.

- For detailed identification of problems and to build models for further research.

- To append to the points emerging from a pilot or major survey.

- For providing verbatim remarks and anecdotes from participants, so that the outcome of the research can be made available to the client.

- For testing the functioning of a questionnaire, by going through questions related to routing, signposting, understanding and ambiguity.

Table 3.1 Comparison of Qualitative and Quantitative Researches

Criteria	Qualitative Research	Quantitative Research
Purpose	Describes individuals and events in their natural setting	Explores, describes, tests or assesses phenomena
Group studied	Smaller and not randomly selected	Larger and randomly selected
Variables	Study of the whole, not variables	Specific variables studied
Types of data collected	Words, images, or objects	Numbers and statistics
Forms of data collected	Qualitative data such as open-ended responses, interviews, participant observations, field notes, and reflections	Quantitative data based on precise measurements using structured and validated data-collection instruments

Type of data analysis	Identify patterns, features, themes	Identify statistical relationships
Objectivity and subjectivity	Subjectivity is expected	Objectivity is critical
Role of the researcher	Active (immersion)	Passive(immersion optional)
Results	Particular or specialized findings that are less generic	Generic findings that can be applied to other populations
Scientific method	Exploratory or bottom–up: the researcher generates a new hypothesis and theory from the data collected	Confirmatory or top-down: the researcher tests the hypothesis and theory with the data
View of human behaviour	Dynamic, situational, social and personal	Regular and predictable
Most common research objectives	Explore, discover and construct	Describe, explain and predict
Focus	Wide-angle lens: examines the breadth and depth of phenomena	Narrow-angle lens: tests a specific hypotheses
Nature of observation	Study behaviour in a natural environment	Study behaviour under controlled conditions; isolated causal effects
Nature of reality	Multiple realities; subjective	Single reality; objective
Final report	Narrative report with contextual description and direct quotations from research participants	Statistical report with correlations, comparisons of means and statistical significance of findings

Quantitative and qualitative methods primarily differ in terms of their flexibility. In general, quantitative methods are quite rigid. When the researcher makes use of quantitative methods like surveys and questionnaires, for instance, all participants are presented with the same questions, in the same order. The types of responses given by the participants may either be closed-ended or fixed. The benefit of this rigidity is that it allows significant comparison of responses between participants and study sites. Nevertheless, it is important that the questioner knows which important questions to put across, the best way to frame them and the range of possible responses.

Qualitative methods are characteristically more elastic; i.e., they stir more spontaneous responses and adapt to the interface between the researcher and the study participant. For instance, qualitative methods comprise mostly open-ended questions that are not necessarily worded in exactly the same way for each participant. With open-ended questions, participants are free to write what they think and their responses are comprehensive.

The primary debate in social research pertains to these approaches to social research: quantitative and qualitative methods. Both the approaches not only have different uses and benefits, but also have different underlying assumptions. The question as to which ideology trumps the other is a complex one and not just a simple dichotomous choice. In fact, it will not be too far-fetched to say that a significant number of researchers now apply a consciously selected combination of the two approaches. There are varied schools which also believe that these two approaches are not polar opposites and can be substituted for each other while finding answers. It, ultimately depends on the researcher, the purpose of research and the research design as to which approach should be selected.

3.5 POSITIVISM IN SOCIOLOGY

The approach of positivism was formulated by Auguste Comte. Auguste Comte was born in France in 1798 during the height of the French Revolution, a period of chaos and unrest. His parents were devout Catholics and ardent royalists. Comte was a brilliant student excelling in physics and math with an unusual memory. His early career was poorly organized and a rather self-destructive affair in which he proceeded to 'shoot himself in the foot' several times. Along with 14 others he was expelled from school after a student uprising over a geometry instructor, thus dashing hopes of an otherwise promising academic career.

He did, nonetheless, manage to become secretary to Henri St. Simon, another prominent thinker with whom Comte shared many ideas. He met, and later married, a nineteen-year-old prostitute but had an unhappy married life. He had a falling out with St. Simon and organized on his own a subscription series of lectures on the 'Positive Philosophy'. Comte attempted suicide by throwing himself into the Seine and was rescued by a passer-by. Comte interpreted this Samaritan act as a sign that his mission in life was to complete and disseminate his positive philosophy.

In 1829, Comte completed the series of lectures, and between 1830 and 1842, published his *Cours de Philosophie Positive* in six volumes. In 1832, he managed to achieve a minor appointment at the Ecole Poly-technique, but, in 1844, he wrote a scathing article on St. Simon and the Ecole and was dismissed. During the same year, two other important events also occurred. Comte obtained a small stipend from the English philosopher, John Stuart Mill, who had been impressed by his *Positive Philosophy*, and he also began an affair with Madame Clotilde de Vaux. In 1846, she died in his arms and Comte was later to credit her with teaching him about the affective tendencies of human nature, a consideration which was to inform his suggestion for a 'religion of humanity'.

In fact, Comte was to see this religion of humanity as part of the practical application of his philosophy as recommended in his works—*The System of Positive Polity or Treatise of Sociology: Instituting the Religion of Humanity*. *Positive Philosophy* was the work in which he outlined his preferred way of knowing the world, and the *Positive Polity* contained his ideas about how to improve society, and how to establish what was, in his view, the best society possible by applying this knowledge.

According to Comte, a stable social order rested on a consistent form of thought. He saw his own thought as leading to the establishment of a more stable, industrial order. He saw this relationship between thought and practice as a natural, rather than a causal one and saw thought as evolving naturally towards the kind of philosophy which he was formulating and recommending. Ways of thinking, of philosophizing, of knowing the world, were, in his view, primary, both in the history of humankind and in his own practice. In other words, Comte believed that people acted in such a way as to correspond with the way they thought. In different societies or periods of history, furthermore, a person's way of thinking, of knowing their world, was responsible for producing the kind of society in which they lived.

Science of Sociology

According to Comte, sociology is a social, organic science. Sociology is a relatively new, evolving science dependent upon all the foregoing theories in science. However, it is quite clear that sociology is gradually moving towards the goal of a *definite* science.

Comte had a very wide conception of sociology. According to him, all other social sciences are subsumed under it. He believed in a unified integral study of all social sciences taken together. He posited that the subject matter of sociology is society. It studies the structure of the society and the set of rules governing its functions.

Since sociology tries to explore the principles which help society to stay integrated and in order, it is essential that the law of sociology should be scientific. In order to make the societal laws scientific, they should pass through the full circle of making of scientific laws, namely *observation*, *experimentation*, *comparison* and *classification*. What needs to be emphasized here is the fact that in making these societal laws, use of full scientific technique is necessary.

Comte maintained that the positive science of society called sociology must pursue the method which was followed by definite sciences like astronomy, physics, chemistry, and biology. He insisted that the new science must be *positive*. Positive means pursuing scientific methods of analysis and prognosis. The method of sociology includes observation as well as experimentation. Observation must be guided by a theory of social phenomena. Experimentation is controlled observation. In sociology, experimentation involves the study of pathological cases. According to him, central to sociology is the comparison of different co-existing states of human society on the various parts of the earth's surface. By this method, he argued, the different stages of evolution may be allowed once. These conventional methods of science, like observation, experimentation and comparison, must be used in combination with the historical method.

Law of Human Progress

The law of human progress is one of Comte's the most important central ideas. He proposed that the evolution of the human mind is parallel to the evolution of any individual mind. The development of the individual human organism is termed as ontogeny. This forms the basis for the development of phylogeny or the development of the human race. In our childhood, we all believed in imaginary worlds; when we become adults, we start accepting the world with its vices and virtues. Mankind has also undertaken quite a similar journey; from believing in the make-belief to the maturity of adulthood.

According to Comte:

> Each of our leading conceptions—each branch of our knowledge passes successively through three different theoretical conditions: the Theological or fictitious; Metaphysical or abstract; and the Scientific or positive. In theological state, human mind, seeking the essential nature of beings, the first and final causes (the origin of purpose) of all effects supposes all phenomena to be produced by the immediate action of supernatural beings. In Metapysical state the mind supposes abstract forces, veritable entities (that is personified abstractions) capable of producing all phenomena. In the final, the positive state, the mind has given over the vain search after Absolute notions, the origin of destination of the universe, and the causes of phenomena, and applies itself to the study of their laws, that is, their invariable relations of succession and resemblance

Theological or Fictitious State: Law of Three Stages

According to Comteian proposition, all theoretical conceptions, whether general or special, bear a supernatural influence. This kind of thinking is found among the primitive people and sometimes the thinking of children. At this state, there is substantial lack of logical and orderly thinking. However, Comte argues that the primitive man as well as children do have scientific outlook also. Owing to theological state of their minds, their understanding is characterized by an unscientific outlook. The main subject matter of

the theological state is natural events. The unusual and unintelligible events of nature tend man towards theological or fictitious interpretation of events. Unable to discover the natural causes of various happenings, the primitive man attributed them to imaginary or divine force. The explanation of natural events in non-natural, divine or imaginary conditions is known as theological or fictitious state. The theological state implies belief in the other world wherein reside divine forces which control the events in this world. It is clear that theological state implies a belief in divine and extraterrestrial forces. Comte has classified the theological state further in three stages:

- **Fetishism:** The first and primary stage in the theological state is that of *fetishism*. Fetishism is a belief that there is some living spirit in non-living objects. This is also known as *animism*. The concept of animism signifies that the inanimate objects are not dead but are possessed by living spirits. One can argue that in India, particularly rural and tribal areas, there is a widespread belief that some deities reside in tree, stones and mountains. Therefore, it has been seen that people engage in the worship of a particular tree, stone mountain.

- **Polytheism:** With the gradual development in human thinking, there occurred a change in the form of thinking. *Polytheism* is the next stage to fetishism. In this stage, man had classified god and every natural force had a presiding deity. Each god had some definite function and his scope and area of action was determined.

- **Monotheism:** The last and most developed form of theological state is seen manifested in *monotheism*. As the very term monotheism implies, at this level of human thinking a belief in one god had replaced the earlier belief in many gods. The monotheistic thinking symbolizes the victory of human intellect and reason over non-intellectual and irrational thinking. In monotheism, it is believed that one God is supreme and that he is responsible for the maintenance of order and system in the world.

Metaphysical or abstract state

The metaphysical or abstract thinking marks the second stage in the evolution of human mind. According to Comte, each successive stage is an improvement upon the earlier stage. With the gradual improvement in human mind, human problems also become more intricate. The theological state was not adequate to tackle these improvements efficiently. The appearance of conflicting and opposite forces in the world presented problems which could not be successfully tackled by monotheism. It was difficult to believe that the same god was responsible for prehistoric creation as well as destruction. A single god could not account for simultaneous creation and destruction. In order to resolve this intellectual query, metaphysical thinking was developed. Under metaphysical thinking, people believe that an abstract power or force guides and determines the events in the world. Metaphysical mind disregards belief in the presence of several gods.

Scientific or positive state

This state is the most advance and developed form of the human mind. All metaphysical knowledge is based upon speculation and is at best inferential knowledge. There are no direct means to confirm the findings of metaphysical knowledge; it is purely a matter of belief or temperament. The modern temperament of man is such that it cannot remain satisfied with mere guesswork; it craves for positive knowledge which can be scientifically confirmed. The positive and scientific knowledge is based upon facts, and these facts are gathered by observation and experience. The observation and classification of facts

are the beginning of scientific knowledge. From these facts we generalize and draw conclusions. These conclusions, in turn, are subjected to verification. Once verified, these become established laws, which can be relied upon in gathering and classifying the facts. Scientific thinking is thoroughly rational and in it there is no place for any belief or superstition. According to Comte, the human mind before reaching the state of positivism, must have passed through the two earlier stages of theological and metaphysical states.

The three stages suggested by Comte have a strong idealistic basis. Yet he correlated every stage of evolution of the human mind with social organizations present in that period. The theological stage that corresponds roughly with the ancient age is dominated by the rule of the army and priests. In the metaphysical state, society was dominated by clergy and lawyers. This state roughly falls during the Middle Ages and the Renaissance. The modern era marks the beginning of the positive state and is generally ruled by industrialists and scientific moral guides. In the first state, the family takes centrestage, while in the second, the State rises to prominence. In the third state, however, the entire civilization has become an operative social unit.

3.6 SUMMARY

- Science refers to organized knowledge, but this knowledge and these facts are seldom conclusive. New experiences and additional information constantly alter the previous findings and replace them with generalizations that confirm the latest findings.

- The method which integrates the most important aspects of the deductive and inductive methods of research is recognized as *scientific method*. It is generally attributed to Charles Darwin.

- Although, in practice, scientific method involves a double movement of reasoning from induction to deduction, in its simplest form, it consists of working inductively from observations to hypotheses and then deductively from the hypotheses to the logical implications of the hypotheses in relation to what is already known.

- Scientific method differs from other methods of acquiring knowledge on the basis of generalizations from authority, tradition, experience, and syllogism. It also differs from the methods of chance, of trial-and-error, and of intuition. When using the scientific method, one engages himself in a thinking process called *reflective thinking*.

- Scientific theories serve as means and ends in the development of science. As means, they provide a framework which guides scientist in making observation and discovery. Theories summarize knowledge and put in order the knowledge within a given field. They also clarify and provide meaning to this summarized knowledge comprising of isolated empirical findings.

- The steps involved in a scientific process are as follows: purposeful observation, analysis-synthesis, selective recall, hypothesis, and verification by inference and experiment.

- Characteristics of a good research: interdisciplinary team approach, methodological process, objectivistic approach, economical approach and economic in nature.

- It is important for a study to be objective in nature, no matter which discipline it represents. Whether it is natural or social science's research, a study which is

subjective in nature does not produce results which can be dependent upon. That is because being objective in social science means being able to check and examine the evidences which the study has produced.

• Value neutrality is the duty of sociologists to strive to be impartial and overcome their biases as they conduct their research.

• Qualitative methodology is familiar with the fact that the subjectivity of the researcher is a deep ingredient of his scientific research. Everything is guided by subjectivity, beginning from the choice of topic that one studies, to devising hypotheses, to selection of methodology and interpretation of data.

• Many qualitative researchers make subjectivity and objectivity counterbalance each other. It is believed that objectivity negates subjectivity. The reason for this is because it makes the observer an inert recipient of external information, with an agency.

• Objectivism puts subjectivity and objectivity together since it reasons that objective knowledge needs active, sophisticated and subjective processes like: perception, analytical reasoning, synthetic reasoning, logical deduction and the ability to differentiate essences from appearances. On the other hand, subjective processes can augment objective conception of the world.

• Quantitative and qualitative methods primarily differ in terms of their flexibility. In general, quantitative methods are quite rigid. Qualitative methods are characteristically more elastic; i.e., they stir more spontaneous responses and adapt to the interface between the researcher and the study participant.

• The primary debate in social research pertains to these approaches to social research: quantitative and qualitative methods. Both the approaches not only have different uses and benefits, but also have different underlying assumptions. The question as to which ideology trumps the other is a complex one and not just a simple dichotomous choice. It, ultimately depends on the researcher, the purpose of research and the research design as to which approach should be selected.

• According to Comte, sociology is a social, organic science. Sociology is a relatively new, evolving science dependent upon all the foregoing theories in science. However, it is quite clear that sociology is gradually moving towards the goal of a *definite* science.

• The positive and scientific knowledge is based upon facts, and these facts are gathered by observation and experience. The observation and classification of facts are the beginning of scientific knowledge. From these facts we generalize and draw conclusions. These conclusions, in turn, are subjected to verification. Once verified, these become established laws, which can be relied upon in gathering and classifying the facts.

3.7 KEY TERMS

• **Scientific method:** It refers to the process of working inductively from observations to hypotheses and then deductively from the hypotheses to the logical implications of the hypotheses in relation to what is already known.

• **Deductive reasoning:** It is one where the researcher moves from the general to the specific part of the research.

- **Objective study:** It refers to a study which is free from personal biases and beliefs of the researcher.
- **Value neutrality:** It is the duty of sociologists to strive to be impartial and overcome their biases as they conduct their research.
- **Positivism:** It is a sociological philosophy as per which positive means pursuing scientific methods of analysis and prognosis. The method of sociology includes observation as well as experimentation. Observation must be guided by a theory of social phenomena. Experimentation is controlled observation.

3.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The knowledge obtained from science is seldom conclusive because the new experiences and additional information constantly alter the previous findings and replace them with generalizations that confirm the latest findings**.**

2. Scientific method differs from other methods of acquiring knowledge on the basis of generalizations from authority, tradition, experience, and syllogism. It also differs from the methods of chance, of trial-and-error, and of intuition. When using the scientific method, one engages himself in a thinking process called *reflective thinking*.

3. The analysis-synthesis step in the scientific process includes the following points: the essential elements in a problematic situation must be selected by analysis, similarities as well as dissimilarities must be isolated and exceptions are to be given special attention.

4. Being objective in social science means being able to check and examine the evidences which the study has produced. An objective study is free from personal biases and beliefs of the researcher. Objective studies are not speculative in nature but are dominated by actual observations.

5. Biases can enter during the supply of information because many times information is collected by those who may carry their own prejudices. They may present information which is suited to their needs, thus keeping the wholesome information out of reach of the researcher.

6. Objectivism is the highest form of the subjects being studied as part of research methodology.

7. Qualitative methods are characteristically more elastic; i.e., they stir more spontaneous responses and adapt to the interface between the researcher and the study participant.

8. The main subject matter of the theological state is natural events. The unusual and unintelligible events of nature tend man towards theological or fictitious interpretation of events. Unable to discover the natural causes of various happenings, the primitive man attributed them to imaginary or divine force.

9. The metaphysical or abstract thinking marks the second stage in the evolution of human mind. According to Comte, each successive stage is an improvement upon the earlier stage. With the gradual improvement in human mind, human problems also become more intricate.

10. The first and primary stage in the theological state is that of fetishism. Fetishism is a belief that there is some living spirit in non-living objects. This is also known as animism.

3.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is scientific method? List and describe its various steps.

2. List the qualities of good research.

3. State the various ways by which subjectivity can be removed from research.

4. What is value neutrality?

5. How is sociology moving towards definite science?

Long-Answer Questions

1. What is science? Discuss the role of scientific theories in the development of science.

2. Critically analyse the debate on objectivity and subjectivity in the field of research.

3. Compare the quantitative and qualitative approach in research.

4. Explain Comte's theory of positivism in sociology.

5. Assess the law of three stages.

3.10 FURTHER READING

Chawla, D. and N. Sondhi. 2011. *Research Methodology*. New Delhi: Vikas Publishing House.

Paneerselvam, R. 2009. *Research Methodology*. New Delhi: Prentice Hall of India.

Kumar, Ranjit. 2011. *Research Methodology: A Step-by-Step Guide for Beginners*. Australia: SAGE.

Cooper, Donald R. 2006. *Business Research Methods.* New Delhi: Tata McGraw-Hill Publishing Company Ltd.

Malhotra, N K. 2002. *Marketing Research – An Applied Orientation*. 3rd edn. New Delhi: Pearson Education.

Krishnaswami, O R. 2013. *Methodology of Research in Social Sciences*. New Delhi: Himalaya Publishing.

Ritzer, George. 1988. *Sociological Theory*, Second edition. New York: McGraw-Hill.

UNIT 4 TYPES OF RESEARCH

Structure

4.0   INTRODUCTION

In this unit, we shall discuss the different types of research in detail. Research design is a detailed plan used by the researcher to understand the methods by which he can achieve the objectives set forth in the study. It involves:

- Selection of the type of research design

- Selecting appropriate data sources

- Selecting the sampling plan

- Determining the sample size

- Understanding the issues related to measurement and scale selection

There are varied types of research design available for the research to choose. Each one has different purpose and sample size apart from the techniques. Additionally, each type of research also has certain advantages and limitations. It depends on the researcher, the topic of research and the tools available which play a crucial role in determining the selection of type of research. The unit will provide you a basic understanding of basic, applied, conceptual and empirical research along with descriptive, exploratory and experimental type of research.

4.1   UNIT OBJECTIVES

After going through this unit, you will be able to:

- Identify the concept of basic and applied research

- Describe conceptual and empirical research

- Discuss the purpose, advantages and disadvantages of historical research

- Assess exploratory research

- Analyse descriptive and experimental research

4.2 BASIC, APPLIED AND EMPIRICAL RESEARCH

Research can be either fundamental (basic or pure) or action-oriented (applied) research. Fundamental research focuses on finding generalizations and formulating theories. It is the research done for knowledge enhancement; the research which does not have immediate commercial potential; and the research which is done for human welfare, animal welfare and plant kingdom welfare. For example, research on the institution of marriage came into being is an example of basic or fundamental research. Here the main motivation is to expand man's knowledge and not to create or invent something. Basic research lays down the foundation for the applied research.

Applied research is designed to solve practical problem of the modern world, rather than to acquire knowledge for the sake of knowledge. Its goal is to improve the human condition. It focuses on analysis and solving social and real life problems. This research is usually conducted on large scale basis and is expensive. Thus, it is often conducted with the support of some financing agency like government, public corporation, World Bank, UNICEF, UGC, etc. Examples of applied research topics include persuasion, eyewitness memory, clinical treatments of psychological disorders, behavioral interventions for children with autism, decision making, etc.

4.2.1 Empirical Research

Conceptual research is that which is related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. In a conceptual research, a concept is chosen for examination, and the research involves quantifying and tallying its presence. 'Scoping of Vulnerability Definitions of Polio' ——the research done by the United Nations University Institute for Environment and Human Security (UNU-EHS) is an example of conceptual research. As part of the Global Pulse initiative's design and development phase, the UNU-EHS conducted a series of research to help the project gain a better understanding of how different communities of practice use the term 'vulnerability'. As part of the research, UNU-EHS analysed 76 definitions of vulnerability used by UN agencies, NGOs, scientific organizations and academia. In addition, the Institute's researchers scanned through 68 reports to draw out key lessons for the development of vulnerability indicator sets.

The development of psychoanalysis as a science and clinical practice is another example of conceptual research. Research has clarified, formulated and reformulated psychoanalytic concepts permitting to better shape the findings emerging in the clinical setting. By enhancing clarity and explicitness in concept usage it has facilitated the integration of existing psychoanalytic thinking as well as the development of new ways of looking at clinical and extra-clinical data.

Empirical research, on the other hand, relies only on real experiences and observations. It is data-based research and its conclusions can be verified by observations or experiments. It is also called experimental type of research. In empirical research, all facts are obtained at first hand, at their source, and at times by stimulating the production of desired information. To prove a given hypothesis, the evidence gathered through empirical studies and experiments is considered to be the most powerful and accurate. Research design varies by field and by the question being investigated. Many researchers combine qualitative and quantitative forms of analysis to better answer questions which cannot be studied in laboratory settings, particularly in the social sciences and in education.

In some fields, empirical research may begin with a research question. For example: 'Does listening to vocal music during the learning of a word list have an effect on later memory for these words?' This question is tested through experimentation in a lab. Usually, a researcher has a certain theory regarding the topic under investigation. Based on this theory, some statements or hypotheses are proposed, for example, 'listening to vocal music has a negative effect on learning a word list'. From these hypotheses, predictions about specific events are derived, for example, 'people who study a word list while listening to vocal music will remember fewer words on a later memory test than people who study a word list in silence'. These predictions are then tested with a suitable experiment. Depending on the outcomes of the experiment, the theory on which the hypotheses and predictions were based will be supported or not.

4.3 HISTORICAL RESEARCH

History is a meaningful record of past events. It is a valid integrated account of social, cultural, economic and political forces that had operated simultaneously to produce historical events. It is not simply a chronological listing of events but an integrated assessment of the relationship between people, events, times and places. It is used to understand the present on the basis of what we know about past events and developments.

Historical research attempts to establish facts so as to arrive at a conclusion concerning past events. It is a process by which a researcher is able to come to a conclusion as to the likely truth of an event in the past by studying objects available for observation in the present. Historical research is a dynamic account of the past, which seeks to interpret past events in order to identify the nuances, personalities and ideas that have had an influence on these events.

According to Kerlinger: 'Historical research is the critical investigation of events, developments, and experience of the past, the careful weighing of the evidence of the validity of sources of information of the past, and the interpretation of the weighed evidence.'

According to Gay (1981): 'Historical research is the systematic collection and objective evaluation of data related to past occurrences in order to test hypotheses concerning causes, effects, or trends of those events which may help to explain present events and anticipate future events.'

Therefore, it can be concluded that true historical research is a process of reconstructing the past through systematically and objectively collecting, evaluating, verifying and synthesizing evidence relating to the past events to establish facts and defensible conclusions, often in relation to particular hypotheses (if appropriate), to arrive at a scholarly account of what happened in the past.

Nature and Value of Historical Research

The main aim of historical research is to obtain an exact account of the past to gain a clearer view of the present. Historical research tries to create facts to arrive at conclusions concerning past events. It is usually accompanied by an interpretation of these events at the end of their relevance to present circumstances and what might happen in the future. This knowledge enables us, at least partially, to predict and control our future existence.

- Historical research as many other types of research, includes the delimitation of a problem, formulating hypothesis or tentative generalization, gathering and

analysing data, and arriving at conclusions or generalizations, based upon deductive-inductive reasoning. However, the historian faces greater difficulties than researchers in any field.

- The job of the historian becomes more complicated when he derives truth from historical evidence. The major difficulty lies in the fact that the data on which historical facts are based cannot be substantiated and is relatively inadequate.

- It may be difficult to determine the date of occurrence of a certain historical event partly because of changes brought in the system of calendar and partly due to incomplete information. The historian lacks control over both treatment and measurement of data.

The historical method has taken two principal forms. In the first form, the method concentrates upon the issues of origin, development and transformation of social institutions, societies and civilizations. In the other form, the historical changes of social structure and types of society are investigated and compared i.e., both causal explanation and historical interpretation takes place. In sociological research, the comparative (or cross-cultural) method is based on the assumption that a society (or a social system) cannot be fully understood without comparing it with other societies or systems.

Types of Historical Research

The various types of historical research are:

- **Legal research:** It is of immense value and interest to educational administrators. It seeks to study the legal basis of social institutions run by different religions and castes, central and state schools, school finance, etc. But this type of research need special training in the field of law. Anybody without this training is not competent to do this type of research.

- **Biographic research:** It aims at determining and presenting truthfully the important facts about the life, character and achievements of famous and important figures in history.

- **Studying the history of ideas:** This involves the tracing of major philosophical or scientific thoughts from their origins through their different stages of development. It aims at tracing changes in popular thought and attitudes over a given period of time.

Advantages and Disadvantages of Historical Research

The advantages of historical research are:

- The researcher is not physically involved in the situation under study.

- No danger of experimenter-subject interaction.

- Documents are located by the researcher, data is gathered, and conclusions are drawn out of sight.

- Historical method is much more synthetic and eclectic in its approach than other research methods, using concepts and conclusions from many other disciplines to explore the historical record and to test the conclusions arrived at by other methodologies.

- Perhaps more than any other research method, historical research provides librarians with a context. It helps to establish the context in which librarians carry

out their work. Understanding the context can enable them to fulfil their functions in society.

- It provides evidence of ongoing trends and problems.
- It provides a comprehensive picture of historical trends.
- It uses existing information.

Historical research suffers from several limitations, some are natural due to the very nature of the subject and others extraneous to it and concerning the capabilities of the researcher.

- Good historical research is a slow, painstaking and exacting process. An average researcher finds it difficult to cope with these requirements.
- Historical research requires high level of knowledge, language skills and art of writing on the part of the researcher.
- Historical research requires a great commitment to methodological scholarly activity.
- Sources of data in historical researches are not available for the direct use of the researcher and historical evidence is, by and large, incomplete.
- Interpretation of data is very complex.
- It is difficult to predict the future, through historical research.
- Scientific method cannot be applied to historical evidence.
- Modern electronic aids (like computers) have not contributed much towards historical research.
- It is not possible to construct 'historical laws' and 'historical theories'.
- Man is more concerned with the present and future and there is a tendency to ignore the past and the importance of historical research.
- It is time-consuming.
- Resources in historical research are scarce.
- Data in historical research can be contradictory.
- The historical research may not be conclusive.
- Gaps in data cannot be filled as there are no additional sources of information available in historical research.

A historian can generalize but not predict or anticipate, can take precautions but not control; can talk of possibilities but not probabilities.

Process of Historical Research

Historical research includes the delimitation of a problem, formulating hypothesis or tentative generalizations, gathering and analysing data, and arriving at conclusions or generalizations based upon deductive-inductive reasoning. However, according to Ary, *et al*., (1972) the historian lacks control over both treatment and measurement of data. The historian has relatively little control over sampling and he has no opportunity for replication. As historical data is the closed class of data located along a fixed temporal locus, the historian has no choice of sampling his data. He is supposed to include every type of data that comes his way. Historical research is not based upon experimentation, but upon reports of observation, which cannot be authenticated. The historian handles

data which are mainly traces of past events in the form of various types of documents, relics, records and artefacts, which have a direct or indirect impact on the event under study.

In deriving the truth from historical evidence, the major difficulty lies in the fact that the data on which historical research is based are relatively inadequate. It may be difficult to determine the data of occurrence of a certain historical event partly because of changes brought out in the system of calendar and partly due to incomplete information. Historical research attempts to establish facts to arrive at a conclusion concerning the past events.

Steps in Historical Research

The steps involved in undertaking a historical research are not different from other forms of research. But the nature of the subject matter presents a researcher with some peculiar standards and techniques. In general, historical research involves the following steps:
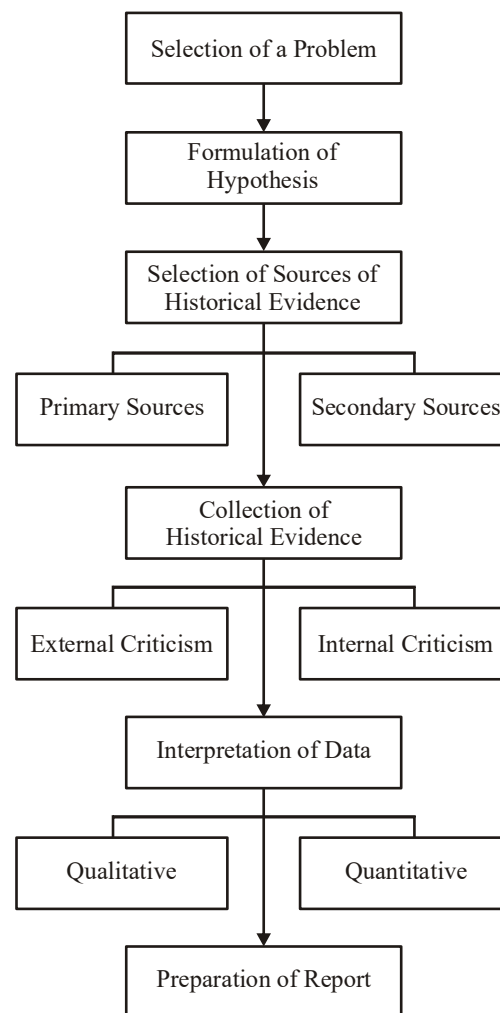
Fig. 4.1 Steps in Historical Research

Step1: The first step is to make sure the subject falls in the area of the history of society. One topic could be the study of the various social systems and how they have changed with the passing of time. The researcher may be interested in a historical investigation of

those aspects of society that have not been touched upon by any studies yet. Moreover, the researcher may be interested in re-examining the validity of current interpretations of certain historical problems which have already been studied.

Step 2: This necessitates that a thought is given to the various aspects of the problem and various dimensions of the problem are identified. Hypothesis also needs to be formulated. The hypothesis in historical research may not be able to be tested, they are written as explicit statements that tentatively explain the occurrence of events and conditions. While formatting a hypothesis, a researcher may formulate questions that are most appropriate for the past events he is investigating. Research is then directed towards seeking answers to these questions with the help of the evidence.

Step 3: Collection of historical evidence involves following two sub-steps:

 (i) Selection of sources of historical evidence

 (ii) Cutting out the historical evidence from them

Historical evidence is hidden broadly in two types of historical sources and is useful to the researcher in many respects. The primary sources, however, are closest to the researcher's heart and kept at the highest pedestal.

Step 4: Historical evidence collected must be truthful; hence for establishing the validity of these sources, the dual processes of external and internal criticism are used. External criticism is undertaken to establish the authenticity of the documents of source, correctness of author or builder, data or period to which it belongs, etc. Internal criticism is done to judge the correctness of the contents of sources.

Step 5: Though statistical testing of hypothesis is not possible, the relationship among various facts still needs to be established, and synthesis and integration of the facts in terms of generalization needs to be done.

Three strategies are used:

 (i) **Generic Analysis:** Identifies the essential meanings of a concept and isolates those elements that distinguished the concept from other words.

 (ii) **Differential Analysis:** Is used when a concept means to have more than one standard meaning and the basis for differentiating between meanings is unclear.

 (iii) **Conditions Analysis:** Involves identification of the context condition in which it can be safely said that the concept was present. Such conditions are rejected, revised and new conditions added.

In this type of investigation, the researcher must be very cautious while dealing with the 'cause and effect' relationship.

Step 6: The final stage of the study is the preparation of a systematic and comprehensive report. It is not just the data which is of significance in such a study. Of prime relevance are the ideas and insights of the researcher, particularly his assessment of the interaction between the data and the ideas that are used to explain the data.

Sources of Data in Historical Research

In this section, we discuss the three sources of data in historical research: (i) Primary sources, (ii) Secondary sources, and (iii) Tertiary sources.

 (i) **Primary Sources:** Primary sources are eye witness accounts and are the only firm basis of historical enquiry. Good, Barr and Scates (1941) have called them the 'first witness to a fact'.

Direct observation, and reporting or recording of the same, comprise primary sources of data. These provide first-hand information about events that have occurred in the past. Some of the main types of primary sources are:

- Verbal narratives written by the participants or observers. These may take various forms, such as official minutes or records, biographies, letters, contracts, deeds, wills, certificates, magazines or newspaper accounts, maps, pictures, books, etc.

- Personal primary sources which are typically a person's observation of events in which he has participated.

- Physical artefacts like museum collections, artefacts in historical spots such as remains or relics, as well as various other types of institutions.

- Mechanical artefacts represent information that is observed through the medium of non-natural items like photographs, films, and audio cassettes.

(ii) **Secondary Sources:** Secondary sources of data basically refer to information that is obtained second-hand. For instance, the person from whom information is obtained neither participated nor witnessed the events. Some types of secondary sources are magazine and newspaper articles, interviews referred to in the articles, research papers, research reports, documentaries, etc.

While carrying out historical studies, primary sources of data have highest credibility when they are used to authenticate presented facts. However, second-hand information that is available, should also be considered in order to develop a more holistic view.

Advantages of Secondary Sources

(a) They may acquaint a researcher with major theoretical issues in his field and to the work that has been done in the area of study.

(b) They may suggest possible solutions of the problem and working hypotheses and may introduce the researcher to important primary sources.

Some type of data may be primary sources for some purposes and secondary sources for another. For example, a high school textbook in Indian history will be ordinarily classified as secondary source, but the book would be a primary source of data if one were making a study of the changing emphasis on national integration in high school history textbooks.

(iii) **Tertiary Sources:** These sources include bibliographies, catalogues and indexes that guide a researcher to primary and secondary sources.

Evaluation of Data

The main feature of historical research is the evaluation of historical data. The backbone of historiography is the authenticity of data collected through different sources. Even when the data are collected through different sources, doubts can be raised about their validity, reliability and relevance. The process of judging validity, reliability and relevance of data is carried out through two devices viz., (a) External criticism and (b) Internal criticism.

(a) External Criticism

External criticism is also known as lower criticism. It involves testing the sources of data for integrity, i.e., every researcher must test the information received to ensure that any source of data is in fact what it seems to be. External criticism helps to determine whether it is what appears or claims to be and whether it reads true to the original so as to save the researcher from being the victim of fraud. On the whole, the general criteria followed for such criticism depends on:

- A good chronological sense, a versatile intellect, common sense, an intelligent understanding of human behaviour, and plenty of patience and persistence on the part of the researcher.
- Recent validation of the quality of the source.
- A good track record of the source.

This information may be found in relevant literature. Thereafter, these literary sources can be verified for genuineness of content by verifying signatures, handwriting, writing styles, language, etc. Further, material sources of information can be verified through physical and chemical tests on the ink, paint, paper, cloth, metal, wood, etc.

(b) Internal Criticism

After the integrity of the data sources are established, the actual data content is subject to verification—this process is known as the internal criticism of the data. It is also called higher criticism which is concerned with the validity, truthfulness, or worth of the content of document.

At the outset, the information obtained through a particular source is examined for internal consistency. The higher the internal consistency, the greater the accuracy. The researcher should establish the literal as well as the real meaning of the content within its historical context.

This is followed by an evaluation of the external consistency of the data. This is important because, although the authorship of a report is established, the report may comprise distorted pictures of the past. For verifying that the content is accurate, the researcher should firstly compare the information received through two independent sources, and secondly match new information obtained with the information already on hand which has been tested for reliability. Fox (1969) suggested three major principles that need to be followed in order to establish external consistency of the data: (i) Data from two independent sources to be matched for consistency, (ii) Data must have been obtained from at least one independent primary source, and (iii) Data should not be gathered from a source that has a track record of providing contradictory information. It is recommended that the researcher apply his professional knowledge and judgment to make a final evaluation in case it is not possible to find matching information from two comparable sources.

The following series of questions have been listed by Good, Barr and Scates (1941) to guide a researcher in the process of external and internal criticism of historical data:

- Who was the author, not merely what his name was but what his personality, character and position were like, etc.?
- What were his general qualifications as a reporter—alertness, character and bias?

- What were his special qualifications as a reporter of the matters here treated?

- How was he interested in the events related?

- Under what circumstances was he observing the events?

- Had he the necessary general and technical knowledge for learning and reporting the events?

- How soon after the events was the document written?

- How was the document written, from memory, after consultation with others, after checking the facts, or by combining earlier trial drafts?

- How is the document related to other documents?

- Is the document an original source—wholly or in part? If the latter, what parts are original, what borrowed? How credible are the borrowed materials? How accurately is the borrowing done? How is the borrowed material changed and used?

Perpetually, the researcher needs answers for all these questions and, therefore, he has to depend, somewhat, upon evidence he can no longer verify. At times, he will have to rely on the inferences based upon logical deductions in order to bridge the gaps in the information.

Purpose of Historical Research

Historical research is carried out to serve the following purposes:

- **To discover the context of an organizational situation:** In order to explore and explain the past, a historian aims to seek the context of an organization/a movement/ the situation being studied.

- **To answer questions about the past:** There are many questions about the past to which we would like to find answers. Knowing the answers can enable us to develop an understanding of past events.

- **To study the relationship of cause and effect:** There is a cause and effect relationship between two events. A historian would like to determine such a relationship.

- **To study the relationship between the past and the present:** The past can often help us get a better perspective about current events. Thus, a researcher aims to identify the relationship between the past and the present, whereby we can get a clear perspective of the present.

- **To reorganize the past:** A historian reconstructs the past systematically and objectively, reaching conclusions that can be defended.

- **To discover unknown events:** There are some historical events that could have occurred in the past that are not known. A historian seeks to discover these unknown events.

- **To understand significance of events:** There may be significant events that could have been responsible for shaping the organization/movement/situation/ individual being studied by a historian.

- **To record and evaluate the accomplishments of individuals, institutions and other kinds of organizations:** Historians are greatly interested in recording and evaluating the accomplishments of leading individuals and different kinds of organizations including institutions and agencies as these influence historical events.

- **To provide understanding of the immediate phenomenon of concern:** A researcher may be investigating a phenomenon. Historical perspective can enable him to get a good understanding of the immediate phenomenon of concern.

Problems in Historical Research

The problems encountered in historical research are:

- **Amount of data:** Often, it is difficult to decide as to how much data is sufficient to reach meaningful conclusions.
- **Selection of data:** A historian must avoid improper or faulty selection of data which may be the result of relying too heavily on some data, ignoring other data, etc. This can result in a bias in the study.
- **Evaluation of historical data and their sources:** Inadequate evaluation of data and their sources can lead to misleading results.
- **Synthesis of data into a narrative account:** Due to the very nature of historical research, it becomes most fruitful, if a researcher is able to successfully synthesize or integrate the facts into meaningful generalizations. Thus, a failure on the part of a researcher to interpret data adequately is considered a serious setback.

There are four problems at the stage of synthesis and in report preparation as given below:

(i) The ability to establish causation from interrelated events is the first problem. It is incorrect to infer that one event caused the other just because they occurred simultaneously.

(ii) The second problem is to accurately define the keywords and terms such that ambiguity is avoided and the correct connotation is established.

(iii) Distinguishing between evidence indicating how people should behave *vs.* how they did behave is the third problem.

(iv) The fourth problem involves distinguishing between the intent and the outcome. This means that educational historians ensure that the consequences of some activity or policy were actually the intended consequences.

Historical synthesis and interpretation are considered an art, which is subjective in nature. This raises a serious problem of subjectivity. 'Historical synthesis is necessarily a highly subjective art. It involves the intuitive perception of patterns and relationships in the complex web of events, as well as the art of narrative writing. Explanations and judgments may be called for, that will involve the historian's own personality, experience, assumptions, and moral values. Inevitably there are personal differences among historians in this respect, and prolonged academic disputes among historians of different schools or nationalities have arisen over practically every event. The initial reduction of complex events of the recent past to comprehensible pattern is particularly difficult and subjective…'. Since the very process of writing a narrative is a human one, therefore, total objectivity is almost impossible. As a consequence, bias and distorting of facts to fit preconceived notions or ideas are not unusual. It may also be kept in mind that historical conclusions are conditioned by place, time and the author. In order to overcome some of these inherent weaknesses, the writer must clearly indicate the underlying assumptions in his approach. In case he belongs to a particular school of thought, the same must be stated clearly.

NOTES

Check Your Progress

1. What is another name for empirical research?
2. State the aim of historical research.
3. What is conditions analysis?
4. Give some examples of tertiary sources of data in historical research.

4.4 EXPLORATORY RESEARCH

Exploratory research is essentially a method of collecting data either by using available secondary sources, or by using primary data that has been collected by unstructured method. Exploratory research is generally carried out to know more about the problem, particularly if the available information is insufficient to proceed further. It is primarily qualitative in nature, and the purpose of the entire research design is to get more elaborate information. Therefore, methods such as focus groups, depth interviews, experience interviews, and projective techniques are used to gather data. These methods typically entail obtaining detailed information about relevant aspects of the problem in hand from a small number of respondents. Since the number of respondents is less, and the data is loosely structured, its reliability cannot be tested. However, rich information is obtained by probing.

The primary disadvantage of the exploratory design is that the sample size is small. Therefore, the results of the research cannot be generalized for the entire population. A further research (usually of a descriptive nature) will have to be undertaken to reach conclusive results.

Usually, exploratory design is used when additional information is needed to define the problem more clearly. It is also used when the researcher wants to gather attitudinal or perceptual data regarding the respondents.

As stated earlier, exploratory research involves the use of qualitative-research methods. While quantitative research methods involve the use of structured research procedures by selecting an appropriate sample, size that is representative to the population under study, qualitative research primarily uses a loosely structured research technique. The basic purpose of the qualitative research is to enhance the existing knowledge of the marketer regarding the problem at hand. When secondary data is not adequate to clearly formulate the problem, or state the objectives of the research, an exploratory research becomes necessary before proceeding further.

Data is obtained by probing the respondent. Depth interviews and focus groups are typical methods of obtaining information in qualitative research. The information obtained through these methods is difficult to assimilate. This is because the information is obtained by probing, or by asking open-ended questions. Therefore, codification of the responses is difficult. The interpretation of data obtained from this form of research requires experts.

Advantages

- Qualitative researches use smaller samples. Therefore, they can be completed at a lesser cost and on time.

- The information obtained from this type of research method is rich and extremely useful in understanding the problem better. In-depth information can be collected from respondents.

- Most of the information obtained through qualitative research gives insights about consumers to the researcher. Additionally, since the researcher does not go with any preset structured format of presenting questions to the respondent, he comes across information that he might not have anticipated, but is useful for the problem in hand. This is absent in case of a highly structured, quantitative research study.

- Methods like observation allow the researcher to see and understand the actual behaviour of consumers. Researchers can, for instance, observe consumers in situations such as in a retail store.

Disadvantages

- Qualitative research suffers from some drawbacks such as the inability of the researcher to draw conclusions based on the data obtained. This is because of the limitation of the sample size, and the nature of the research, as it is unstructured. Unstructured research makes it extremely difficult to classify the data obtained from respondents, as the information does not have a common structured format, which can be used for classification.

- It is also more expensive to conduct qualitative research on a larger scale as compared to quantitative research.

- Since statistical tools cannot be used in interpreting data, it becomes difficult to understand the extent of differences among respondents. In quantitative research, the magnitude of differences can be gauged.

- It is also difficult to find researchers who are experts in conducting qualitative research studies, and have the ability to accurately interpret the information obtained.

Experience Surveys

An experience survey involves the use of experts to obtain information about a problem at hand. The researcher based on his judgement identifies the experts. The expert is selected based on his ability to guide the researcher in the right direction by providing appropriate information pertaining to an industry, product, market or customer segment.

Projective Techniques

As the name indicates, the respondent is asked to project his thought process about a particular object in a certain manner indicated by the researcher. The technique finds its underpinnings in psychology. These techniques are, for instance, particularly useful in the area of motivational research. Some of the projective techniques are:

- *Word association test:* This test involves the use of words to elicit appropriate responses from the respondents. The respondent reads a word, and writes the first word that comes to his mind. That word is presumed to be most strongly associated with the given word stimulus for the respondent. Thus, the researcher can use the word association test for designing advertising.

- *Sentence completion test:* Respondents are asked to complete incomplete sentences about the desired object. The test is often used to understand customers' choice criteria, and brand image.

- *Picture tests:* Respondents are given pictures and asked to write a story based on the given picture. These stories are used to understand customer perceptions, attitudes or preferences.

Depth Interview

A depth interview involves the researcher to conduct a detailed interview with a selected respondent. The method extensively employs the use of probing to elicit responses. The information is obtained by using a unstructured questionnaire, that has a detailed outline

of all the information that is required by the researcher from the respondent. However, the questionnaire allows for probing for seeking clarifications from the respondent. The basic purpose of the depth interview is to collect detailed information. The objective of using this method is to make the respondent talk as much about the object as possible, so that his implicit feelings, attitudes, perceptions or motivations are revealed. These are not possible with a structured questionnaire. The interviewer constantly asks the respondent questions such as 'why,' 'how,' 'why not the other option,' etc.

The interviewer will be able to achieve the desired results only if he is skilled in interviewing. He must put the respondent at ease, and establish a connection with him in the beginning. This helps the respondent to open up, thus, lending more information to the interviewer. The interviewer should also be very clear about the required information. Therefore, if the respondent moves away from the main object of discussion, the interviewer has to bring him back to the track. Therefore, the interviewer should be a very good listener.

The main advantage of the depth interview is that the information obtained is unparalleled in terms of its detail. In addition, the method is extremely flexible.

The biggest disadvantage of the method is the cost involved. In addition, the method is heavily dependent on the skill level of the interviewer. Thirdly, the information obtained cannot be generalized.

Focus Group

A focus group is formal gathering of respondents in groups of about eight to twelve members each. Each group discusses certain issues posed to them by a group moderator. The moderator has an unstructured questionnaire that details the information to be elicited from the participants. Each group discussion can last for an hour to about three hours depending on the amount of information solicited and the extent of participation of the respondents.

The advantages of the focus group are:

- The focus group solicits in-depth information. It resembles the depth interview in this sense. This is also the main advantage of the focus group. In addition, there can be opinions and counter-opinions that form the outcome of the discussion. Interesting and fruitful revelations can come up during the course of the focus groups.

- The focus group can be used to gain clearer understanding about the problems that require decision-making.

- It also enables the marketer to generate new ideas for advertising, new products, or for improving existing products.

- The marketer uses this method to elicit comparisons between his product and those of his competitors', and understand customer perceptions about his offerings.

- Hidden motivations, perceptions, or attitudes can be brought forth using this method. This, however, requires the use of a skilled focus group moderator.

- He can also understand emerging consumer trends among the target customers, or understand how consumer preferences are changing.

Participants for the focus group should be selected very carefully. They should be representative of the population. For every sub-group (age, gender, income), separate focus groups should be conducted. The participants should be encouraged to voice their

opinions. The role of the moderator is most important in the discussion, as no individual should be allowed to dominate the proceedings, and if any individual is very silent, he should be encouraged to talk. The focus group can be conducted in a central hotel, or a convention centre. The participants can be rewarded for their participation at the end of the session.

The data thus, gathered suffers from the same limitations as the other qualitative data. It cannot be generalized, and is difficult to interpret. However, focus groups can be recorded and observed. This should be done with the consent of the participants. It allows the researcher to not only record the information, but also allows him to observe the nonverbal communication of the participants. The transcripts of the focus groups are thereafter, documented verbatim first, and then results are drawn from it. As in the case of other qualitative researches, an expert must do the data interpretation.

4.5 DESCRIPTIVE RESEARCH

A descriptive research design involves the use of quantitative studies and use of statistical tools to understand the target customers of a company. Usually, this type of research design involves a survey based framework. The purpose of this research design is to reach conclusions about the existing preferences of the target customer, or understand their reactions to changes such as introduction of new products, changes in product features, pricing or communication. Since the sample size of respondents is large the results thus, obtained can be generalized for the population. Managers can draw inferences from the study, and can use the results to formulate marketing strategies. However, descriptive does not establish a cause-effect relationship between variables. It only informs the manager the extent to which one variable can influence the other, as the effect of other variables in the environment has not been controlled. For instance, the manager can find out the extent to which price influence sales, however, he cannot determine if price change is the cause of changes in sales. Generally, descriptive research are used to understand consumer attitudes, perceptions, buying intentions, comparisons among competing brands, choice criteria, and evaluations of any other element of the marketing mix.

One of the most important aspects of any research is time. Some studies are done at a particular point in time. Such a research is referred to as a cross-sectional study. It is called so because the researcher is only attempting to study a slice, or cross-section of a particular phenomenon. Longitudinal study is when a researcher studies a problem over a period of time.

One of the biggest disadvantages of cross-sectional study is that since it involves the completion of the research by the respondents only once at a particular time, it is prone to systematic error (that arises from the data from a single source). This can be eliminated by using longitudinal research design. Cross-sectional designs are also limited in their ability to result in causal inferences. In general, when the theoretical constructs are well developed, when the scales used are heterogeneous, relationship between constructs are strong, the likelihood of intervening events are high, and likelihood of alternative explanations are low, the cross-sectional design should be the preferred method.

Causal Research Design/Experimental Research Design

The main purpose of the causal research is to establish a cause-effect relationship among variables. This type of research design is extremely useful for marketers when they want to establish the exact cause outcomes. For instance, if the marketer wants to

Check Your Progress

5. Mention some of the methods used to gather data in exploratory research.
6. When is exploratory design used?
7. What does an experience survey involve?

understand the exact impact of the increase in prices by 10 per cent on sales of a particular product, he can use experimental research. However, the causal research is extremely difficult and expensive to conduct. There are several variables in the environment that cannot be controlled. In order to conduct an accurate causal research, it is extremely important that all variables other than the ones under study be kept constant, so that the marketer can understand exactly how much one variable influences the other. In real market circumstances, this is not possible. Even if it were made possible under artificial laboratory conditions, the results would not be accurate. Therefore, causal designs are used rarely in marketing research.

A descriptive research design uses the survey based research technique to arrive at conclusions that can enable marketers to take appropriate decisions. This type of research design is best suited for evaluating marketing decisions, particularly in ascertaining customer perceptions and acceptance. The results of the research can be generalized for the population as a large sample size is chosen for doing the research. The sample chosen is representative of the population.

The qualitative (exploratory) and quantitative (descriptive) research designs are not mutually exclusive. In fact, many a times, the exploratory research precedes the descriptive research. The descriptive research is used after the nature of the problem has been well understood, or more information has been collected, so that the research objectives can be clearly formulated. Thereafter, the descriptive research begins, as for a descriptive research design, clear problem or objective formulation is absolutely necessary.

Advantages

One of the biggest advantages of the survey design is the use of the large sample, and the ability of the manager to draw results from the study. A quantitative research can also be generalized for the population, unlike the qualitative research, out of which conclusions cannot be drawn. Another important advantage is the low cost of conducting the research as compared to qualitative research.

It is also easy to carry out the survey, as the survey instrument (usually a questionnaire) is structured. Therefore, the data obtained is standardized, and there is no need for an expert to interpret even the raw data. Respondents also find it easier to answer questions that are well formed and structured.

Rich analysis of the data is extremely important for the marketer. He can analyse data across various socio- economic variables, and use sophisticated statistical package to draw conclusions.

Disadvantages

A very important problem with survey research is the sampling and non-sampling errors that arise in the survey process. Unless the sampling plan is accurate, and the sample is sufficiently representative, the results drawn from the research would be completely erroneous and misleading. The survey instrument should be reliable and validated to ensure accuracy of data collected.

Survey researches also involves highly structured questionnaires, thus, eliminating the possibility of probing the respondent. Therefore, deep insights into the customer's behaviour cannot be obtained. Most of the times, even the questions are close-ended to facilitate codification and analysis of data. This, however, restricts the responses of the respondent.

All survey methods rely on the veracity of the respondent. In case the respondent is not telling the truth, there is not much that the surveyor can do about it. Also, most surveys rely on the memory of the respondent. Therefore, surveys should not solicit information that happened long back.

Mall-Intercept Survey

Due to the amount of money and time that needs to be spent for a survey method, a mall intercept survey is becoming more acceptable method. In this, an interviewer stands in a shopping mall, and conducts face- to- face interviews with suitable respondents who visit the mall.

The method is definitely less expensive than the self- administered survey, as the respondents are captive at a single physical location. It is thus, convenient for the interviewer as well. It also enables the administration of the questionnaire at the point of purchase, and may elicit better responses from the respondents.

However, the biggest disadvantage is that the customers who visit the mall may not be the best or the only respondents representative of the population. Also, they may be very inclined to answer questions while they are in the mall.

Telephone Survey

The telephone is a cost-effective method of conducting surveys. This is the primary advantage of the phone-based survey. It is also faster, and a large number of respondents can be reached using the phone in a relatively shorter time at low cost. It is also possible to monitor the calls made by the interviewers. The researcher can call back in case of unavailability of the respondent. This process is also cheaper and faster than a face- to- face survey. In case respondents are scattered in far flung areas with access to phone facilities, this may be the best survey method available. Also, the phone surveys are most suitable for those questions where respondents may not want to disclose their identity.

However, the primary disadvantage of the phone survey is that a lengthy survey is not possible over the phone. The respondent may disconnect the phone. Also, it is not possible to use show cards, or pictures while conducting the survey. Even if there are many options in a question, the respondent may be forced to recollect all of them. Therefore, only short questions and instruments can be suitably used in a phone-based survey. Telemarketing may also be considered to be a source of intense displeasure by many respondents.

If the phone-based survey is used, first the respondents need to be selected by using a telephone directory. The selection of the respondents should be random. Usually a random selection of the telephone numbers of respondents can be done by using an appropriate software.

Self-Administered Survey

In a self-administered survey, the respondent reads the questions himself, and answers the survey instrument. The advantage for the interviewer is that he has to spend less time with the respondents. Also, the interviewer's bias gets reduced if the respondent fills up the instrument himself. However, the disadvantage of this aspect is that the interviewer does not have the opportunity to clarify any doubts that the respondent might have while filling up the survey instrument.

Direct-Mail Survey

One of the methods of a self-administered survey is the direct-mail survey. In this, the questionnaires are mailed to the respondent. The respondent upon receiving the survey instrument fills it up and sends it back to the researcher. Direct mail is the best method of research in case a strong database of relevant respondents and their addresses are available.

The main problem with the direct-mail survey is that the response rate is very low. A response rate of around 2 per cent is considered to very good in case of a direct-mail survey. The main reason for the low response rate is the respondent's inertia to fill in the questionnaire. However, researchers have tried to overcome this drawback by offering an incentive to the respondent, for instance, in the form of a discount coupon. Also, they can be sent a self- addressed envelope in which the respondent can send back the completed questionnaire. This measures can increase the response rate to some extent.

Survey Methods

Online Survey: An online survey is the latest method of conducting research. The online survey uses the Internet as the medium for receiving responses to the survey instrument. The Internet based survey can be done at low cost and is less time consuming to undertake the survey. The basic problem with the online survey is that it is not possible to find out whether the person taking the survey is a correct respondent. But this problem can be overcome by sending the questionnaires to identified respondents.

While selecting the appropriate survey-research method, the researcher must consider several factors such as the availability of time and money. The difficulty of the questionnaire is an important consideration to find out if a self- administered survey or a telephonic survey would work, or a person administered survey will have to be undertaken. The length of the questionnaire is also an important determinant of the same decision.

Person-Administered Surveys: In the person-administered survey, the survey instrument or the questionnaire is administered by field interviewers who received training about the procedure. The interviewers are trained to identify the right respondent, elicit his interest in answering the questionnaire, filling up the data properly and to avoid any interviewer biases.

4.6   EXPERIMENTAL RESEARCH

Experimentation is the method wherein the cause-effect relationship is examined between variables. Descriptive research does not involve the establishment of cause and effect. It involves the understanding of the impact of the independent variable on the dependent variable. Experimental designs also use independent and dependent variables. However, the independent variable is used to directly measure the impact of its variation on the outcome of the dependent variable. The difference between the experimental research design and the descriptive research design is that in experimental research design, only the independent variable is varied, while all other variables that can cause variations are kept constant. Thus, the marketer can exactly judge the cause-effect relationship. Whereas in descriptive research, the independent variable only impacts the dependent variable, though the researcher knows that other variables in the environment can also cause variations in the dependent variable.

The researcher has to keep other variables constant in order to understand the causal relationship. The variables that are kept constant are called control variables. The researcher tries his best to keep the other variables constant in the environment. Some variables cannot be controlled by the researcher. These include any variations in external, uncontrollable factors. These factors are called extraneous variables.

Test marketing is a typical example of an experimental design. Test marketing involves the launch of the new product in one or few geographical areas chosen to be representative of its intended market. A new product is sold into distribution outlets so that performance can be gauged face-to-face with rival products. The product is promoted as it would be in a national launch and consumers are asked to choose it against competitors' products as they would if the new product went national. By projecting test marketing results to the full market, an assessment of the new product's likely success can be gauged. Test marketing does have problems. Test towns and areas may not be representative of the national market and thus, sales projections may be inaccurate. Competitors may invalidate the test market by giving distribution incentives to stock their product, thereby denying the new product shelf space. Test markets need to be long enough to measure the repeat purchase rate for the product. This can mean a delay in national launch stretching to many months and years. In the meantime, competitors that are more aggressive can launch a rival product nationally and therefore, gain pioneer advantage. Getting the cooperation of the distributors is important. Sometimes, they refuse to take part in test marketing activities or charge heavy fees.

The advantage of test marketing is that the information provided by test marketing facilitates the go / no go national launch decision. Sometimes, a number of test areas are used with different marketing mix combinations to predict the most successful launch strategy. Its purpose is to prevent a costly and embarrassing national launch mistake. Test marketing is commonly used with FMCG goods. For very expensive equipments it is impractical. On global scale, companies roll out products from one country to another. They gain some of the benefits of test marketing in that the lessons learnt from an early launch in a country market can be applied to later launches.

4.7   SUMMARY

- Research can be either fundamental (basic or pure) or action-oriented (applied) research. Fundamental research focuses on finding generalizations and formulating theories. It is the research done for knowledge enhancement; the research which does not have immediate commercial potential; and the research which is done for human welfare, animal welfare and plant kingdom welfare.

- Applied research is designed to solve practical problem of the modern world, rather than to acquire knowledge for the sake of knowledge. Its goal is to improve the human condition. It focuses on analysis and solving social and real life problems.

- Conceptual research is that which is related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones.

- Empirical research relies only on real experiences and observations. It is data-based research and its conclusions can be verified by observations or experiments. It is also called experimental type of research.

- Historical research attempts to established facts so as to arrive at a conclusion concerning past events. The true historical research is a process of reconstructing the past through systematically and objectively collecting, evaluating, verifying and synthesizing evidence relating to the past events to establish facts and defensible conclusions, often in relations to particular hypotheses, to arrive at a scholarly account of what happened in the past.

- Historical research includes the delimitation of a problem, formulating hypothesis or tentative generalizations, gathering and analysing data, and arriving at conclusions or generalizations based upon deductive-inductive reasoning.

- Steps in historical research include: making sure that the subject falls in the area of the history of society, giving thought to the various aspects of the problems and the various dimensions of the problems, collecting historical evidence, checking the authenticity of the historical evidence selected, establishing relationship between various facts and preparing a systematic and comprehensive report.

- There are three sources of data in historical research: primary, secondary and tertiary.

- The process of judging validity, reliability and relevance of data in historical research is carried out through two devices: external criticism and internal criticism.

- The major problems in historical research are: amount of data, selection of data, evaluation of historical data and their sources and synthesis of data into a narrative account.

- Exploratory research is a method of collecting data either by using available secondary sources, or by using primary data that has been collected by unstructured method.

- The methods such as focus groups, depth interviews, experience interviews, and projective techniques are used to gather data.

- An experience survey involves the use of experts to obtain information about a problem at hand. The researcher based on his judgement identifies the experts.

- In projective techniques, the respondent is asked to project his thought process about a particular object in a certain number indicated by the researcher.

- A depth interview involves the researcher to conduct a detailed interview with a selected respondent. The method extensively employs the use of probing to elicit responses.

- A focus group is formal gathering of respondents in groups of about eight to twelve members each. Each group discusses certain issues posed to them by a group moderator.

- The descriptive research design involves the use of quantitative studies and use of statistical tolls to understand the target customers of a company.

- The main purpose of causal research is to establish a cause-effect relationship among variables. This type of research design is extremely useful for marketers when they want to establish the exact cause outcomes.

- In a mall-intercept survey, an interviewer stands in a shopping mall, and conducts face-to-fcae interviews with suitable respondents who visit the mall.

- Experimentation is the method wherein the cause-effect relationship is examined between variables. Descriptive research does not involve the establishment of cause and effect. It involves the understanding of the impact of the independent variable on the dependent variable.

4.8 KEY TERMS

- **Fundamental research:** It refers to the research that focuses on finding generalizations and formulating theories.

- **Applied research:** It refers to the research which is designed to solve practical problems of the modern world rather than to acquire knowledge for the sake of knowledge.

- **Historical research:** It is a process by which a researcher is able to come to a conclusion as to the likely truth of an event in the past by studying objects available for observation in the present.

- **Exploratory research:** It is essentially a method of collecting data either by using available secondary sources, or by using primary data that has been collected by unstructured method.

- **Experimental research:** It refers to research activity wherein the manipulation of variables takes place, and the resultant effect on other variables is studied.

- **Descriptive research:** involves the understanding of the impact of the independent variable on the dependent variable.

4.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Empirical research is also called experimental type of research.

2. The main aim of historical research is to obtain an exact account of the past to gain a clearer view of the present.

3. Conditional analysis is a strategy of hypothesis testing which involves identification of the context condition in which it can be safely said that the concept was present. Such conditions are rejected, revised and new conditions added.

4. Some examples of tertiary sources of data in historical research are bibliographies, catalogues and indexes.

5. Some of the methods used to gather data in exploratory research are: focus groups, depth interviews, experience interviews, and projective techniques.

6. An exploratory design is used when additional information is needed to define the problem more clearly. It is also used when the researcher wants to gather attitudinal or perceptual data regarding the respondents.

7. An experience survey involves the use of experts to obtain information about a problem at hand.

8. The descriptive research design involves the use of quantitative studies and use of statistical tools to understand the target customers of a company.

9. The main purpose of the causal research is to establish a cause-effect relationship among variables. This type of research design is extremely useful for marketers when they want to establish the exact cause's outcomes.

10. One of the biggest advantages of the survey design is the use of the large sample, and the ability of the manager to draw results from the study.

11. Experimentation is the method wherein the cause-effect relationship is examined between variables.

12. Test marketing is a typical example of an experimental design.

4.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the limitations of historical research.
2. What are the two types of historical research?
3. Give two advantages of historical research.
4. Write a short note on exploratory design.
5. What are experience surveys?
6. Write a short note on projective techniques.
7. Define descriptive research design.

Long-Answer Questions

1. Explain the concept of conceptual research and empirical research.
2. Discuss the steps involved in historical research.
3. Compare and contrast exploratory, descriptive and causal research designs.
4. Discuss the advantages and disadvantages of the exploratory research design.
5. Discuss the characteristics of focus groups. List at least four characteristics and explain why it is important to consider each of them.

4.11 FURTHER READING

Wrenn, Bruce, David L. Loudon and Robert E. Stevens. 2001. *Marketing Research: Text and Cases*. New York: Routledge.

Kumar, Vineet, David A. Aaker and George S. Day. 1999. *Essentials of Marketing Research*. New Jersey: John Wiley and Sons.

Bhattacharya, D. K. 2003. *Research Methodology.* New Delhi: Excel Books.

Kothari, C.R. 2004. *Research Methodology: Methods and Techniques.* New Delhi: New Age International Publishers.

UNIT 5 TECHNIQUES OF DATA COLLECTION

Structure

5.0 INTRODUCTION

This unit discusses the techniques of data collection. To understand the multitude of choices available to a researcher for collecting the project/ study-specific information, one needs to be fully cognizant of the resources available for the study and the level of accuracy required. To appreciate the truth of this statement, one needs to examine the gamut of methods available to the researcher. The data sources could be either contextual and primary or historical and secondary in nature.

Primary data as the name suggests is original, problem- or project-specific, collected for the specific objectives and needs to be spelt out by the researcher. The authenticity and relevance is reasonably high. The monetary and resource implications of this are quite high and sometimes a researcher might not have the resources or the time or both to go ahead with this method. In this case, the researcher can look at

alternative sources of data which are economical and authentic enough to take the study forward. These include the second category of data sources—namely the secondary data.

Secondary data, as the name implies, is that information which is not topical or research specific and has been collected and compiled by some other researcher or investigative body. The said information is recorded and published in a structured format, and thus, is quicker to access and manage. Secondly, in most instances, unless it is a data product, it is not too expensive to collect. As suggested in the opening vignette, the data to track consumer preferences is readily available and the information required is readily available as a data product or as the audit information which the researcher or the organization can procure and use for arriving at quick decisions. In comparison to the original research-centric data, secondary data can be economically and quickly collected by the decision maker in a short span of time. Also the information collected is contextual; what is primary and original for one researcher would essentially become secondary and historical for someone else.

Qualitative research, thus, is presumed to go beyond the obvious of constructs and variables that are not visible or measurable; rather they have to be deduced by various methods like observation, schedule, questionnaire, interview, projective, case study, focus groups and content analysis. There are a variety of such methods which will be discussed in detail in this unit. However, common premise of all these are that they are relatively loosely structured and require a closer dialogue or interaction between the investigator and the respondent. The information collected is more in-depth and intensive and results in rich insights and perspectives than those delivered through a more formal and structured method. However, since the element of subjectivity is high, they require a lot of objectivity on the part of the investigator while collecting and interpreting the data. Conducting a qualitative research is an extremely skillful task and requires both aptitude and adequate training in order to result in valuable and applicable data.

5.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Explain observation schedule as a tool of data collection
- Discuss the concept of questionnaire as a method of data collection
- Assess schedule as a method of data collection
- Describe how conducting interviews are an important source of data collection
- Explain projective method as a tool of data collection
- Discuss how case studies and content analysis are important sources of data collection

5.2 OBSERVATION

Observations have led to some of the most important scientific discoveries in human history. Charles Darwin used his observations of animal and marine life at the Galapagos Islands to formulate his theory of evolution which he described in *On the Origin of Species*. Today, social scientists, natural scientists, engineers, computer scientists, educational researchers and many others use observation as a primary research method.

The kind of observations one makes depends on the subject being researched. Traffic or parking patterns on a campus can be observed to ascertain what kind of improvements can be made. Clouds, plants or other natural phenomena can be observed as can people, though in the case of the latter, one may often have to ask for permission so as to not violate any privacy issue.

Observation may be defined as '*a process in which one or more persons monitor some real-life situation and record pertinent occurrences*'. It is used to evaluate the overt behaviour of the individual in controlled and uncontrolled situations.

According to Marie Jahoda: 'Observation method is a scientific technique to the extent that it (a) serves a formulated research purpose, (b) is planned systematically rather than occurring haphazardly, (c) is systematically recorded and related to more general propositions than presented as a set of interesting curious, and (d) is subjected to checks and controls with respect to validity, reliability, and precision much as is all other scientific evidence.'

According to Good and Hatt: 'Observation may take many forms and is at once the most primitive and the most modern of research techniques. It includes the most casual, uncontrolled experiences as well as the most exact film records of laboratory experimentation.'

5.2.1 Types of Observation

Observation can be of the following types:

1. **Participant observation:** In the process of 'participant observation', the observer becomes more or less one of the group members and may actually participate in some activity or the other of the group. The observer may play any one of the several roles in observation, with varying degrees of participation, as a visitor, an attentive listener, an eager learner or as a participant observer.

2. **Non-participant observation:** In the process of 'non-participant observation', the observer takes a position where his/her presence is not felt by the group. He/She may follow the behaviour of an individual or characteristics of one or more groups closely. In this type of observation, a one-way 'vision screen' permits the observer to see the subject but prevents the subject from seeing the observer.

Observation may also be classified into the following categories:

- *Natural observation:* Natural observation involves observing the behaviour in a normal setting and in this type of observation, no efforts are made to bring any type of change in the behaviour of the observed. Improvement in the collection of information can be done with the help of natural observation.

- *Subjective and objective observation:* All observations consist of two main components, the subject and the object. The subject refers to the observer, whereas the object refers to the activity or any type of operation that is being observed. Subjective observation involves the observation of one's own immediate experience, whereas the observation involving an observer as an entity apart from the thing being observed is referred to as 'objective observation'. Objective observation is also known as 'retrospection'.

- ***Direct and indirect observation:*** With the help of the direct method of observation, one comes to know how the observer is physically present, in which type of situation is he/she present and then this type of observation

monitors what takes place. Indirect method of observation involves studies of mechanical recording or recording by some other means like photographic or electronic. Direct observation is relatively straightforward as compared to indirect observation.

- *Structured and unstructured observation:* Structured observation works according to a plan and involves specific information of the units that are to be observed and also about the information that is to be recorded. The operations that are to be observed and the various features that are to be noted or recorded are decided well in advance. Such observations involve the use of special instruments for the purpose of data collection that are also structured in nature. But in the case of unstructured observation, its basics are diametrically against the structured observation. In such observations, the observer has the freedom to note down what he/she feels is correct and unlike point of study. This approach of observation is very suitable for exploratory research.

- *Controlled and non-controlled observation***:** Controlled observations are the observations made under the influence of some external forces. Such observations rarely lead to improvement in the precision of the research results. However, these observations can be very effective if these are made to work in coordination with mechanical synchronizing devices, film recordings, etc. Non-controlled observations are made in the natural environment, and unlike to the controlled observation, these observations involve no influence or guidance of any type of external force.

5.2.2 Recording Techniques of Observation

Many different techniques may be employed to study and document a subject's behaviour. The data collection techniques are all accurate but may be suitable for different purposes. While certain methods help gather detailed descriptions of behaviour, certain others facilitate documenting behaviour promptly with bare minimum description.

- **Anecdotal records:** Anecdotal records refer to a few sentences jotted down in a notebook. These sentences pertain to what the subject is engaged in at a particular moment. Only those behaviours that can be seen or heard and that can be counted are documented while creating an anecdotal record.

- **Narrative description:** Narrative description is also known as running behaviour record and specimen record, and is a formal method of observation. When following this technique, one is supposed to record continuously and in as much detail as possible, like what the subject is doing and saying when alone or when interacting with other people. In its methodology, it is similar to anecdotal record but is definitely more detailed. The researcher studies the context setting, the behaviour patterns, and the order in which they take place. The main aim of this technique is to gain an objective description of a subject's behaviour without conjecture, analysis, or assessment.

- **Checklists:** Checklists are usually standardized forms which list specific skills and behaviours based on standard levels or are specifically compiled by the researcher for a particular research study.

- **Interviewing:** In this observation technique, the researching team tries to identify the feelings and beliefs of the subjects, that are not visible through simple

observation. During the process of interviewing, everything that the subject says must be recorded exactly as it is. The interviewer should avoid any kind of editing of the interview subscript.

- **Time sampling:** This method is distinct from others in two ways—it monitors and keeps an account of a few chosen samples of subject's behaviour, and is carried out only during prearranged periods of time. When a behaviour pattern is seen during the specified time interval, it is recorded. This technique therefore helps to gather representative examples of behaviour.

- **Frequency counts:** In some cases, a researcher may be more interested in studying the frequency of an occurrence or behaviour or another pattern, such as how often a consumer buys a particular product or how often an individual started a conversation with a colleague. To get this data, the researcher will have to keep a count of the frequency of the particular behaviour and study how long the behaviour lasts. This is usually done by simply marking an occurrence on a chart each time the behaviour is repeated.

- **Event sampling:** This technique is focused on observing specific behaviours or events in a subject's behaviour pattern. However, it does not take into account the frequency or the length of the recording interval.

5.2.3 Advantages and Disadvantages of Observation

The advantages of observation are as follows:

- The observer through participant observation is able to participate in the group, he has access to a vast body of information.

- The observer is able to give a context to the behaviour displayed by the members of the group, which is far better than the information received through a questionnaire and interview.

- The observer through the method of observation gains information as to what the members of a group actually do rather than what they say, which is considered a more credible information about the behaviour.

- The observation method like participant observation used by ethnographers takes a long period of time. This allows the researcher to dig deeper and uncover varied deep rooted aspects related to the research question.

- Observation is far more flexible than other methods of data collection as it is not rigid and based strictly on a set and pre-defined questions. The researcher has the freedom to be more open minded.

The disadvantages of observation are as follows:

- It is very difficult to establish the validity of observations.

- Many items of observation cannot be defined.

- The problem of subjectivity is involved.

- Observation may give undue stress to aspects of limited significance simply because they can be recorded easily, accurately and objectively.

- Various observers observing the same event may concentrate on different aspects of a situation.

- The observers have little control over the physical situation.
- There are certain situations which observers are not allowed to observe, and are expected to produce an accurate account.
- It may not be feasible to classify all the events to be observed.
- Observation is a slow and laborious process.
- There may be lack of agreement among the observers.
- The data to be observed may be unmanageable.
- Observation needs competent observers and it may be difficult to find them.
- Observation is a costly affair. It involves lot of expenses on travelling, staying at the places where the events are taking place and purchasing sophisticated equipment to help in observation.

5.2.4 Characteristics of Observation for Research

The characteristics of observation for research are as follows:

- Observation schedule should be specific.
- The steps should be systematic.
- It should be quantitative.
- It should be recorded immediately.
- It should be made by experts.
- Schedule should be scientific. We should be able to check and substantiate the results.

Jennifer Symonds gives a list of nine essential characteristics of good observation, which are as follows:

- Good eyesight
- Alertness
- Ability to estimate
- Ability to discriminate
- Good physical condition
- An immediate record
- Good perception
- Freedom from preconceptions
- Emotional disinterest

Planning Administration Aspect of Observation

This includes the following:

- Securing an appropriate group of persons to observe
- Deciding and arranging any special conditions for the group
- Determining the length of each observation period, the interval between periods and the number of periods

Points to be Considered while Defining the Activities

These are as follows:

- Inclusion of those activities which are true representatives of the general category one is studying
- Defining those activities very carefully.

While arranging for the record, the following points should receive attention:

- Deciding the form for recording so as to make note-making easy and rapid
- Deciding the use of appropriate symbols, abbreviations and some use of shorthand

One can train oneself by:

- Training oneself to observe others as perception improves with practice
- Studying manuals that list observation techniques

Planning Effective Observation

This includes the following:

- Sampling to be observed should be adequate; appropriate group of subjects should be there
- Units of behaviour should be defined as accurately as possible
- Method of recording should be simplified
- Detailed instructions may be given to observers to eliminate the difference in the perspectives of observers
- Too many variables may not be observed simultaneously
- Excessively long periods of observation without interspersed rest periods should be avoided
- Observers should be fully trained
- Observers should be well equipped
- Conditions of observation should remain constant
- Number of observations should be adequate
- Records of observation must be comprehensive
- Length of each observation period, interval between periods, and number of periods should be clearly stated
- Interpretations should be carefully made

5.3 QUESTIONNAIRE TOOLS

A questionnaire is 'a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event'. Although they are often designed for statistical analysis of the responses, this is not always the case. This method was the invention of Sir Francis Galton. Questionnaire is used when factual information is desired. When opinion rather than facts are desired, an opinionative or attitude scale is

---

Check Your Progress

1. Define observation.
2. What are the main components of the observation method?
3. List two disadvantages of the observation method of data collection.

---

used. Of course, these two purposes can be combined into one form that is usually referred to as 'questionnaire'.

Questionnaire may be regarded as a form of interview on paper. The procedure for the construction of a questionnaire follows a pattern similar to that of the interview schedule. However, because the questionnaire is impersonal, it is all the more important to take care of its construction.

A questionnaire is a list of questions arranged in a specific way or randomly, generally in print or typed and having spaces for recording answers to the questions. It is a form which is prepared and distributed for the purpose of securing responses. Thus, a questionnaire relies heavily on the validity of the verbal reports.

According to Goode and Hatt, 'in general, the word questionnaire refers to a device for securing answers to questions by using a form which the respondent fills himself.'

Barr, Davis and Johnson define questionnaire as, 'questionnaire is a systematic compilation of questions that are submitted to a sampling of population from which information is desired' and Lundberg says, 'fundamentally, questionnaire is a set of stimuli to which literate people are exposed in order to observe their verbal behaviour under these stimuli.'

5.3.1 Types of Questionnaire

Figure 5.1 depicts the types of questionnaires that are used by researchers.
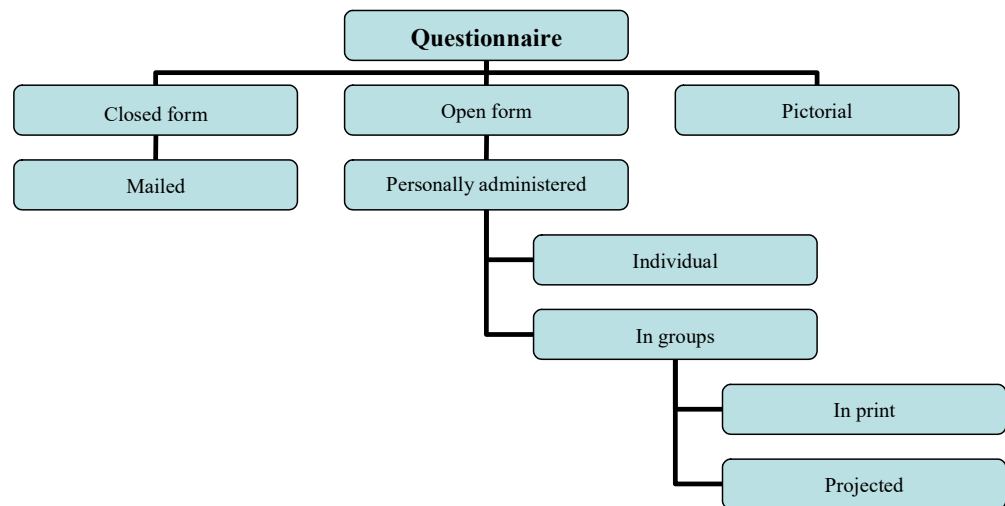
***Fig. 5.1** Types of Questionnaires*

Commonly used questionnaires are:

1. **Closed form:** Questionnaire that calls for short, check-mark responses are known as closed form type or restricted type. They have highly structured answers like mark a 'yes' or 'no', write a short response or check an item from a list of suggested responses. For certain types of information, the closed form questionnaire is entirely satisfactory. It is easy to fill out, takes little time, keeps the respondent on the subject, is relatively objective and is fairly easy to tabulate and analyse.

These types of questionnaires are very suitable for research purposes. However, construction of such a type of questionnaire requires a lot of labour and thought. It is generally lengthy as all possible alternative answers are given under each question.

2. **Open form:** The open form or unrestricted questionnaire requires the respondent to answer the question in their own words. The responses have greater depth as the respondents have to give reasons for their choices. The drawback of this type of questionnaire is that not many people take the time to fill these out as they are more time consuming and require more effort, and it is also more difficult to analyse the information obtained.

No alternative or plausible answers are provided. The open form questionnaire is good for depth studies and gives freedom to the respondents to answer the questions without any restriction.

Limitations of open questionnaire are as follows:

- Difficult to fill out

- Respondents may never be aware of all the possible answers

- Takes longer to fill

- Returns are often few

- Information is too unwieldy and unstructured, and hence difficult to analyse, tabulate and interpret

Some investigators combine the approaches and the questionnaires carry both the closed and open form items. In the close ended questions, the last alternative is kept open for the respondents to provide their optimum response.

3. **Pictorial form:** Pictorial questionnaires contain drawings, photographs or other such material rather than written statements and the respondents are to choose answers in terms of the pictorial material. Instructions or directions can be given orally. This form is useful in working with illiterate persons, young children and persons who do not know a specific language. It keeps up the interest of the respondent and decreases subjects' resistance to answer.

5.3.2 Questionnaire Administration Modes

The main modes of questionnaire administration are as follows:

- **Through mail:** Mailed questionnaires are the most widely used and also perhaps the most criticized tool of research. They have been referred to as a 'lazy person's way of gaining information'. The mailed questionnaire has a written and signed request as a covering letter and is accompanied by a self-addressed, written and stamped envelope for the return by post. The method of mailing out the questionnaire is less expensive in terms of time, funds required; it provides freedom to the respondent to work at his/her own convenience and enables coverage of a large population.

- **Personal contact/face-to-face:** Personally administered questionnaires both in individual and group situations are helpful in some cases and have the following advantages over the mailed questionnaire: (i) the investigator can establish a rapport with the respondents; (ii) the purpose of the questionnaire can be explained; (iii) the meaning of the difficult terms and items can be explained to the

respondents; (iv) group administration when the respondents are available at one place is more economical in time and expense; (v) the proportion of non-response is cut down to almost zero; and (vi) the proportion of usable responses becomes larger. However, it is more difficult to obtain respondents in groups and may involve administrative permission which may not be forthcoming.

- **Computerized questionnaire:** It is a mode of questionnaire administration where the questions need to be answered on the computer.

- **Adaptive computerized questionnaire:** It is a mode of questionnaire administration presented on the computer where the next questions are adjusted automatically according to the responses given as the computer is able to gauge the respondent's ability or traits.

5.3.3 Appropriateness of Questionnaire

The qualities and features which make questionnaires an effective instrument of research and help to elicit maximum information are discussed below:

- **Type of information required:** The usefulness and effectiveness of a questionnaire is determined by the kind of information sought. Not every type of questionnaire can be elicited through it. A questionnaire which will consume more than 10–20 minutes is unlikely to get good response. Also, the questions should be explicit and capable of clear-cut replies.

- **Type of respondent reached:** A good deal depends upon the types of respondents covered by the questionnaire. All types of individuals cannot be good respondents. Only literate and socially conscious individuals would give any consideration to a questionnaire. Also, the respondent must be competent to answer the kind of questions contained in a particular questionnaire.

- **Accessibility of respondents:** Questionnaires sent by e-mail can help to survey the opinion of the people living in far-flung places.

- **Precision of the hypothesis:** Appropriateness of the questionnaire also depends upon how realistic is the hypothesis in the mind of the researcher. The researcher must frame questions in such a manner that they elicit responses needed to verify the hypothesis.

5.3.4 Types of Questions

There are many types of questions that can be asked, but the way to get to the correct answer is to know which is the right question. It requires knowledge and expertise to design the correct type of questionnaire.

The following is a list of the different types of questions which can be included in a questionnaire design:

- **Open format questions:** Open format questions are those which give the respondent a chance to communicate their individual opinions. There are no set answers to choose from. Responses from open format questionnaires are insightful and even unexpected. Qualitative questions are an example of open format questions. An ideal questionnaire is one which ends with an open format question giving the respondents the chance to state their opinion or ask for their suggestions.
    Example: State your opinion about the reservation system.

A respondent's answer to an open-ended question is coded into a response scale afterwards. An example of an open-ended question is a question where the person being tested has to complete a sentence (sentence completion item).

- **Closed format questions:** Multiple choice questions are the best example of closed format questions. Closed format questions generate responses that can be statistics or percentages in nature. Preliminary analysis can also be performed with ease. Closed format questions have the added advantage of being able to monitor opinions over a period of time as they can be put to different groups at different intervals.
- **Leading questions:** These types of questions force the audience to give a particular type of answer.
  For example, asking a question with answer options: fair, good, excellent superb.
- **Likert questions:** Likert questions can help you ascertain how strongly your respondent agrees with a particular statement. Likert questions can also help to assess liking and disliking.
  Example: How often do you see beggars at the traffic light?
- **Rating scale questions:** In rating scale questions, the respondent is asked to rate a particular issue on a scale that may range from poor to good. Rating scale questions usually have an even number of choices, so that respondents are not given the choice of a middle option.
  For example, questions with answers options: good, fair, poor or very poor.

Questions to be Avoided during Preparation of a Questionnaire

The following questions should be avoided when preparing a questionnaire:

- **Embarrassing questions:** Embarrassing questions are those that ask respondents about their personal and private life. Embarrassing questions are mostly avoided.

- **Positive/Negative connotation questions:** While defining a question, strong negative or positive overtones must be avoided. Depending on the positive or negative association of our question, we will get different data. Ideal questions should have neutral or subtle overtones.

- **Hypothetical questions:** Hypothetical questions are questions that are based on assumption and hope. An example of a hypothetical question would be 'If you were a director in the Tribal Affairs department, what changes would you bring about?' These types of questions force the respondents to give their ideas on a particular subject. However, these kinds of questions do not give consistent or clear data.

5.3.5 Steps for Preparing and Administering the Questionnaire

The steps involved in preparing and administering the questionnaire are as follows:

- **Planning the questionnaire:** One should get all the help possible in planning and constructing the questionnaire. Other questionnaires should be studied and items should be submitted for criticism to other members of the class or faculty.

- **Modifying questions:** Items can be refined, revised or replaced by better items. If a computer is not readily available for easily modifying questions and rearranging the items, it is advisable to use a separate card or slip for each item. This procedure

also provides flexibility in arranging items in the most appropriate psychological order before the instrument is finalized.

- **Validity and reliability of questionnaire:** Questionnaire designers rarely deal with the degree of validity and reliability of their instrument. There are ways to improve both validity and reliability of questionnaires. Basic to the validity of a questionnaire is asking questions in the least ambiguous way. The meaning of all terms must be clearly defined so that they have the same meaning to all respondents. The panel of experts may rate the instrument in terms of how effectively it samples significant aspects of content validity. The reliability of the questionnaire may be tested by a second administration of the instrument with a small sub-sample, comparing the responses with those of the first. Reliability may also be estimated by comparing the responses of an alternate form with the original.

- **Try out or pilot testing:** The questionnaire should be tried on a few friends and acquaintances. What may seem perfectly clear to the researcher may be confusing to the other person who does not have the frame of reference that the researcher has gained from living with and thinking about an idea over a long period. It is also a good idea to pilot test the instrument with a small group of persons similar to those who will be used in the study. They may reveal defects that can be corrected before the final form is printed.

- **Information level of respondents:** It is important that the questionnaire be sent only to those who possess the desired information and are likely to be sufficiently interested to respond objectively and conscientiously. A preliminary card asking whether the individual would respond is recommended by some research authorities.

- **Getting permission:** If the questionnaire is to be used in a public school, it is essential that approval for the project is secured from the Principal. Students should be informed that participation is voluntary. If the desired information is delicate or intimate in nature, the possibility of providing for anonymous responses should be considered. The anonymous instrument is most likely to produce objective and honest responses.

- **Cover letter:** A courteous, carefully constructed cover letter should be included to explain the purpose of the study. The cover letter should assure the respondent that all information will be held in strict confidence. The letter should promise some sort of inducement to the respondent for compliance with the request. In educational circles, a summary of questionnaire results is considered an appropriate reward, a promise that should be scrupulously honoured after the study has been completed.

- **Follow-up procedures:** Recipients are often slow to return completed questionnaires. To increase the numbers of returns, a vigorous follow-up procedure may be necessary. A courteous postcard reminding the recipient may bring in some additional responds. A further step in follow-up may involve a personal letter or reminder. In extreme cases, it may be appropriate to send the copy of questionnaire with a follow-up letter.

- **Analysing and interpreting questionnaire responder:** Data obtained by the questionnaire is generally achieved through calculation and counting. The total is converted into proportion or percentages. Calculation of contingency coefficient of correlation is often made in order to suggest probability of relation among data. Computation of chi-square statistics in it is also advisable.

Improving the Validity of a Questionnaire

The validity of the information collected through a questionnaire can be improved by using the following techniques:

- The questions should be relevant to the subject or problem.
- The questions should be perfectly clear and unambiguous.
- The questions should be retroactive and not repulsive.
- Check whether the information has been collected from a reasonably good proportion of respondents.
- The information should show a reasonable range of variety.
- The information should be consistent with what is already known or is expected.
- Use another external criterion like consultation of documents or interview with a small group of respondents to cross check the truthfulness of the information given through the questionnaire.

The question sequence should be the following:

- Questions should flow logically from one to the next.
- The researcher must make sure that the answer to a specific question is not prejudiced by earlier questions.
- Questions should flow from the more general to the more specific.
- Questions should follow an order which starts from the least sensitive to the most sensitive.
- Questions should flow from factual and behavioural questions to attitudinal and opinion questions.
- Questions should flow from unaided to aided questions.

  The three stages theory (also known as the sandwich theory) should be applied when sequencing questions. The order to be followed should be first, screening and rapport questions; second, the product specific questions; and third, demographic questions.

Questionnaire Construction Issues

The following problems are faced by a researcher while constructing a questionnaire.

- It is very important to know exactly how you are going to use the information received from the research conducted. If the research or information cannot be implemented or acted upon, then the research would just have been a waste of time, money and effort.
- Clear parameters regarding the research's aims and scope should be drawn before starting the research. This would include the questionnaire's time frame, budget, manpower, intrusion and privacy.
- The target audience selected will depend on how arbitrarily one has chosen the respondents and what the selection criteria are.
- The framework of expected responses should be clearly defined so that the responses received are not random.
- Only relevant questions should be included in the questionnaire as unrelated questions are a burden on the researcher and respondent.

- If you have formed a hypothesis which you want to study then you will know what questions need to be asked.

- The respondents' background and education should not influence the way they answer the questions.

- The type of scale, index, or typology to be used shall be determined.

- The questions asked (closed, multiple-choice, and open) should adhere to the statistical data analysis techniques available and the goals of the study.

- Questions and prepared responses to choose from should not be biased. A biased question or questionnaire influences the responses given.

- The order in which the questions are presented or asked is also important as the earlier questions and their responses may influence the later ones.

- The language should be kept simple to avoid ambiguity. Ambiguous words may cause misunderstanding, possibly invalidating questionnaire results. Double negatives should also be avoided.

- Questions should address only one issue at a time so that the respondent is not confused as to what response is required.

- The list of possible responses should be comprehensive so that respondents should not find themselves without a suitable response. A solution to this would be to add the category of 'other' in the options.

- Categories in the questionnaire should be kept separate. For example, in both the 'married' category and the 'single' category—there may be a need for separate questions on marital status and living situation.

- Writing style should be informal yet to the point and suitable for the target audience.

- Personal questions about age, income, marital status, etc., should be placed at the end of the survey so that even if the respondent is hesitant to give out personal information, they would still have answered the other questions.

- Questions which try to trick the respondent may end in inaccurate responses.

- Presentation which is pleasing to the eye with the use of colours and images can end up distracting the respondent.

- Numbering the questions would be helpful.

- Whoever administers the questionnaire, be it research staff, volunteers or whether self-administered by the respondents, it should have clear, detailed instructions.

Factors Affecting Reliability of Answers

Factors affecting reliability of answers are as follows:

- **Confusing questions:** If the questions are not easily understood or they are capable of being interpreted in more than one way, the answers might be unreliable because the answer may be the result of misinterpretation of the questions not intended by the researcher.

- **Prejudice regarding sample:** The responses received from the sample may not be true representations of the sample.

- **Lack of coverage to illiterates:** This method is inapplicable to illiterates and semi-illiterates as they will be unable to read the questions.

- **Response selectivity:** The respondents of a questionnaire may belong to a selected group. Therefore, the conclusions lack the kind of objectivity and representativeness essential for its validity.

5.3.6 Importance and Limitations of Questionnaire Method

As a matter of fact, this method can be applied in a very narrow field. It can be used only if the respondents are educated and willing to cooperate. However, it is still widely used, owing to the following merits:

- **Economical:** The questionnaire requires paper, printing and postage only. There is no need to visit the respondents personally or continue the study over a long period.

- **Time saving:** Besides saving money, the questionnaire also saves time. Data can be collected from a large number of people within a small time frame.

- **Most reliable in special cases:** It is a perfect technique of research in some cases.

- **Research in wide area:** Mailed questionnaire comes very handy if the sample comprises people living at great distances.

- **Suitable in specific type of responses:** The information about certain problems can be best obtained through the questionnaire method.

Limitations of the Questionnaire Method

Like all other methods, the questionnaire is also limited in value and application. This means that it cannot be used in every situation and that its conclusions are not always reliable. Key limitations of the method are as follows:

- **Limited response:** As noted earlier, this method cannot be used with illiterate or semi-illiterate groups. The number of persons who cooperate and respond to the questionnaire is very small.

- **Lack of personal contact:** There is very little scope of personal contact in this method. In the absence of personal contact, very little can be done to persuade the respondents to fill up the questionnaire.

- **Useless in-depth problems:** If a problem requires deep and long study, it is obvious that it cannot be studied by the questionnaire method.

- **Possibility of wrong answers:** A respondent may not really understand a question or may give the answer in a casual manner. In both cases, there is a strong likelihood of misleading information being given.

- **Illegibility:** Some persons write so badly that it is difficult to read their handwriting.

- **Incomplete response:** There are people who give answers which are so brief that the full meaning is incomprehensible.

5.4 SCHEDULES

A schedule is a questionnaire containing a set of questions that are required to be answered to collect data about a particular item. A schedule is generally used in a face-to-face situation. The following are the objectives for which a schedule is created:

Check Your Progress

4. What is a questionnaire?

5. What are the limitations of open form questionnaire?

6. What is the three stages theory when sequencing questions?

7. List two limitations of the questionnaire method.

- It is created for a definite item of inquiry. A schedule sets the boundaries for the subject under study.

- It acts as an aid to memorize the information being collected by the interviewer from various respondents. It helps to avoid being confused while analysing and tabulating the data.

- It helps in tabulating and analysing the data in a systematic and standardized manner.

Characteristics of a Good Schedule

The essential characteristics of a good schedule are as follows:

- The information or questions included in the schedule should be accurate and should enable the respondent to understand properly the context in which the questions are being asked.

- The schedule should be pre-arranged and structured in such a manner that the information gathered or collected should be accurate and tenable. For this, the following points must be considered:

  o The size of the schedule should be accurate.

  o The questions in the schedule should be understandable and definite.

  o The questions should not contain any biased evaluation.

  o All the questions of the schedule should be properly interlinked.

  o The information gathered should be organized in a table so that it can be easily used for statistical analysis.

Suitability of the Schedule Method

The schedule method is mostly applied in the following situations:

- When the field of investigation is wide and dispersed

- When the researcher requires quick results at lesser cost

- When the respondents are well-trained and educated

5.4.1 Types of Schedules

There are five types of schedules, which are as follows:

- **Observation schedule**: This schedule is used to observe all the activities and record the responses of the respondents under some predefined conditions. The main idea behind examining the activities is to verify the required information.

- **Rating schedule**: It is used to measure and rate the thoughts, preferences, self-consciousness, perceptions and other similar characteristics of the respondent.

- **Document schedule**: It is used for collecting important data and preparing a source list. This schedule is mostly used to attain data from autobiographies, diaries or government records regarding written facts and case histories.

- **Institution survey schedule**: It is used for studying the problems of institutions.

- **Interview schedule**: It is used to ask the interviewee questions and record the responses in the space provided in the questionnaire itself.

Merits and Limitations of the Schedule Method

The merits of the schedule method are as follows:

- In this method, the researcher is always there to help the respondents. So, the response rate is high as compared to other methods of data collection.

- The presence of the researcher not only removes doubts present in the mind of the respondent, but also avoids false replies from the respondent due to fear of cross-checking.

- In this method, there is personal contact between the researcher and the respondent. Thus, the data can be collected easily and can also be relied upon.

- This method helps to better understand the personality, living conditions and values of the respondents.

- It is easy for the researcher to detect and rectify defects in the schedule during sampling.

Limitations of the Schedule Method

The limitations of this method are as follows:

- It is a costly and time-consuming method.

- It requires well-trained and experienced field workers for conducting interviews of the respondents.

- Sometimes, the respondent may not be able to speak out due to the physical presence of the researcher.

- If the field of research is dispersed, it becomes difficult to organize the various activities of the research.

5.4.2 Organization of the Schedule

The schedule is prepared by performing the following steps:

- **Selection of respondents**: Usually, the sampling method is used for the selection of respondents. The sample should be representative of the respondents and should contain all the relevant information about the respondents.

- **Selection and training of field workers**: Since the field workers interview the respondents and collect the required data, this should be done carefully and proper training should be provided to them.

- **Conducting interviews**: For a successful interview and correct results, the following points must be kept in mind:
  - **Follow correct approach**: The field worker should go to the respondent with the correct approach so that the respondent can clearly understand the purpose of the interview.
  - **Generating accurate responses**: For proper and accurate response from the respondents, the respondents should not be misunderstood in their perspective and context.

5.4.3 Difference between Questionnaire and Schedule

When you work with questionnaires and schedules, you will observe that there are several similarities between the two. However, there are prominent differences also, which are as follows:

- A questionnaire is mostly sent by the interviewer to the interviewee by mail and is filled by the interviewee, whereas a schedule is filled by the interviewer at the time of interview.

- Data collection through a questionnaire is cheaper as compared to a schedule, as money is spent only in preparing the schedules and mailing them. In the schedule method, extra money is spent on appointing interviewers and imparting training to them.

- In the case of a questionnaire, response is generally low because many people do not respond. On the other hand, response is high in the case of schedules since the interviewer fills them at the time of the interview.

- The identity of the respondent is not always clear in the case of a questionnaire, whereas in the case of schedules, the identity of the interviewee or respondent is known.

- The questionnaire method is time consuming as the respondent may not return the questionnaire in time. There is no such problem with the schedule method because the schedule is filled at the time of the interview.

- The questionnaire method does not allow personal contact with the respondent but the schedule method does.

- The questionnaire method is useful only if the respondent is literate, while in the case of a schedule, it is not necessary for the interviewee to be literate.

- The risk of incomplete and incorrect information is more in a questionnaire, while in a schedule, the information collected is complete and more accurate.

5.5 INTERVIEW

One of the main methods of data collection is conducting interviews. It takes place as a two-way conversation between the researcher and the respondent, whereby information is gathered by asking topic related questions.

We learn not only from the respondents' responses but also his/her gestures, facial expressions and pauses. Interviewing can be conducted either face-to-face or over the telephone by skilled personnel by using a structured schedule or an unstructured guide.

According to Rummel J. Francis: 'The interview method of collecting data requires the actual physical proximity of two or more persons, and generally requires that all the normal channels of communication be open to their use. It is necessary to see one another, to hear each other's voices, to understand one another's language, and to use all that is psychologically inherent in physical proximity. It usually entails a non-reciprocal relation between the individuals concerned. One party desires to get information from another—one party interviews the other—for a particular purpose.'

Check Your Progress

8. What is a schedule?
9. When is the interview schedule used?

Theodore L. Torgerson has stated that the interview method of study extends certain aspects of the observational technique.

Thus, the interview method permits the gathering of development data to supplement the cross-sectional data obtained from observations. The interviewer can probe into casual factors, determine attitudes, discover when the problem started, enlist the interviewee in an analysis of his own problem and secure his support of the therapy to be applied.

5.5.1 Types of Interviews

The different types of interviews are as follows:

- **Group interview:** A proper setting for group interviews requires a group of not more than 10 to 12 persons with some social, intellectual, and educational homogeneity, which ensures effective participation by all. For a full spontaneous participation of all, it is better to arrange a circular seating arrangement.

- **Diagnostic interview:** Its purpose is to locate the possible causes of an individual's problems, getting information about his past history, family relations and personal adjustment problem.

- **Clinical interview:** Such an interview follows after the diagnostic interview. It is a means of introducing the patient to therapy.

- **Research interview**: Research interview is aimed at getting information required by the investigator to test his/her hypothesis or solve his/her problems of historical, experimental, survey or clinical type.

- **Single interview or panel interviews:** For the purpose of research, a single interviewer is usually present. In case of selection and treatment purposes, panel interviews are held.

- **Directed interview:** It is structured, includes questions of the closed type and is conducted in a prepared manner.

- **Non-directive interview:** It includes questions of the open-ended form and allows much freedom to the interviewee to talk freely about the problem under-study.

- **Focused interview:** It aims at finding out the responses of individuals to exact events or experiences rather than on general lines of enquiry.

- **Depth interview:** It is an intensive and searching kind of interview. It emphasizes certain psychological and social factors relating to attitudes, emotions or convictions.

It may be observed that on occasions several types are used to obtain the needed information.

Other classifications of interviews are as follows:

- Intake interview, as the initial stage in clinic and guidance centres

- Brief talk contacts as in schools and recreation centres

- Single hour interview

- Clinical psychological interview, stressing psychotherapeutic counselling and utilizing case history data and active participation by the counsellor in the re-education of the client

- Psychiatric interviews, similar to psychological counselling, but varying with the personality and philosophical orientation of the individual worker and with the setting in which used

- Psychoanalytic interviews

- Interview form of test

- Group interviews for selecting applicants for special course

- Research interview

5.5.2 Important Elements of Research Interview

The important elements of research interview are as follows:

1. Preparation for Research Interview

- Decide the category and number of persons that you would like to interview.

- Have a clear conception of the purpose and the information required.

- Prepare a clear outline, a schedule or a checklist of the best sequence of questions that will systematically bring out the desired information.

- Decide the type of interview that you are going to use, i.e., structured or non-structured interview.

- Have a well thought-out plan for recording responses.

- Fix up the time well in advance.

- Procure the tools to be used in recording responses.

2. Executing an Interview

- Be friendly and courteous and put the respondent at ease so that he talks freely.

- Listen patiently to all opinions and never show surprise or disapproval to a respondent's answer.

- Assume an interested manner towards the respondent's opinion, and as far as possible do not divulge your own.

- Keep the direction of the interview in your own hands and avoid irrelevant conversation and try to keep the respondent on track.

- Repeat your questions slowly and with proper emphasis in case respondent shows signs of failing to understand a particular question.

3. Obtaining the Response

Perhaps the most difficult part of the job of an interviewer is to obtain a specific, complete response. People can often be evasive and answer 'do not know' if they do not want to make an effort of thinking. They can also misunderstand the question and answer incorrectly in which case the interviewer would have to probe more deeply.

An interviewer should be skilled in the technique as only then can the interviewer gauge whether the answers are incomplete or non-specific. Each interviewer must fully understand the motive behind asking particular questions and whether the answer is giving the information required. The interviewer should form the habit of asking himself/herself, 'Does that completely answer the question that I just asked?'

Throughout the interview, the interviewer must be extremely careful as to not suggest a possible reply. The interviewer should always content himself with mere repetition (if the question is not understood to answer).

4. Reporting the Response

There are two chief means of recording opinion during an interview. If the question is preceded, the interviewer only needs check a box or circle or code, or otherwise indicate which code comes closest to the respondent's opinion. If the question is not preceded, the interviewer is expected to record the response verbatim.

The following points may be kept in view in this respect:

- Quote the respondents directly, just as if the interviewers were newspaper reporters taking down the statement of an important official without paraphrasing the reply, summarizing it in the interviewer's own words, 'polishing up' any slang or correcting bad grammar that distorts the respondent's meaning and emphasis.

- Ask the respondent to wait until the interviewer gets down 'that last thought'.

- Do not write as soon as you have asked the question and do not write while the respondent talks. Wait until the response is completed.

- Use common abbreviations.

- Do not record and evaluate the responses simultaneously.

5. Closing the Interview

It should be accompanied by an expression of thanks giving recognition to the respondent's generosity in sparing time and effort.

6. Use of Tape Recorder in Interview

- It reduces the tendency of the interviewer to make an unconscious selection of data favouring the interviewer's biases.

- The tape recorded data can be played more than once, and thus it permits a thorough study of the data.

- Tape recorder speeds up the interview process.

- Tape recorder permits the recording of some gestures.

- The tape recorder permits the interviewer to devote full attention to the respondent.

- No verbal productions are lost in a tape recorded interview.

- Other things being equal, the interviewer who uses a tape recorder is able to obtain more interviews during a given time period than an interviewer who takes notes or attempts to reconstruct the interview from memory after the interview has been completed.

5.5.3 Indifferent Attitude of the Respondent and the Role of the Research Worker

It is observed that the research worker is likely to encounter several problems arising out of the apathy of the respondents. In such a situation, the following points may be kept in view:

- When the respondent is really busy and has no time, the field worker may request for a more convenient time.

- When the respondent simply wants to avoid the interview and is not inclined to be bothered about it, the field worker should try to explain to the respondent the importance of the study, and how the respondent's own response is of material value in the case.

- When the respondent is afraid to give the interview as it affects his boss or the party to which he belongs or any other cause which is likely to harm his interest, the field worker must assure the respondent that absolute secrecy would be maintained by the researcher and the organization.

- When the respondent does not hold a high opinion about the outcome of such interviews in general, or has a poor opinion about the research organization or institution conducting it, it is the duty of the research worker at such times to explain to the respondent the importance of the problem, and convince the respondent regarding the status of the research body.

- When the respondent is suspicious and he thinks that the enquiry is either from the income tax department or some other secret agency, at such times he may generally ask such questions. Who are you? Who told you our name? Have you interviewed the neighbour? etc. The research worker should try to eliminate the respondent suspicion. A letter of authority, the letter head or the seal of the research body would prove to be useful on such occasions.

- When the respondent is unsocial or otherwise confined to his own family (such a tendency is mostly found in the case of newly married couples), the research worker at such times will try to create the respondent's interest in the subject of investigation.

- When the respondent is too haughty and thinks it below his dignity to grant an interview to petty research workers, the investigator should get a letter of introduction from an influential person.

5.5.4 Advantages and Disadvantages of Interview Method

The advantages of the interview method over other techniques are as follows:

- A well-trained interviewer can obtain more data and greater clarity by altering the interview situation. This cannot be done in a questionnaire.

- An interview permits the research worker to follow-up leads as contrasted with the questionnaire.

- Questionnaires are often shallow and they fail to dig deeply enough to provide a true picture of opinions and feelings. The interview situation usually permits much greater depth.

- It is possible for a skilled interviewer to obtain significant information through motivating the subject and maintaining rapport, other methods do not permit such a situation.

- The respondents when interviewed may reveal information of a confidential nature which they would not like to record in a questionnaire.

- Interview technique can be used in the case of children and illiterate persons who cannot express themselves in writing. This is not possible in a questionnaire.

- The percentage of response is much higher than in case of a mailed questionnaire.

- The field worker is personally present to remove any doubt or suspicion regarding the nature of enquiry or meaning of any question or term used. The answers are, therefore, not biased because of any misunderstanding.

- The field worker may create a friendly atmosphere for proper response. The field worker may start a discussion, and develop the interest of the respondent before showing the schedule. A right atmosphere is very conducive for getting correct replies.

- The interviewee may disclose personal and confidential information which the interviewee would not ordinarily place in writing on paper. The interviewee may need the stimulation of personal contacts in order to be drawn out.

- The interview enables the investigator to follow-up leads and to take advantage of small clues, in dealing with complex topics and questions.

- The interview permits an exchange of ideas and information. It permits 'give and take'.

- It is useful in the case of some categories of persons. The interview enables the interviewee to deal with young children, illiterates and those with limited intelligence or who's state of mind is not quite normal.

- Interviews are also used for pupil counselling, for selection of candidates for instructional purposes, for employment, for psychiatric work, etc.

- The respondent does not feel tired or bored. Supplementary questions may be put to enliven the whole discussion.

- The difficulties of bad handwriting of the respondent, use of pencil, etc., are also avoided as every schedule is filled in by the interviewer.

- A probe into life pattern is possible. The personal contact with the respondent enables the field worker to probe more deeply into the character, living conditions and general life pattern of the respondent. These factors have a great bearing in understanding the background of any reply.

- The information gathered through interviews has been found to be fairly reliable.

- It is possible for the interviewer to probe into attitudes, discover the origin of the problem, etc.

- Interview technique is very close to the teacher. It is generally accepted that no research technique is as close to the teacher's work as the interview.

- Sometimes interviews can be held at suitable intervals to trace the development of behaviour and attitudes.

- Interviews can be used for student counselling, occupational adjustment, selection of candidates for educational courses, etc.

- Interviews can be used for all kinds of research methods—normative, historical, experimental, case studies and clinical studies.

- Interview techniques provide scope for cross questioning.

- This technique allows the interviewer to remain in command of the situation throughout the investigation.

- Through the respondent's incidental comments, facial expression, bodily movements, gestures, etc., an interviewer can acquire information that could not be obtained easily by other means.

- Cross questioning by the interviewer can enable him/her to judge the sincerity, frankness and insight of the interviewee.

Disadvantages of Interview Method

The method of interview, in spite of its numerous advantages has the following limitations:

- **Very costly:** It is a very costly affair. The cost per case is much higher in this method than in case of mailed questionnaires. Generally speaking, the cost per questionnaire is much less than the cost per interview. A large number of field workers may have to be engaged and trained in the work of collection of data. All this entails a lot of expenditure and a research worker with limited financial means finds it very difficult to adopt this method.

- **Biased information:** The presence of the field worker while encouraging the respondent to reply, may also introduce a source of bias in the interview. At times the opinion of the respondent is influenced by the field worker and his replies may not be based on what he thinks to be correct but what he thinks the investigator wants.

- **Time consuming:** It is a time consuming technique as there is no guarantee how much time each interview can take, since the questions have to be explained, interviewees have to assured and the information extracted.

- **Expertness required:** It requires a high level of expertise to extract information from the interviewee who may be hesitant to part with this knowledge.

Among the important qualities to be possessed by an interviewer are objectivity, insight and sensitivity.

5.6 PROJECTIVE TECHNIQUES

The idea of projecting oneself or one's feelings on ambiguous objects is the basic assumption in projective techniques. The 19th century saw the origin of these techniques in clinical and developmental psychology. However, it was after second World War that these techniques were adopted for use in advertising agencies and market research firms. Ernest Dichter (1960) was one of the pioneers who used these techniques in consumer and motivational research. Consumer surveys and research were considered incomplete if they did not make use of projective techniques (Henry, 1956; Rogers and Beal, 1958; Newman, 1957). However, with the advent of technology and computer-aided analysis, these subjective methods were generally forgotten.

It was only in the 1990s that work done on semiotics, in-depth interviews and renewed interest in human emotions and needs, especially the latent needs and brand personalities led to resurgence of these methods (Belk et al., 1997 and Zaltman, 1997).

Unlike the other approaches discussed in the unit, these methods involve indirect questioning. Instead of asking direct questions, the method involves a relatively ambiguous stimuli and indirect questions related to imaginary situations or people. The purpose of the research is to present a situation to the respondents to project their underlying needs, emotions, beliefs and attitudes. The ambiguity of the situation is non-threatening and thus a person has no hesitation in revealing his/her true inner motivations and emotions. The more the degree of ambiguity, the more is the range of responses one gets from the respondents. In the theoretical sense, projective techniques unearth beliefs, attitudes and feelings that might underlie certain behaviour or interaction situations. Thus, the respondents' attitudes are uncovered by analysing their responses to the scenarios that

are deliberately constructed to stimulate responses from the right side of the brain, which is stated to be the affective side. The second premise of projective techniques is to uncover the different levels of consciousness (Freud, 1911). Generally, the structured methods look at primary motivations; however, it is the underlying latent needs which might drive the individual to behave in a certain manner. The third is to reveal data that is inhibited by socially-desirable and correct responses. Sometimes individuals hesitate to express their prejudices or feelings towards other individuals, groups or objects. Indirect and ambiguous stimuli might reveal startling results in such cases. In psychology, there are a wide variety of techniques available. These can be categorized on the basis of the conduction process. Some of these techniques are briefly discussed below.

1. **Association techniques:** These are the most frequently used methods in management research. They essentially involve presenting a stimulus to the respondent and he needs to respond with the first thing that comes to his mind. The method is essentially borrowed from clinical psychology, the most well-known being the Rorschach Inkblot test. The set of inkblots are ambiguous in nature, however, these are standardized blots symmetrical in nature. The first few are in shades of black and white and the others are coloured. Each of these is presented in a sequence to the consumer. The responses, time taken, the direction in which the blot is turned, are noted. There are norms and scores available for evaluating the personality of the individual. They require a considerable amount of training in conduction and interpretation and, thus, are not commonly used. A technique based on the same principle is called the word association test. This found its earliest uses in 1936 by Houghton for advertising evaluations. The technique involves presenting a group of words and the respondent needs to respond instantly with the first thing that comes to his mind. The critical words are disguised and come after a few neutral or mundane words. The idea is that the element of surprise will reveal associations that lie in the subconscious or the unconscious mind. The words which are selected to address the objectives of the study are called test words and the others are called fillers.

   For example, to attest the extent of eco-friendly attitude of a community, one could have a number of words like 'environment', 'plastic', 'water', 'earth', 'tigers', 'clean', etc. These would be embedded in the fillers to see the extent to which the consumer is aware. The person's exact response is either noted or recorded; in case one is doing this manually, it is critical to note the reaction time of the person, as hesitating would mean that there was a latent response which the person was not comfortable about revealing. In this case, the response needs to be discarded or evaluated through other responses. Another variation of the test used in individual and brand personality is to ask the person to think of an animal/object that one associates with a brand or a person.

   For example, the word 'wall' is associated with a famous Indian cricketer.

   The obtained answers are measured in terms of:

   - Similarity of responses given to a test word by a number of respondents

   - Unique responses

   - Time taken for a response

   - Non-response

In case a person does not respond at all, it is assumed that there is emotional block hampering. A person's attitudes and feelings related to the topic can be measured by this technique.

Illustration: Talking to elders: A popular pharmaceutical firm produces a range of expensive products meant for old age consumers. The company plans to use television advertising to create awareness about the products. Word association was used to study old people's attitudes towards medication and supportive therapy. Six men and six women were selected to administer the test; they were matched on income, class, age, education and current status of living with their married sons/daughters. The test words used and the responses obtained are in Table 5.1.

Table 5.1 Test Word Used and Responses Obtained

Test words	Responses		
Health	Care (3)	Bad (2)	Good (1)
Life	Difficult (2)	Relaxed (3)	Good (1)
Medicines	Necessity (4)	Prevention (2)	Avoid (1)
Walking stick	Support (3)	Avoid (2)	Carved ivory (1)
Adult diapers	Embarrassment (4)	Necessity (2)	
Treatment	In time (2)	Expensive (4)	
Bones	Weak (3)	Brittle (3)	
Death	The end (1)	Inevitable (5)	

The major responses are highlighted and reveal that the seniors are not afraid of dying, are realistic about failing health and supportive medicines or walking stick. However, they have clearly stated that they do not want to be embarrassed. Thus, talking about their health problems on a public platform and offering solutions would not be welcome. They are conscious and positive about medicines being essential, however, their dignity must be kept intact.

This research was taken as a reflection of the attitude of the elderly at large and the company does not use television advertising at all, rather it relies on doctors and chemists to push the product.

An extension of the association technique is the completion technique.

2. **Completion techniques:** These techniques involve presenting an incomplete object to the respondent, which can be completed by the respondent in any way.

Old age is ....

Sentence completion is the most popular of all projective techniques and is inevitably used in almost all measuring instruments as an open-ended question. However, the incomplete sentence of a typical projective test needs to be more ambiguous than a typical open-ended question. Generally, they are given a single word or phrase and asked to fill it in, for example:

Working at IBM is. Or ....

McDonald is. ....

Another extension of the technique is story completion. Here, the individual is given an incomplete story or idea. One provides a backdrop and a background for a possible topic. However, the possible end is left open-ended. The subject

is supposed to complete the story and provide a conclusion. The theoretical assumption is that the completion of the story/sentence reflects the underlying attitude and personality traits of the person.

3. **Construction techniques:** These techniques might appear similar to completion technique, however here, the focus is on the completed object, which could be a story, a picture, a dialogue or a description. Here again, the level of ambiguity and scope for letting loose the respondents' imagination is vast.

Clinical psychology has a whole range of construction techniques, but here we will refer only to the ones which are actively used in business research. These are:

- **Story construction tests:** The most often used test is the Thematic Apperception Test (TAT) developed by Henry (1956). There are a total of 20 pictures, most of them having the profile of a man, woman or child either clearly visible or diffused. The pictures are given to the respondent and he is asked about what is happening here? What happened or led to this? What do you think is going to happen now? The assumption is, that in most instances the person puts himself into the shoes of the protagonist and actually indicates how he would respond in the given situation. The story gives an indication of the person's personality and need structure. For example, an individual may be characterized as extroverted, or a pessimistic or high on creativity or high on dogmatism, and so on. The TAT is used extensively, in parts (a few selected pictures) or in totality in a number of organizations, including the armed forces. The usage is majorly done for selection and recruitment process.

- **Cartoon tests:** The tests make use of animated characters in a particular situation (Masling, 1952). They are considered ambiguous as the figures bear no resemblance to a living being and thus are considered non-threatening. The cartoon usually has a picture that has two or more characters talking to each other; usually the statement/question by one character is denoted and one needs to fill in the response made by the other character. The picture has a direct relation with the topic under study and is assumed to reveal the respondent's attitude, feelings or intended behaviour. They are one of the easiest to administer, analyse and score.

4. **Choice or ordering techniques:** These techniques involve presenting the respondents with an assortment of stimuli—in the form of pictures or statements—related to the study topic. The subject is supposed to sort them into categories, based on the study instructions given. For example, in a study on measuring desired supervisor–subordinate relations, a set of Tom and Jerry cartoon pictures were used, some in which Tom is overpowering Jerry, some neutral pictures where they are carrying out their respective tasks and others where Jerry, the mouse outwits Tom. The respondent needs to sort them into good, neutral and bad picture piles.

These sets are not similar to cartoon tests as they do not require completion or closure. These require sorting, in order to measure any stereotyped or typical behaviour of the respondent. The pictures that have been given to the person carry an expert score (that is they have been categorized on a rating scale to

reveal different degrees of the attitude). The higher the selection of pictures with extreme scores, the more rigid is the respondent's attitude and in case modification or enhancement is required, the task would be more difficult. The test is used to measure attitudes and the strength of the existing attitude.

5. **Expressive techniques:** The focus on the other five techniques was on the end result or the output. However, in expressive techniques, the method or means or expressions used in attempting the exercise are significant. The subject needs to express not his/her own feelings and opinions but those of the protagonist(s) in a given verbal or visual situation. Again the presumption is that people are uncomfortable giving personal opinion on a sensitive issue, but, do not mind or are less inhibitive when it is in the third person. There are many examples: Clay modelling—here the emphasis is on the manner in which the person uses or works with clay and not on the end result.

Psychodrama (Dichter, 1964)—here the person needs to take on the roles of living or inanimate object, like a brand(s) and carry out a dialogue.

Object personification (Vicary, 1951)—here the person personifies an inanimate object/brand/organization and assigns it human traits.

Role playing is another technique that is used in business research. The respondents are asked to play the role or assume the behaviour of someone else. The details about the setting are given to the subject(s) and they are asked to take on different roles and enact the situation.

The third-person technique is again considered harmless as here, the respondent is presented with a verbal or visual situation and needs to express what might be the person's beliefs and attitudes. The person may be a friend, neighbour, colleague, or a 'typical' person. Asking the individual to respond in the third person reduces the social pressure, especially when the discussion or study is about a sensitive issue. For example, no respondent even when assured of anonymity, would own up to being open to an extra-marital affair; however, if asked whether a colleague/friend/person in his/her age group might show an inclination for the same, the answers might be starkly different.

5.6.1 Evaluating Projective Techniques

As can be seen from the description of the techniques available to the researcher, the projective techniques are unsurpassed in revealing latent yet significant responses. These would not surface through a more structured or standardized techniques like focus group discussions or interviews. The ambiguity and the third-person setting give the respondent a sufficient camouflage and confidence to feel comfortable about revealing attitudes, interests and beliefs about sensitive issues. There might also be instances where the respondent is unaware of his underlying motivations, beliefs and attitudes that are operating at a subconscious level. Projective techniques are helpful in unearthing these with considerable ease and expertise.

However, this richness of data also has its disadvantages. The conduction and analysis of the technique requires specialists and trained professionals. This is also the reason why the tests are expensive and time consuming in usage. Most of the techniques require varying degrees of ambiguity and the higher the ambiguity, the richer is the response. But, at the same time, it makes the analysis and interpretation difficult and

subjective. Role playing and psychodrama require interaction and participation by the subject, thus the person who volunteers to participate in the study, might be unusual in some way. Therefore, generalizing the results of the analysis might be subject to error.

5.7 CASE STUDY AND CONTENT ANALYSIS

Case studies are discussions of individual cases under topics of discussion which help researchers to corroborate known facts proved previously through research. Social scientists, in particular, used the case study method to conduct research for many years. A variety of disciplines used this method of research to corroborate their findings in real life situations. Researcher Robert K. Yin defines the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used (Yin, 1984, p. 23).

However, critics feel that the case study method is not reliable enough for establishing a rule or principle as it portrays only a minuscule population which forms not even a part of the entire population. Some feel that this method is only a reliable exploratory tool. Literature supports reports of carefully planned and crafted studies of the case study method. Robert E. Stake, Helen Simmons, and Robert Yin are renowned researchers who have written about the utility of case studies in social sciences. They have prescribed six steps that should be used when utilizing the case study method. These are:

- Determine and define the research questions
- Select the cases and determine data gathering and analysis techniques
- Prepare to collect the data
- Collect data in the field
- Evaluate and analyse the data
- Prepare the report

1. Determine and define the research questions

Before a case study research is undertaken, cementing a research focus is important so that the researcher can refer to it during the course of study. The research object is often a person, an organizational policy, a group of people, etc. A number of data gathering methods are used by the researcher who studies every case study in depth. The researcher reads the available literature to understand where the topic stands in terms of prior research and undertakes a thorough planning before embarking on the actual case study. Literature and previous studies help him to decide where to look for evidence to corroborate his findings on the concerned topic. These help in designing the blueprint for the current study.

2. Select the cases and determine data gathering and analysis techniques

While designing the study, researchers finalize the approaches, methods of data extraction and data gathering for real-life cases that they need to study. While using multiple cases, each case is treated as a single case. The conclusions of these cases can then be utilized for underlining various facets of their study. The researchers need to discriminate positively

for the case study that they want to utilize for corroborating their findings. Researchers should decide whether they want to study cases that are conventional or extraordinary while conducting the study. In case they are hesitant, they may go back to the purpose of the study that they had enumerated before beginning the research. The decision to choose a single or multiple case studies is an important one, while a single case study may be examined for analysing more than one inherent principle. These types of case studies involve two different levels of analysis which increases the complexity of data collected. Multiple sources and techniques in the data collecting process is a key strength of the case study method. Researchers need to determine what data they would wish to gather by examining a case and how to analyse the data collection. The tools they may use are interviews, surveys, documentation review, observation and collection of physical artefacts. During the design phase of the research, researchers should make sure that the study ensures construct validity, external validity, internal validity and reliability. Researchers need to use the correct measures for ensuring construct validity. Internal validity is ensured when the conditions may be used over and over again to prove validity of the case. External validity is ensured when the findings may be generalized beyond the case or cases. A case study is said to be more externally valid when it can withstand more people, places and procedures. Techniques known as within-case examination and cross-case examination and literature review help ensure the validity of the case.

3. Prepare to collect the data

Researchers using the case study method generally gather a large amount of data from a number of sources. Organizing this data in a systematic manner is a challenge in itself. Researchers should plan ahead to prevent getting overwhelmed by this data. They might even lose sight of the original purpose of gathering the data. Researchers sort, categorize, store and retrieve data for analysis with the help of databases. Extraordinary cases help researchers by providing an efficient training programme, establishing proper protocols and conducting a pilot study before entering fieldwork. The training programme covers the concept to be studied, terminology, processes, methods, etc. The researchers also learn the application of techniques used in the study. In order to gather data from the interviewed population, researchers have to be skilled enough to retain or record the interviews without the gadget coming in the interviewee's way. Researchers should know how to steer conversation towards the questions they intend to ask next. They should be trained in analysing body language and interpret answers not expected by them. Researchers need to read between the lines and in case the topic is sensitive, understand a respondent's hesitation and silence. Researchers should not feel threatened by missed appointments and lack of space for holding the interview or unexpected turns of events during the interview; for example, a respondent may break down while answering a sensitive question. Researchers should be humane, understanding and flexible in approach. They should revisit the research design that they had created before starting the case studies and make changes as and when required.

4. Collect data in the field

Researchers should be trained to collect and store multiple sources of evidence in various formats while going about studying the case. Though case study research is flexible, any change that comes up needs to be documented carefully. The multiple storing of data is required so that converging lines of enquiry and patterns may be discovered. Field notes

may be used for recording intuitions, hunches, feelings, and also for documenting the work in progress. Illustrations, anecdotes and special records may be written in the field notes so that the researcher may refer to it when making case study reports. The data and the field notes should be kept separately for analysis. The researcher needs to document, classify and cross-refer all evidence so that these could be efficiently recalled for examination and sorting as and when required.

5. Evaluate and analyse the data

The raw data gathered by the researchers need to be interpreted at different levels to find linkages between the objectives of the research and the outcome of studying the case. Researchers must remain open to new insights and opportunities throughout the evaluation and analysis process. They can triangulate data with the help of different techniques and collection methods inherent to the case study method. Researchers will be provided with new insights and conflicting data by case studies which are extraordinary. They would need to categorize, tabulate and combine data to address the purpose of the study. In order to cross-check data collected, short, repeated interviews need to be conducted. Placing information into arrays, creating matrices of categories, making flow charts or other displays, etc., may be used by the researcher as specific techniques. The quantitative data collected may be used to corroborate the qualitative data collected during interviews. Many research organizations may also use multiple researchers to verify the data collected. When these multiple observations converge, researchers may become more confident of their findings. Conflicting observations need in-depth study of the findings. The cross-case search technique requires that researchers look at data from different angles and do not reach a premature conclusion. Across all cases investigated, the cross-case search divides data by type. When a pattern from one data is vouched for by another data, the finding is stronger. When these evidences do not form a data, a further probe is essential.

6. Prepare the report

An exemplary case study report transforms the manner in which a complex issue is presented. Case study reports are often published so that readers may apply the experience in their real-life situations. Case studies mostly display evidences to gain the confidence of the readers. Researchers also underline the boundaries of the case and draw the attention of the readers to conflicting propositions. Many researchers present case study reports in the form of a chronological account. Some may treat a case as a fresh chapter. Once a report is completed, the researcher should always edit and examine it for loopholes. Representative audience group is used for comments and criticisms and the valid criticisms are incorporated in the next draft. Since case studies involve multiple sources of data, or may include more than one case within a study, they often become complex. The case study method is generally used by researchers from various disciplines to build upon a theory, to produce a new theory, to challenge or dispute a theory, to explore new horizons, to apply solutions to situations, to describe a phenomenon, etc. There are a number of advantages of the case-study method. These are: applicability to real life situations, to contemporary social situations and easy accessibility to its published reports. Case studies help common man understand a complex theory through easy, real-life situations that are used to exemplify the principle being discussed.

Some examples of case studies are given below.

Case Study 5.1

Danish International (C)

Shameem was returning back after an exhaustive session with P&Y consultants. The lady consultant had reviewed the information that he had provided about the working atmosphere at Danish.

The consultant had also conducted a couple of visits to the office and had submitted her report. She had pointed out clearly that the indifference she had observed was a matter of serious concern. No benchmarked data would help as the problem was peculiar to the unit. She had advised that the attitude and emotions of the members would have to be analysed. She had told Shameen that they had a couple of standardized tests that she could administer and prepare an action plan.

Shameem was not convinced as he knew that the issue needed to be handled at a different level. Then he remembered the lady he had met from Transcend, the research beyond group, who had made a presentation yesterday about seeking the latent to work on the manifest. He recalled the book that he had read by Sigmund Freud and how it had made a lot of sense about why people reacted in a certain way. Yes, there was merit in the surreal. But this was business, should he go for the subjective?

He reached office, read the P& Y report, thought about what he believed and picked up his phone and made the call ………….

1. Who do you think he called? Why?

2. Are there any alternative technique(s) he could use? Explain by providing a template for collecting the information.

Case Study 5.2

What's in a Car?

Shridhar from Bengaluru, had developed an electric car—VERVE (It is a fully automatic, no clutch, no gears), two-door hatchback, easily seating two adults and two children with a small turning radius of just 3.5 metres. It runs on batteries and as compared to other electric vehicles, has an onboard charger to facilitate easy charging which can be carried out by plugging into any 15 amp socket at home or work. A full battery charge takes less than seven hours and gives a range of 80 km. In a quick-charge mode (two-and-a-half hours) 80 per cent charge is attained which is good enough for 65 km. A full charge consumes just about 9 units of electricity. Somehow the product did not take off the way he expected. He is contemplating about repositioning the car. As he stood looking at the prototype, he knew that there were a couple of questions to which he must find answers before he undertook the repositioning exercise. Who should be the target segment—old people, young students just going to college, housewives, or …….? What should be the positioning stance?What kind of image would these customers relate to? Was a new name or punch-line required? How should the promotions be undertaken? Hyundai had done it with Shah Rukh Khan, should he also consider a celebrity? If yes who?

1. What kind of research study should Shridhar undertake? Define the objectives of his research.

2. Do the stated objectives have scope for a qualitative research?

3. Which method(s) would you recommend and why?

4. Can you construct a template for conducting the study? What element would you advice Shridhar to keep in mind, and why?

Case Study 5.3

Candy-Ho! (A)

The evening sky was overcast. Looking out from the window of his office on the 12th floor, Sagar Ahuja could still see the etched out skyline of New Delhi. Sighing wearily, he turned his thoughts back to his comfortable job at Indore where he was marketing spicy Gujarati namkeen, and wondered what on earth he was doing in an alien city whose complexities and multiplicities seemed to defy any description to his simple mind. Having been a star performer at his regional office, and responsible for the launch of two revolutionary products for his company, he had been approached by head hunters to join Nefertiti—the famous global confectionary company in India. As his first assignment he had been given the job of swimming in deep waters and launch a new bubblegum that had been developed.

The Product

It was a sugar-coated, round-shaped, centre-filled liquid gel bubblegum in two flavours—strawberry and blueberry. The product was packed in mono pillow packs and was going to be priced at ₹1.00 per piece. The name of the product was to be *Moondrops.*

He had in front of him the results of a research conducted by Offspring Research Agency—a market research company specializing in child research studies.

Research Objectives

- To understand the meaning of a candy/bubblegum in a child's life.

- To analyse the response to two advertisements that had been created to market the bubblegum.

- To arrive at a decision on how to position and market the gum, and the advertisement that would be more suitable for the purpose.

Weighted base: Those whose favourite category is bubblegum and chewing gum	771
Like the taste/like to eat it	87
Soft to chew	26
Easily available everywhere	18
Helps in passing time/kills boredom/overcomes feeling of restlessness	18
Freshens breath	17
Taste you never get tired of/can keep eating repeatedly	11
Has variety of flavours	11
Not costly/Does not cost much	11
Improves taste of mouth/removes bad taste in mouth	10
Can be had any time of the day	10
Makes me feel happy/fun to have	9
Liked by my friends	7
Worth the price I pay for it/value for money	6

Data Source: Primary Research carried out by Nefertiti Company. Random Interviews with SEC A and B consumers equally split between male and female respondents, in the top eight cities, total sample size was 1,000 respondents.

FGD Analysis

The result of 24 focus groups across age groups and metros revealed the following data from a projective technique that involved personifying the bubblegum. The responses are across age groups and are in the decreasing order of most stated.

- *I want to play with my bubblegum*

- *The bubblegum has lots of friends— lot of names*

- *The bubblegum is very naughty—no one can catch him*

- *The bubblegum is my friend and helps me fight the older kids*
- *If all bubblegums were to fight, my bubblegum would win*
- *If I am feeling sad, my bubblegum would make me laugh*
- *My bubblegum is the bravest*

Post the FGC. Select respondents (children) were shown two advertisements. reaction to these are listed below:

(a) The Race Ad

The storyboard was that at a school annual function race, where the 'hero' of the story deliberately loses the race and comes third instead of first to get the third prize of two big jars of Moondrops. Followed by the punchline '*Moondrops ke liye kuch bhi ho sakta hai*'.

Reactions (With loud laughter)

All the kids were involved with the ad while viewing it and liked the storyboard with comments such as:

- '*It was interesting*'.
- '*Main soch raha tha ki yeh ladka ruk kyon gaya*'. (I was wondering why the boy stopped.)

The children enjoyed when the kid smiles with two big Moondrop jars in his hand.

- '*Jab who ladka race mein finish line ke pas aake ruk jata hai*'. (When the boy stops near the finish line.)
- '*Jab use third prize Moondrops milta hai aur use doorse do first and second prize wale ladke ghoor ke dekhte hain*'. (When he gets Moondrops as the third prize and the first and second prize winners stare at him.)
- *We feel proud to win a race even if we do not get any prize.*'
- '*If I win the race then Mummy and Daddy will anyway buy me Moondrops*'.
- '*Mein sirf Moondrops ke liye race nahin haroonga*'. (I'll never lose a race just for Moondrops.)
- '*Woh ladka buddhoo tha, kyonki usne jeeti hui race har di.*' (That boy was a foo1, as he lost a race that he was winning.)

The kids were surprised when the child stops just near the finish line and when the other two children are surprised and shocked that he is getting the Moondrops as the third prize.

Empathy/Relatability

Not many of the kids could relate to the ad. They did not see themselves doing the same just for getting two jars of Moondrops, the underlying reason being that they had to lose (If they could finish first, then why finish third).

(b) Kitty Party Ad

The story starts with a child returning from school to see a kitty party in progress at home (lots of fat aunties chatting and eating samosas and pakoras). One fat aunty pulls his cheek affectionately and much to his disgust, kisses him. He then feels happy when his reward is a Moondrop from the fat aunty. Seeing that he gets a Moondrop when the aunty kisses him, he plays a prank on all the aunties by jumping on the table and the sofa and kissing all the aunties there. His reward is lots of Moondrops. Followed by the punchline, '*Moondrops ke liye kuch bhi ho sakta hai*'.

Reactions

The scene where the fat aunty kisses the boy and they show her fat lips. The boy kissing the aunties by jumping on the sofa, on the table and by hugging an aunty.

- '*Jab who moti aunty ke lips dikhate hain*'. (When they show the fat aunty's lips.)
- '*Jab who moti aunty use kiss karti hain*'. (When the fat aunty kisses him.)
- '*Jab who sari aunties ko kiss karta hai aur aunties hairan ho jati hain*'. (When he surprises all the aunties by kissing them.)

Likeability

- '*Dekhne mein maza aaya*' (It was fun to watch.)
- '*Jab usne aunties ko kiss kiya to bahut accha laga*' (It was really good to see him kissing the aunties.)
- '*Aunty ka face itna funny tha, unko dekh ke hasi aayi*' (Aunty's face was so funny that we felt like laughing.)

Empathy/Relatability

- '*Chii, hum naughty nahin hain*' (Ugh, we are not naughty.)
- '*Aunty ko kiss nahin karenge, beizzati hoti hai.*' (Will not kiss the aunty, it is insulting.)
- '*Ganda lagta hai*'. (Don't like it.)
- '*Aunty ko kis karenge to manjan karna padega'.* (Will have to brush teeth if we kiss aunty.)

1. Can you help Mr Ahuja arrive at a decision?

5.7.1 Content Analysis

This technique involves studying a previously recorded or reported communication and systematically and objectively breaking it up into more manageable units that are related to the topic under study. It is peculiar in its nature that it is classified as a primary data collection technique and yet makes use of previously produced or secondary data. However, since the analysis is original, first hand and problem specific, it is categorized under primary methods. Some researchers classify it under observation methods, the reason being that in this, one is also analysing the communication in order to measure or infer about variables. The only difference being that one analyses communication that is ex-post facto rather than live. One can content-analyse letters, diaries, minutes of meetings, articles, audio and video recordings. The method is structured and systematic and thus of considerable credibility.

The first step involves defining U, or the *universe of content*. For example, in the case of Ritu, who wants to know what makes the young Indian tick, she could make use of the blogs written by youngsters, essays and reality shows featuring the age group. She decides that she wants to assess value systems, attitudes towards others/elders, clarity of life goal and peer influences. This step is extremely critical as this indicates the assumptions or hypotheses the researcher might have formulated.

This universe can be reported in any of five different formats (Berelson, 1954). The smallest reported unit could be a *word*. This is especially useful as it can be easily subjected to a computer analysis. In Ritu's case, the values that she wants to evaluate are individualistic or collectivistic, aggressive or compliant. Thus, she can sift the communication and place words such as 'I' or 'we' under the respective heads. Words like 'hate' 'dislike' go under aggression and 'alright' 'fine' 'maybe not so good' for complacency. Then counts and frequencies are calculated to arrive at certain conclusions.

The next level is a *theme*. This is very useful but, a little difficult to quantify as this involves reporting the propositions and sentences or events as representing a theme.

For example, disrespect towards elders is the theme and one picks out the following as a representative: a young teen's blog which says *my old man (father) has gone senile and needs to be sent to the looney bin for expecting me to become a space scientist, just because he could not become one……*

This categorization becomes more complex as the element of observer's bias comes into play. Thus, this kind of analysis could be extremely useful when carried out by an expert. However, in the case of an untrained analyst, the reliability and validity of the findings would be questionable.

The other units are *characters* and *space* and *time measures*. The character refers to the person producing the communication, for example the young teenager writing the blog. Space and time are more related to the physical format, i.e., the number of pages used, the length of the communication and the duration of the communication.

The last unit is the *item*, which is more Gestaltian in nature and refers to categorizing the entire communication as say 'responsible and respectful' or 'aggressive and amoral'. As in the case of theme, this categorization is equally complex as the observer's bias is likely to be high. Thus, to ensure the reliability of the findings, one may ask another coder to evaluate the same data. Cohen (1960) states the measuring of the percentage of agreement between the two analyses by the following formula:

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Here, Pr(a) is the relative observed agreement between the two raters. Pr(e) is the probability that this is due to chance. If the two raters are in complete agreement, then Kappa is 1. If there is no agreement, then Kappa = 0, 0.21–0.40 is fair, 0.41–0.80 is good and 0.81–1.00 is considered excellent.

Content analysis of large volumes becomes tedious and prone to error if handled by humans. Thus, there are various computer programmes available that can assist in the process. For computers running on Windows, one can use TEXTPACK, this is a dictionary word approach, where it can tag defined words for word frequency by sorting them alphabetically or by frequencies. Open-ended questions can be sorted by a programme called Verbastat (generally used by corporate users) or Statpac, which has an automatic coding module and is of considerable use to individual researchers.

Content analysis is a very useful technique when one has a large quantity of text as data and it needs to be structured in order to arrive at some definite conclusions about the variables under study. Computer assistance has greatly aided in the active usage of the technique. However, it can appear too simplistic, when one reduces the whole data to counts or frequencies.

5.8 SUMMARY

- Observations have led to some of the most important scientific discoveries in human history. Charles Darwin used his observations of animal and marine life at the Galapagos Islands to help him formulate his theory of evolution which he described in *On the Origin of Species.*

- Observation may be defined as 'a process in which one or more persons monitor some real-life situation and record pertinent occurrences'.

- In the process of 'participant observation', the observer becomes more or less one of the group members and may actually participate in some activity or the other of the group.

- All observations consist of two main components, the subject and the object. The subject refers to the observer, whereas the object refers to the activity or any type of operation that is being observed.

- Observation is a costly affair. It involves lot of expenses on travelling, staying at the places where the event is taking place and purchase of sophisticated equipment.

- A questionnaire is 'a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event'.

- Questionnaire that calls for short, check-mark responses are known as closed form type or restricted type. They have highly structured answers like mark a 'yes' or 'no', write a short response or check an item from a list of suggested responses.

- The open form or unrestricted questionnaire requires the respondent to answer the question in their own words.

- The three stages theory (also known as the sandwich theory) should be applied when sequencing questions. The order to be followed should be first, screening and rapport questions; second, the product specific questions; and third, demographic questions.

- Whoever administers the questionnaire, be it research staff, volunteers or whether self-administered by the respondents, it should have clear, detailed instructions.

- The respondents of a questionnaire may belong to a selected group. Therefore, the conclusions lack the kind of objectivity and representativeness essential for its validity.

- A schedule is a questionnaire containing a set of questions that are required to be answered to collect data about a particular item. A schedule is generally used in a face-to-face situation.

- Observation schedule is used to observe all the activities and record the responses of the respondents under some predefined conditions. The main idea behind examining the activities is to verify the required information.

- Data collection through a questionnaire is cheaper as compared to a schedule, as money is spent only in preparing the schedules and mailing them. In the schedule method, extra money is spent on appointing interviewers and imparting training to them.

- One of the main methods of data collection is conducting interviews. It takes place as a two-way conversation between the researcher and the respondent, whereby information is gathered by asking topic related questions.

- A proper setting for group interviews requires a group of not more than 10 to 12 persons with some social, intellectual, and educational homogeneity, which ensures effective participation by all.

- There are two chief means of recording opinion during the interview. If the question is preceded, the interviewer need only check a box or circle or code, or otherwise indicate which code comes closest to the respondent's opinion. If the question is not preceded, the interviewer is expected to record the response verbatim.

- A proper setting for group interviews requires a group of not more than 10 to 12 persons with some social, intellectual, and educational homogeneity, which ensures effective participation by all.

- Research interview is aimed at getting information required by the investigator to test his hypothesis or solve his problems of historical, experimental, survey or clinical type.

- Perhaps the most difficult part of the job of an interviewer is to obtain a specific, complete response. People can often be evasive and answer 'do not know' if they do not want to make the effort of thinking. They can also misunderstand the question and answer incorrectly in which case the interviewer would have to probe more deeply.

- A well-trained interviewer can obtain more data and greater clarity by altering the interview situation. This cannot be done in a questionnaire.

- An interview permits the research worker to follow-up leads as contrasted with the questionnaire.

- The idea of projecting oneself or one's feelings on to ambiguous objects is the basic assumption in projective techniques. The 19th century saw the origin of these techniques in clinical and developmental psychology.

- Ernest Dichter (1960) was one of the pioneers who used these techniques in consumer and motivational research. Consumer surveys and research were considered incomplete if they did not make use of projective techniques.

- In the theoretical sense, projective techniques unearth beliefs, attitudes and feelings that might underlie certain behaviour or interaction situations.

- Cartoon tests make use of animated characters in a particular situation (Masling, 1952). They are considered ambiguous as the figures bear no resemblance to a living being and thus are considered non-threatening.

- Case studies are discussions of individual cases under topics of discussion which help researchers to corroborate known facts proved previously through research.

- Researchers using case study method generally gather a large amount of data from a number of sources. Organizing this data in a systematic manner is a challenge in itself. The researcher should plan ahead to prevent getting overwhelmed by this data.

- An exemplary case study report transforms the manner in which a complex issue is presented. Case study reports are often published so that readers may apply the experience in his or her real-life situations. The case studies mostly display evidences to gain the reader's confidence.

- Content analysis involves studying a previously recorded or reported communication and systematically and objectively breaking it up into more manageable units that are related to the topic under study.

- Content analysis of large volumes becomes tedious and prone to error if handled by humans. Thus, there are various computer programmes available that can assist in the process. For computers running on Windows, one can use TEXTPACK, this is a dictionary word approach, where it can tag defined words for word frequency by sorting them alphabetically or by frequencies.

- Content analysis is a very useful technique when one has a large quantity of text as data and it needs to be structured in order to arrive at some definite conclusions about the variables under study.

5.9 KEY TERMS

- **Observation:** It may be defined as a process in which one or more persons monitor some real-life situation and record pertinent occurrences.

- **Questionnaire:** It is a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event.

- **Schedule:** It is a questionnaire containing a set of questions that are required to be answered to collect data about a particular item; it is generally used in a face-to-face situation.

- **Case studies:** They are discussions of individual cases under topics of discussion which help researchers to corroborate known facts proved previously through research.

5.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Observation may be defined as 'a process in which one or more persons monitor some real-life situation and record pertinent occurrences'.

2. All observations consist of two main components, the subject and the object. The subject refers to the observer, whereas the object refers to the activity or any type of operation that is being observed.

3. The disadvantages of observation are as follows:
   - It is very difficult to establish the validity of observations.
   - Many items of observation cannot be defined.

4. A questionnaire is 'a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event'.

5. Limitations of open questionnaire are as follows:
   - They are difficult to fill out.
   - The respondents may never be aware of all the possible answers.
   - They take longer to fill.

6. The three stages theory (also known as the sandwich theory) should be applied when sequencing questions. The order to be followed should be first, screening and rapport questions; second, the product specific questions; and third, demographic questions.

7. The limitations of the questionnaire method is:
   - **Limited response:** As noted earlier, this method cannot be used with illiterate or semi-illiterate groups. The number of persons who cooperate and respond to the questionnaire is very small.
   - **Lack of personal contact:** There is very little scope of personal contact in this method. In the absence of personal contact, very little can be done to persuade the respondents to fill up the questionnaire.

8. A schedule is a questionnaire containing a set of questions that are required to be answered to collect data about a particular item. A schedule is generally used in a face-to-face situation.

9. The interview schedule is used to ask the interviewee questions and record the responses in the space provided in the questionnaire itself.

10. According to Rummel J. Francis: 'The interview method of collecting data requires the actual physical proximity of two or more persons, and generally requires that all the normal channels of communication be open to their use. It is necessary to see one another, to hear each other's voices, to understand one another's language, and to use all that is psychologically inherent in physical proximity. It usually entails a non-reciprocal relation between the individuals concerned. One party desires to get information from another—one party interviews the other—for a particular purpose.'

11. Research interview is aimed at getting information required by the investigator to test his hypothesis or solve his problems of historical, experimental, survey or clinical type.

12. A tape recorder is very helpful in an interview because:

   - It reduces the tendency of the interviewer to make an unconscious selection of data favouring his/her biases.

   - The tape recorded data can be played more than once, and thus it permits a thorough study of the data.

13. The 19th century saw the origin of these techniques in clinical and developmental psychology.

14. Cartoon tests make use of animated characters in a particular situation (Masling, 1952). They are considered ambiguous as the figures bear no resemblance to a living being and thus are considered non-threatening.

15. Case studies are discussions of individual cases under topics of discussion which help researchers to corroborate known facts proved previously through research.

16. Critics feel that the case study method is not reliable enough for establishing a rule or principle as it portrays only a minuscule population which forms not even a part of the entire population.

17. Content analysis involves studying a previously recorded or reported communication and systematically and objectively breaking it up into more manageable units that are related to the topic under study.

5.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Differentiate between participant and non-participant observation.

2. What are the recording techniques of observation?

3. Enumerate the advantages of the observation method of data collection.

4. What are the types of questionnaires?

5. List the questions that should be avoided during the preparation of a questionnaire.

6. State the importance of the questionnaire method of data collection.

7. What are the characteristics of a good schedule?

8. What are the major differences between a schedule and a questionnaire?

9. What is the interview method of data collection? What are its types?

10. How is the attitude of a respondent significant for a researcher during an interview?

11. What are projective techniques? How is it different from the other methods of data collection?

12. Write a note on the evaluation of projective techniques.

13. How do case studies help the common man?

14. What is the content analysis technique of data collection?

Long-Answer Questions

1. Explain observation schedule as a tool of data collection.

2. Discuss the concept of questionnaire as a method of data collection.

3. Assess schedule as a method of data collection. Also, assess the difference between a schedule and a questionnaire.

4. List the merits and limitations of the schedule method of data collection.

5. Critically analyse how conducting interviews are an important source of data collection.

6. 'The idea of projecting oneself or one's feelings on to ambiguous objects is the basic assumption in projective techniques.' Explain projective method as a tool of data collection.

7. How are case studies an important source of data collection? What are the steps that should be used when utilizing the case study method?

8. Describe content analysis as a technique of data collection.

5.12 FURTHER READING

Chawla, D. and N. Sondhi. 2011. *Research Methodology*. New Delhi: Vikas Publishing House.

Kothari, C. R. 2008. *Research Methodology*. New Delhi: New Age Publishers.

Kumar, B. 2006. *Research Methodology*. New Delhi: Excel Books.

Paneerselvam, R. 2009. *Research Methodology*. New Delhi: Prentice Hall of India.

Payne, Geoff and Judy Payne. 2004. *Key Concepts in Social Research*. London: SAGE Publications.

Kumar, C. Rajender. 2008. *Research Methodology*. Delhi: APH Publishing Corporation.

Kumar, Ranjit. 2011. *Research Methodology: A Step-by-Step Guide for Beginners*. Australia: SAGE.

Gupta, D. 2011. *Research Methodology*. New Delhi: PHI Learning Private Limited.

UNIT 6 ANALYSIS AND USE OF STATISTICS

Structure

6.0 INTRODUCTION

The unit will begin with a discussion on data processing. The processing of data involves analysis and manipulation of the collected data by performing various functions. The data has to be processed in accordance with the outline laid down at the time of developing the research plan. Processing of data is essential for ensuring that all relevant data has been collected for performing comparisons and analyses. Editing of data involves the testing of data collection instruments in order to ensure maximum accuracy. Coding of data can be defined as representing the data symbolically using some predefined rules. You will also learn about the classification of data. Classification of data involves arrangement of data in groups or classes on the basis of some common characteristics. The methods of classification can be divided under the two headings: classification according to attributes and classification according to class intervals. Further, you will learn about the tabulation of data. In simple terms, tabulation means placing the data collected and results from research in a tabular form. Tabulation can be done by hand or mechanically using various electronic devices. Several factors like the size and type of study, cost considerations, time pressures and availability of tabulating machines decide the choice of tabulation. You will also learn about the analysis of data. Analysis of data is the process of transforming data for the purpose of extracting useful information,

which in turn facilitates the discovery of some useful conclusions. Finding conclusions from the analysed data is known as interpretation of data. However, if the analysis is done, in the case of experimental data or survey, then the value of the unknown parameters of the population and hypothesis testing is estimated. Moreover, you will learn about the statistical; tools for data analysis. There are certain basic statistical methods which can be classified into three groups: descriptive statistics, inferential statistics and measures of central tendency and dispersion. The unit will also discuss the measures of central tendency and dispersion.

6.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Describe the concept of data processing
- Explain the analysis of data and coding
- Discuss the tools tables, graphs and diagrams
- Assess the use of statistics including mean, median, mode and standard deviation

6.2 DATA PROCESSING

Research does not merely consist of data that is collected. Research is incomplete without proper analysis of the collected data. **Processing** of data involves analysis and manipulation of the collected data by performing various functions. The data has to be processed in accordance with the outline laid down at the time of developing the research plan. Processing of data is essential for ensuring that all relevant data has been collected for performing comparisons and analyses. The functions that can be performed on data are as follows:

- Editing
- Coding
- Tabulation
- Classification

Usually, experts are of the opinion that processing and analysis of data are inter-related. Therefore, they should be thought as one and the same thing. It is argued that analysis of data generally involves a number of closely-related operations, which are carried out with the objective of summarizing the collected data and organizing it in such a way that they are able to answer the research questions associated with it.

However, in technical terms, the processing of data involves data representation in such a way that it is open to analysis. Similarly, analysis of data is defined as the computation of certain measures along with searching for the patterns of relationship that may exist among data groups.

6.2.1 Editing of Data

Editing of data involves the testing of data collection instruments in order to ensure maximum accuracy. This includes checking the legibility, consistency and completeness of the data. The editing process aims at avoiding equivocation and ambiguity. The collected raw data is also examined to detect errors and omissions, if any. A careful

scrutiny is performed on the completed questionnaires and schedules to assure that the data has the following features:

- Accuracy
- Consistency
- Unity
- Uniformity
- Effective arrangement

The stages at which editing should be performed are as follows:

- **Field editing:** This involves reviewing the reporting forms by the investigator, that are written in abbreviated or illegible form by the informant at the time of recording the respondent's responses. Such type of editing must be done immediately after the interview. If performed after some time, such editing becomes complicated for the researcher, as it is difficult to decipher any particular individual's writing style. The investigator needs to be careful while doing such kind of editing and restrain the researcher from correcting errors or omission by guesswork.

- **Central editing:** This kind of editing involves a thorough editing of the entire data by a single editor or a team of editors. Such editing takes place when all the schedules created according to the research plan have been completed and returned to the researcher. Editors correct errors, such as data recorded in the wrong place or data recorded in months when it should be recorded in weeks. They can provide an appropriate answer to incorrect or missing replies by reviewing the other information in the schedule. At times, the respondent can be contacted for clarification. In some cases, if the answer is inappropriate or incomplete and an accurate answer cannot be determined on any basis, then the editor should delete or remove that answer from the collected data. He/she can put a note as 'no answer' in such a case. The answers that can be easily deciphered as wrong should be dropped from the final results.

Besides using the above-mentioned methods according to the data source, the researcher should also keep in mind certain points while doing the editing which are as follows:

- Familiarity with the instructions given to interviewers and coders
- Know-how of editing instructions
- Single-line striking for deleting an original entry
- Standardized and distinctive editing of data
- Initializing all the answers that have been changed

6.2.2 Coding of Data

Coding of data can be defined as representing the data symbolically using some predefined rules. Once data is coded and summarized, the researcher can analyse it and relationships can be found among various categories.

Checklist for Coding

The checklist enables the researcher to put the responses of the individuals into a limited number of categories or classes, which should possess the following important characteristics:

- Classes should be appropriate and in accordance with the research problem under consideration.
- There must be a class for every data element.
- There should be mutual exclusivity, which means that a specific answer can be placed in one and only one cell of a given category set.
- The classes should be one-dimensional. This means that every class is defined in terms of only one concept.

Significance of Coding

Coding of data is necessary for efficient analysis. It facilitates classification of data into a small number of classes. Thus, only important and critical information that is required for analysis is retained in the research. Coding decisions are usually taken at the designing stage of the questionnaire. This makes it possible to pre-code the questionnaire choices which, in turn, is helpful for computer tabulation.

However, in case of hand coding, some standard method may be used. One such method is to code in the margin with a coloured pencil. Another method is to transcribe data from the questionnaire to a coding sheet. Whatever method is adopted, you should note that coding errors are altogether eliminated or reduced to the minimum level.

6.2.3 Classification of Data

Research studies involve extensive collection of raw data and usage of the data to implement a research plan. To make the research plan easier, the data needs to be classified in different groups for understanding the relationship among different phases of research plan. Classification of data involves arrangement of data in groups or classes on the basis of some common characteristics. The methods of classification can be divided under the following two headings:

- Classification according to attributes.
- Classification according to class intervals.

Figure 6.1 shows the classification of data.
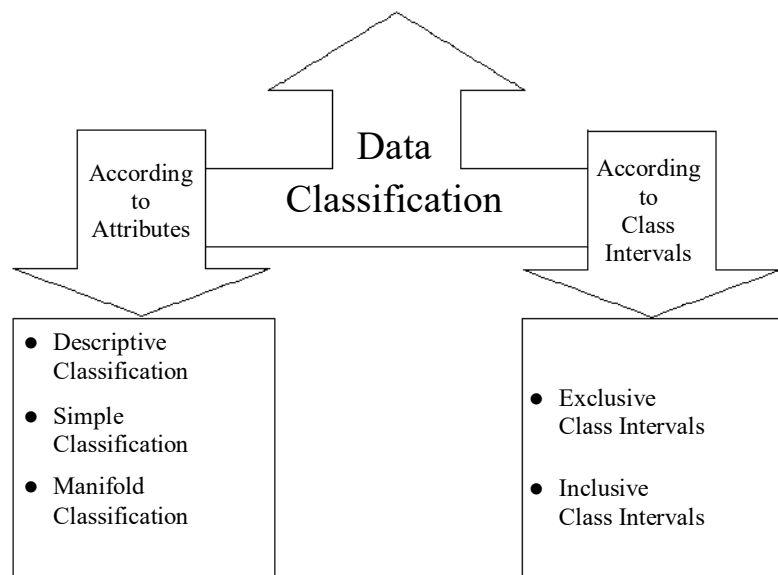
***Fig. 6.1** Data Classification*

Classification of Data According to Attributes

Data is classified on the basis of similar features, as follows:

- *Descriptive Classification:* This classification is performed according to the qualitative features and attributes, which cannot be measured quantitatively. These features can be either present or absent in an individual or an element. The features related to descriptive classification of attributes, can be literacy, sex, honesty, solidity, etc.

- *Simple Classification:* In this type of classification, the elements of data are categorized as those possessing the concerned attribute and those that do not.

- *Manifold Classification:* In this type of classification, two or more attributes are considered simultaneously and the data is categorized into a number of classes on the basis of those attributes. The total number of classes of final order is given by 2^n, where n = number of attributes considered.

Classification of Data According to Class Intervals

Classifying data according to class intervals is a quantitative phenomenon. Class intervals help categorize data which has similar numerical characteristics, such as income, production, age, weight, etc. Data can be measured with certain statistical tools like mean, mode and median. The different categories of data according to class intervals are as follows:

- *Statistics of Variables:* This term refers to measurable attributes, as these typically vary over time or between individuals. The variables can be discrete, i.e., taking values from a countable or finite set, continuous, i.e., having a continuous distribution function or neither. This concept of a variable is widely utilized in the social, natural, medical sciences.

- *Class Intervals:* These refer to a range of values of a variable. This interval is used to calibrate the scale of a variable in order to tabulate the frequency distribution of a sample. A suitable example of such data classification can be the categorizing of birth rate in a country. In this case, babies aged 0–1 year will form a group; those aged 2–5 years will form another group, and so on. The entire data is thus categorized into several numbers of groups or classes or in other words, class intervals. Each class interval has an upper limit as well as a lower limit, which is defined as 'the class limit.' The difference between two class limits is known as class magnitude. The classes can have equal or unequal class magnitudes.

  The number of elements in a given class is called the frequency of the given class interval. All class intervals, with their respective frequencies, are taken together and described in a tabular form called the frequency distribution.

Problems Related to Classification of Data

The problems related to classification of data on the basis of class intervals are divided into the following three categories:

(a) *Number of classes and their magnitude:* There are differences regarding the number of classes into which data can be classified. As such, there are no predefined rules for classification of data. It all depends upon the skill and experience of the researcher. The researcher should display the data in such a way that it should be clear and meaningful to the analyst.

As regards the magnitude of classes, it is usually held that class intervals should be of equal magnitude, but in some cases unequal magnitudes may result in better classification. It is the researcher's objective and judgement that plays a significant role in this regard. In general, multiples of two, five and ten are preferred while determining class magnitudes. H.A. Sturges suggested the following formula for determining the size of class interval:

$$i = R /(1 + 3.3 \log N)$$

Where,

i = Size of class interval.

R = Range (difference between the values of the largest element and smallest element among the given elements).

N = Number of items to be grouped.

Sometimes, data may contain one or two or very few elements with very high or very low values. In such cases, the researcher can use an open-ended interval in the overall frequency distribution. Such intervals can be expressed below two years; or twelve years and above. However, such intervals are not desirable, yet cannot be avoided.

(b) *Choice of class limits:* While choosing class limits, the researcher must determine the mid-point of a class interval. A mid-point is generally derived by taking the sum of the upper and lower limit, of a class and then dividing it by two. The actual average of elements of that class interval should remain as close to each other as possible. In accordance with this principle, the class limits should be located at multiples of two, five, ten, twenty and hundred and such other figures. The class limits can generally be stated in any of the following forms:

o **Exclusive Type Class Intervals:** These intervals are usually stated as follows:

- 10–20
- 20–30
- 30–40
- 40–50

These intervals should be read in the following way:

- 10 and under 20
- 20 and under 30
- 30 and under 40
- 40 and under 50

In the exclusive type of class interval, the elements whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly thirty would be put in 30–40 class interval and not in 20–30 class interval. In other words, an exclusive type of class interval is that in which the upper limit of a class interval is excluded and items with values less than the upper limit, but not less than the lower limit, are put in the given class interval.

o **Inclusive Type Class Intervals:** These intervals are normally stated as follows:

- 11–20
- 21–30
- 31–40
- 41–50

This should be read as follows:

- 11 and under 21
- 21 and under 31
- 31 and under 41
- 41 and under 51

In this method, the upper limit of a class interval is also included in the concerning class interval. Thus, an element whose value is twenty will be put in 11–20 class interval. The stated upper limit of the class interval 11–20 is twenty but the real upper limit is 20.999999 and as such 11–20 class interval really means eleven and under twenty-one. When data to be classified happens to be a discrete one, then the inclusive type of classification should be applied. But when data happens to be a continuous one, the exclusive type of class intervals can be used.

(c) *Determining the frequency of each class:* The frequency of each class can be determined using tally sheets or mechanical aids. In tally sheets, the class groups are written on a sheet of paper and for each item a stroke (a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the element falling in the same group is indicated as a diagonal line through the above said four lines. This enables the researcher to perform the counting of elements in each one of the class groups. Table 6.1 shows a hypothetical tally sheet.

Table 6.1 Sample of a Tally Sheet

Income groups (Rupees)	Tally mark	Number of families (Class frequency)
Below 600	ᚋ ᚋ ᚋ III	15
601-900	ᚋ ᚋ I	9
901-1300	ᚋ ᚋ ᚋ ᚋ ᚋ ᚋ I	25
1301-1500	ᚋ ᚋ ᚋ IIII	16
1501 and above	ᚋ ᚋ II	10
Total		75

In case of large inquiries and surveys, class frequencies can be determined by means of mechanical aids, i.e., with the help of machines. Such machines function either manually or automatically and run on electricity and can sort cards at a speed of around 25,000 cards per hour. Although this method increases the speed, it is an expensive method.

6.2.4 Tabulation of Data

In simple terms, tabulation means placing the data collected and results from research in a tabular form.

Methods of Tabulation

Tabulation can be done by hand or mechanically using various electronic devices. Several factors like the size and type of study, cost considerations, time pressures and availability of tabulating machines decide the choice of tabulation. Relatively, large data requires computer tabulation. Hand tabulation is preferred in case of small inquiries, when the number of questionnaires is small and they are of relatively short length. The different methods used in hand tabulation are as follows:

- *Direct tally method:* This method involves simple codes, where the researcher can directly tally from the questionnaire. The codes are written on a sheet of paper called tally sheet and for each response, a stroke is marked against the code in which it falls. Usually, after every four strokes against a particular code, the fifth response is indicated by drawing a diagonal or horizontal line through the strokes. These groups are easy to count and the data is sorted against each code conveniently.

- *List and tally method:* In this method, code responses may be transcribed into a large worksheet, allowing a line for each questionnaire. This facilitates listing of a large number of questionnaires in one worksheet. The tallies are then made for each question.

- *Card sort method:* This is the most flexible hand tabulation method, where the data is recorded on special cards of convenient sizes and shapes with a series of holes. Each hole in the card stands for a code. When the cards are stacked, a needle passes through a particular hole, thus representing a particular code. These cards are then segregated and counted. In this way, frequencies of various codes can be found out by a repetition of this technique.

Significance of Tabulation

Tabulation enables the researcher to arrange data in a concise and logical order. It summarizes the raw data and displays the same in a compact form for further analysis. It helps in the orderly arrangement of data in rows and columns. The various advantages of tabulation of data are as follows:

- A table saves space and reduces descriptive and explanatory statements to a minimum.

- It facilitates and eases the comparison process.

- The summation of elements and detection of omissions and errors become easy due to a tabular description.

- A table provides a basis for statistical computations.

Checklist for Tables

A table should communicate the required information to the reader in such a way that it becomes easy for him/her to read, comprehend and recall the information when required. There are certain conventions to be followed during tabulation, which are as follows:

- All tables should have a clear, precise and adequate title to make them intelligible enough without any reference to the text.

- Tables should be featured with clarity and readability.

- Every table should be given a distinct number to facilitate easy reference.

- The table should be of an appropriate size and tally with the required information.

- The columns and rows should be headed with bold font letters. It is a general rule to include independent variables in the left column or first row and dependent variables in the bottom row or right column.

- Displaying of numbers should be neat and readable.

- Explanatory footnotes, if any, regarding the table should be placed directly beneath the table, along with the reference symbols used in the table.

- The source of the table should be indicated just below the table.

- The table should contain thick lines to separate data of one class from data of another class and thin lines to separate the different subdivisions of each class.

- All column figures should be properly aligned.

- Abbreviations should be avoided in a table to the best possible extent.

- If the volume of data happens to be large, then it should not be crowded in a single table. It makes the table unwieldy and inconvenient.

Tabulation can also be classified into complex or simple. The former type of tabulation gives information about one or more groups of independent variables, whereas the latter shows the division of data into two or more categories.

6.3   ANALYSIS OF DATA

Analysis of data is the process of transforming data for the purpose of extracting useful information, which in turn facilitates the discovery of some useful conclusions. Finding conclusions from the analysed data is known as interpretation of data. However, if the analysis is done, in the case of experimental data or survey, then the value of the unknown parameters of the population and hypothesis testing is estimated.

Analysis of data can be either descriptive or inferential. Inferential analysis is also known as statistical analysis. Descriptive analysis is used to describe the basic features of the data in a study, such as persons, work groups and organizations. Inferential analysis is used to make inferences from the data, which means that we are trying to understand some process and make some possible predictions based on this understanding.

6.3.1   Types of Analysis

The various types of analyses are as follows:

- **Multiple Regression Analysis:** This type of analysis is used to predict a single dependent variable by a set of independent variables. In multiple regression analysis, the independent variables are not correlated to each other.

- **Multiple Discriminant Analysis:** In multiple discriminant analysis, there is one single dependent variable, which is very difficult to measure. One of the main objectives of this type of analysis is to understand group differences and predict the likelihood that an entity, i.e., an individual or an object, belongs to a particular class or group based on several metric-independent variables.

Check Your Progress

1. State the functions that can be performed on data.
2. When are coding decisions taken?
3. Define class intervals.
4. List the different methods used in hand tabulation.

- **Canonical Correlation Analysis:** It is a method for assessing the relationship between variables. This analysis also allows you to investigate the relationship between two sets of variables.

Univariate, Bivariate and Multivariate Analysis

Many types of analyses are performed according to the variance that exists in the data. Such analyses is carried out to check if the differences between three or more variables are significant enough to evaluate them statistically. There are three types of such analyses, namely univariate, bivariate and multivariate analyses.

(i) **Univariate Analysis:** In this analysis, only a single variable is taken into consideration. It is usually the first activity pursued while analysing the data. It is performed with the purpose of describing each variable in terms of mean, median or mode, and variability. Examples of such analysis are averages or a set of cases that may come under a specific category amidst a whole sample.

(ii) **Bivariate Analysis:** This type of analysis examines the relationship between two variables. It tries to find the extent of association that exists among these variables. Thus, a bivariate analysis may help you; for example, to find whether the variables of irregular meals and migraine headaches are associated and up to what extent. Here, the two variables are thus statistically measured simultaneously.

(iii) **Multivariate Analysis:** This type of analysis involves observation and analysis of three or more than three statistical variables at a time. Such an analysis is performed using statistical tests or even in a tabular format. Thus, for example, you can study the variables of age, educational qualification and annual income of a given set of population at the same time using the multivariate analysis method.

Table 6.2(a) *Bivariate Table*

	1991	**1992**	**1993**
Percentage of students failed	33 per cent	38 per cent	42 per cent
Percentage of students passed	67 per cent	62 per cent	58 per cent

Although the data in both tables is related, except the variable of 'attempts', the multivariate table has been displayed separately in this example. However, you should note that the tables have dealt simultaneously with two or more variables of the data.

Table 6.2(b) *Multivariate Table*

	1991	**1992**	**1993**
	First Attempt	Second Attempt	Third Attempt
Percentage of students who passed in Maths	27 per cent	35 per cent	–
Percentage of students who passed in English	53 per cent	60 per cent	44 per cent

6.3.2 Data Interpretation

Data interpretation refers to the identification of trends in different variables. The researcher uses statistics for this purpose. The researcher is required to be familiar with the knowledge of the scales of measurement. This enables him/her to choose the appropriate statistical method for his/her research project. The scales of measurement facilitate the allotment of numerical values to characteristics adhering to any specific rules. This measurement is also related to such levels of measurement of data like nominal, ordinal and internal and ratio levels. These levels can be explained as follows:

- **Nominal Measurement:** The nominal measurement assigns a numeral value to a specific characteristic. It is the fundamental form of measurement. The nominal measurement calculates the lowest level of data available for measurement.

- **Ordinal Measurement:** This type of measurement involves allotting a specific feature to numeral value in terms of a specific order. The ordinal scale displays the way in which an entity is measured. The ordinal scale of measurement is used to calculate and derive data pertaining to the median, percentage, rank order, correlations and percentile.

- **Interval Measurement:** A researcher can depict the difference between the first aspect of a data and another aspect using this level of measurement. The interval scale of measurement is useful for the researcher in several ways. It can be applied in the calculation of arithmetic mean, averages, standard deviations and determining the correlation between different variables.

- **Ratio Measurement:** In this method, there are fixed proportions (ratio) between the numerical data and the amount of the characteristics that it represents. A researcher should remember while measuring the ratio levels that a fixed zero point exists. The ratio level of measurement facilitates researchers in determining the appropriate data, if the aspects possess any certain characteristic. Almost any type of arithmetical calculations can be executed using this scale of measurement.

The most important feature of any measuring scale is its reliability and validity, which is explained as follows:

- **Reliability:** It is the term used to deal with accuracy. A scale measurement can be said to be reliable, when it exactly measures, only that what it is supposed to measure. In other words, when the same researcher repeats a test, i.e., with a different group but resembling the original group, he/she should get the same results as the former.

- **Validity:** According to Leedy, validity is the assessment of the soundness and the effectiveness of the measuring instrument. There are three types of validity, which can be stated as follows:

  o *Content Validity:* It deals with the accuracy with which an instrument measures the factors or content of the course or situations of the research study.

  o *Prognostic Validity:* It depends on the possibility to make judgements from results obtained by the concerned measuring instrument. The judgement is future oriented.

  o *Simultaneous Validity:* This involves comparing of one measuring instrument with another; one that measures the same characteristic and is available immediately.

Check Your Progress

5. Define bivariate analysis.

6. When is the ordinal scale of measurement used?

6.4 REPRESENTATION OF DATA

There can be a large amount of raw data collected from different sources. Such amount of data becomes quite cumbersome and confusing for the researcher to handle and analyse. It is almost impossible for the researcher to deal with all this data in raw form. Therefore, such data must be presented in a suitable and summarized form without any loss of relevant information so that it can be efficiently used for decision-making. Hence, we construct appropriate tables, graphs and diagrams to interpret, and summarize the entire set of raw data.

In view of the ever-increasing importance of statistical data in business operations and their management, the presentation of data in the form of graphs, tables and diagrams, their importance and use have been discussed in this section.

6.4.1 Tables

Classification of data is usually followed by tabulation, which is considered as the mechanical part of classification.

Tabulation is the systematic arrangement of data in columns and rows. The analysis of the data is done so by arranging the columns and rows to facilitate analysis and comparisons.

Tabulation has the following objectives:

(i) Simplicity is one of the main objectives of tabulation. The removal of unnecessary details gives a clear and concise picture of the data.

(ii) Economy of space and time.

(iii) Ease in comprehension and remembering.

(iv) Facility of comparisons is another main objectives of tabulation. Comparisons within a table and with other tables may be made.

(v) Ease in handling of totals, analysis, interpretation, etc.

Construction of Tables

A table is constructed depending on the type of information to be presented and the requirements of statistical analysis. The following are the essential features of a table:

(i) **Title:** It should have a clear and relevant title, which describes the contents of the table. The title should be brief and self explanatory.

(ii) **Stubs and captions:** It should have clear headings and sub headings. Column headings are called captions and row headings are called stubs. The stubs are usually wider than the captions.

(iii) **Unit:** It should indicate all the units used.

(iv) **Body:** The body of the table should contain all information arranged according to description.

(v) **Headnote:** The headnote or prefatory note, placed just below the title, in a less prominent type, gives some additional explanation about the table. Sometimes, the headnote consists of the unit of measurement.

(vi) **Footnotes:** A footnote at the bottom of the table may clarify some omissions of special features.

(vii) **Source:** A source note gives information about the source used, if any.

(viii) **Arrangement of data:** Data may be arranged according to requirements in chronological, alphabetical, geographical, or any other order.

(ix) **Emphasis:** The items to be emphasized may be put in different print or marked suitably.

(x) **Other details:** Percentages, ratios, etc. should be shown in separate columns. Thick and thin lines should be drawn at proper places.

A table should be easy to read and should contain only the relevant details. If the aim of clarification is not achieved, the table should be redesigned.

Types of Tables

Depending on the nature of the data and other requirements, tables may be divided into various types. They are given below:

(i) **General Tables or Reference Tables:** These contain detailed information for general use and reference, e.g., tables published by government agencies.

(ii) **Specific Purpose or Derivative Tables:** They are usually summarized from general tables and are useful for comparison and analytical purposes. Averages, percentages, etc., are incorporated along with information in these tables.

(iii) **Simple and Complex Tables:** A table showing only one characteristic is a simple table. The tables that show two or more characteristics or groups of items are termed as complex tables (see Tables 6.3 and 6.4).

Table 6.3 Simple Table

Cinema Attendance among Adult Male Factory Workers in Bombay March 1972

Frequency	Number of Workers
Less than once a month	3780
1 to 4 times a month	1652
More than 4 times a month	926

Table 6.4 is the result of a survey on the cinema-going habits of adult factory workers.

Table 6.4 Complex Table

Cinema Attendance among Adult Male Factory Workers in Bombay March 1972

Cinema Attendance Frequency	Single		Married	
	Under 30	Over 30	Under 30	Over 30
Less than once a month	122	374	1404	1880
1–4 times a month	1046	202	289	115
More than 4 times a month	881	23	112	10
Total	2049	599	1805	2005

It is obvious that the tabular form of classification of data is a great improvement over the narrative form.

Frequently, table construction involves deciding which attribute should be taken as primary and which as secondary. For the previous table, we can also consider that whether it would be improved further if 'under 30' and '30 and over' had been the main

column headings and 'single' and 'married' the sub headings. The modifications depend on the purpose of the table. If the activities of age groups are to be compared, it is best left as it stands. But if a comparison between men of different marital status is required, the change would be an improvement.

Advantages of Tabulating of Data

The following are the advantages of tabulating data:

  (i) Tabulated data can be more easily understood and grasped than untabulated data.

 (ii) A table facilitates comparisons between subdivisions and with other tables.

(iii) It enables the required figures to be located easily.

(iv) It reveals patterns within the figures, which otherwise might not have been obvious, e.g., from the previous table, we can conclude that regular and frequent cinema attendance is mainly confined to younger age group.

 (v) It makes the summation of items and the detection of errors and omissions, easier.

(vi) It obviates repetition of explanatory phrases and headings and hence takes less space.

6.4.2 Graphs

In a graph, the independent variable should always be placed on the horizontal or x-axis and the dependent variable on the vertical or y-axis.

Line Graphs

Here, the points are plotted on paper (or graph paper) and joined by straight lines. Generally, continuous variables are plotted by the line graph.

Example 6.1: The monthly averages of Retail Price Index from 1996 to 2003 (Jan. 1996 = 100) were as follows:

Year	1996	1997	1998	1999	2000	2001	2002	2003
Retail Price Index	100	105.8	109.0	109.6	110.7	114.5	119.3	122.3

Draw a diagram to display these figures.

Solution: Here, years are plotted along the horizontal line and the retail price index along the vertical line.

Erect perpendiculars to horizontal line from the points marked as retail price index for the years 1997, 1998, ..., 2003 and cut off these ordinates according to the given data and thus various points will be plotted on the paper. Join these points by straight lines.
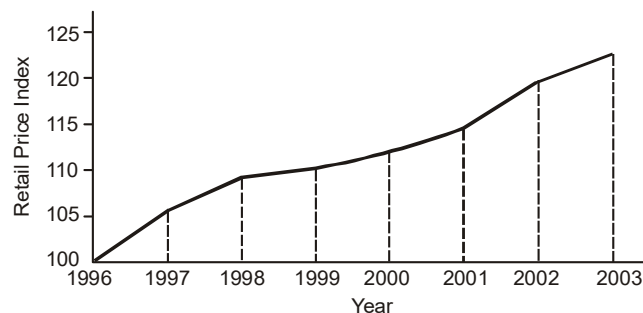
Frequency Polygon

A frequency polygon is a line chart of frequency distribution in which, either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines. Since the frequencies generally do not start at zero or end at zero, this diagram as such would not touch the horizontal axis. However, since the area under the entire curve is the same as that of a histogram which is 100 per cent of the data presented, the curve can be enclosed so that the starting point is joined with a fictitious preceding point whose value is zero. This ensures that the start of the curve is at horizontal axis and the last point is joined with a fictitious succeeding point whose value is also zero, so that the curve ends at the horizontal axis. This enclosed diagram is known as the frequency polygon.

We can construct the frequency polygon from the table presented for the ages of 30 workers as shown in Figure 6.2.
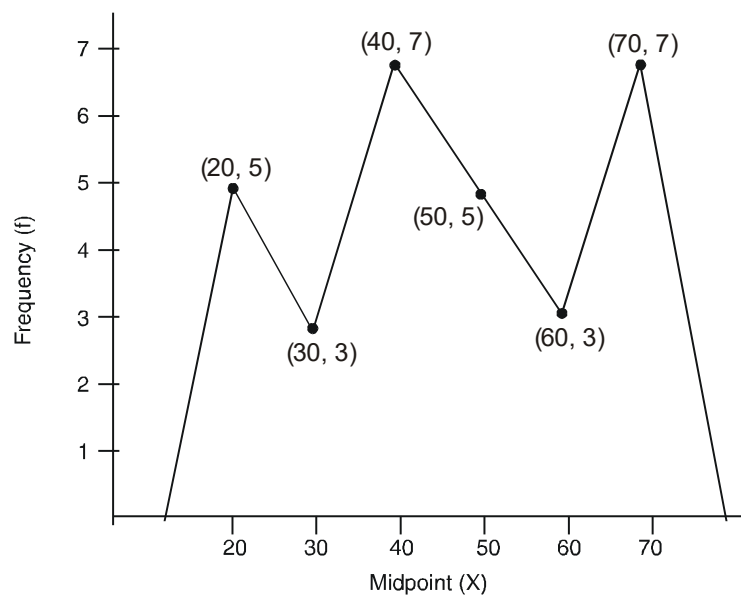
Fig. 6.2 *Frequency Polygon*

Relative Frequencies

In a frequency distribution, if the frequency in each class interval is converted into a proportion, dividing it by the total frequency, we get a series of proportions called *relative frequencies*. A distribution presented with relative frequencies rather than actual frequencies is called a *relative frequency* distribution. The sum of all relative frequencies in a distribution is 1.

Ogive Curves

Cumulative frequency curve or ogive is the graphic representation of a cumulative frequency distribution. Ogives are of two types. One of these is less than and the other one is greater than ogive. Both these ogives are constructed based upon the following table of our example of 30 workers.

Class Interval (years)	Mid-point	(f)	Cum. Freq. (less than)	Cum. Freq. (greater than)
15 and upto 25	20	5	5 (less than 25)	30 (more than 15)
25 and upto 35	30	3	8 (less than 35)	25 (more than 25)
35 and upto 45	40	7	15 (less than 45)	22 (more than 35)
45 and upto 55	50	5	20 (less than 55)	15 (more than 45)
55 and upto 65	60	3	23 (less than 65)	10 (more than 55)
65 and upto 75	70	7	30 (less than 75)	7 (more than 65)

(*i*) **Less than Ogive:** In this case, the less than cumulative frequencies are plotted against the upper boundaries of their respective class intervals. Less than Ogive is shown in Figure 6.3.
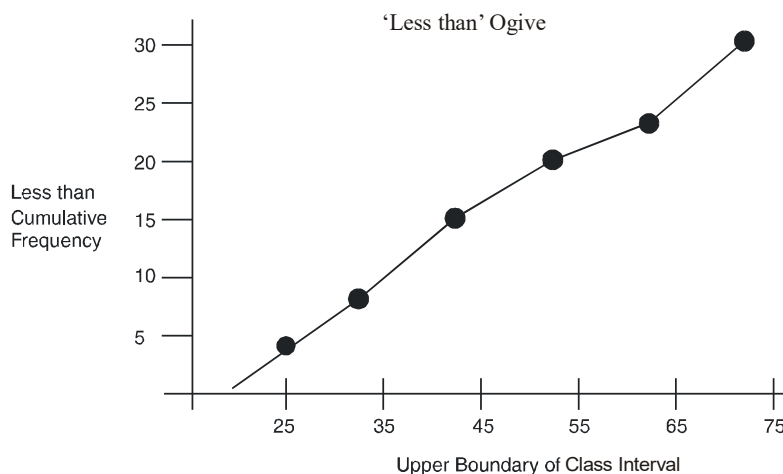
Fig. 6.3 *Less than Ogive*

(*ii*) **Greater or More than Ogive.** In this case, the greater than cumulative frequencies are plotted against the lower boundaries of their respective class intervals. Greater than ogive is shown in Figure 6.4.

These ogives can be used for comparison purposes. Several ogives can be drawn on the same grid, preferably with different colours for easier visualization and differentiation.

Although, diagrams and graphs are powerful and effective media for presenting statistical data, they can only represent a limited amount of information and they are not of much help when intensive analysis of data is required.
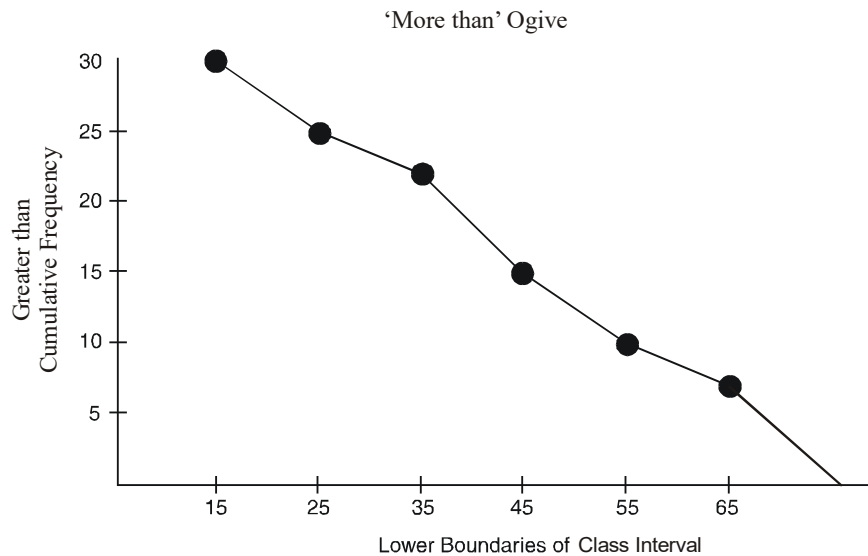
Fig. 6.4 Greater than Ogive

Histograms

A histogram is the graphical description of data and is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.

The word histogram is derived from the Greek word 'histos' which means 'anything set upright' and 'gramma' which means 'drawing, record, writing'. It is considered as the most important basic tool of statistical quality control process.

In this type of representation, the given data is plotted in the form of a series of rectangles. Class intervals are marked along the *X*-axis and the frequencies along the *Y*-axis according to a suitable scale. Unlike the bar chart, which is one dimensional, meaning that only the length of the bar is important and not the width, a histogram is two-dimensional in which both the length and the width are important. A histogram is constructed from a frequency distribution of a grouped data, where the height of the rectangle is proportional to the respective frequency and the width represents the class interval. Each rectangle is joined with the other and any blank spaces between the rectangles would mean that the category is empty and there are no values in that class interval.

As an example, let us construct a histogram for our example of ages of 30 workers. For convenience sake, we will present the frequency distribution along with the midpoint of each interval, where the midpoint is simply the average of the values of the lower and the upper boundary of each class interval. The frequency distribution is shown in Table 6.5.

Table 6.5 Frequency Distribution Table

Class Interval (years)	Mid-Point	(f)
15 and upto 25	20	5
25 and upto 35	30	3
35 and upto 45	40	7
45 and upto 55	50	5
55 and upto 65	60	3
65 and upto 75	70	7

The histogram of this data would be shown in Figure 6.5.
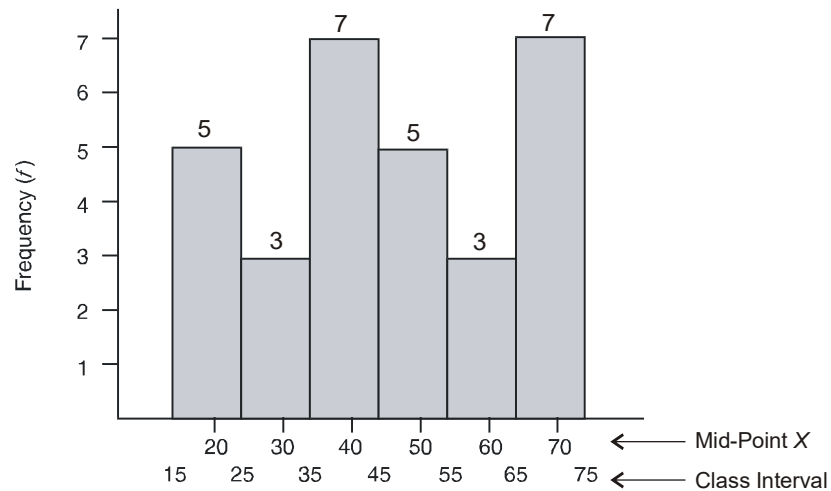
Fig. 6.5 Histogram

6.4.3 Diagrams

The data we collect can often be more easily understood for interpretation if it is presented graphically or pictorially. Diagrams and graphs give visual indications of magnitudes, groupings, trends and patterns in the data. These important features are more simply presented in the form of graphs. Also, diagrams facilitate comparisons between two or more sets of data.

The diagrams should be clear and easy to read and understand. Too much information should not be shown in the same diagram; otherwise, it may become cumbersome and confusing. Each diagram should include a brief and self explanatory title dealing with the subject matter. The scale of the presentation should be chosen in such a way that the resulting diagram is of appropriate size. The intervals on the vertical as well as the horizontal axis should be of equal size; otherwise, distortions would occur.

Diagrams are more suitable to illustrate the data which is discrete, while continuous data is better represented by graphs. The following are the diagrammatic and the graphic representation methods that are commonly used.

One-Dimensional Diagrams

Bars are simply vertical lines where the lengths of the bars are proportional to their corresponding numerical values. The width of the bar is unimportant but all bars should have the same width so as not to confuse the reader of the diagram. Additionally, the bars should be equally spaced.

Example 6.2: Suppose that the following were the gross revenues (in $100,000.00) for a company *XYZ* for the years 1989, 1990 and 1991.

Year	*Revenue*
1989	110
1990	95
1991	65

Construct a bar diagram for this data.

Solution: The bar diagram for this data can be constructed as follows with the revenues represented on the vertical axis and the years represented on the horizontal axis.

A Bar Diagram
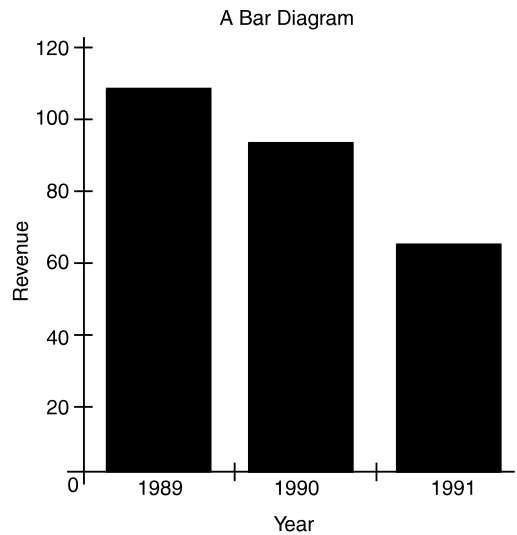
When each figure is made up of two or more component figures, the bars may be subdivided into components. Too many components should not be shown.
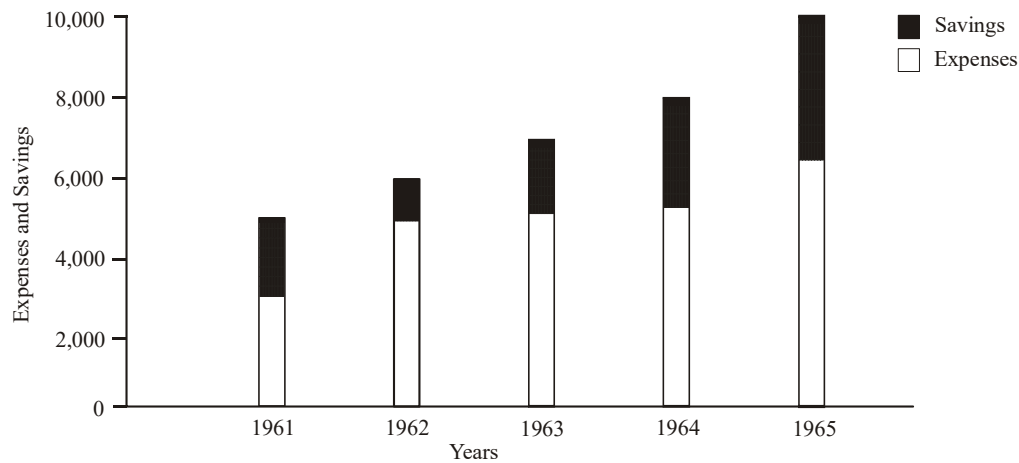
Component Bar Chart Showing Expenses and Savings of Mr X

The following shows the Annual Income, Expenses and Savings of Mr X:

Year	Amounts in ₹ of			Percentages of		
	Income	Expenses	Savings	Income	Expenses	Savings
1961	5000	3000	2000	100.0	60.0	40.0
1962	6000	5000	1000	100.0	83.3	16.7
1963	7000	5000	2000	100.0	71.4	28.6
1964	8000	5000	3000	100.0	62.5	37.5
1965	10000	6000	4000	100.0	60.0	40.0

The bars drawn can be further subdivided into components depending upon the type of information to be shown in the diagram (see Figure 6.3 and Table 6.4). This will be clear by the following example in which we present three components in a bar.

Multiple Bar Charts

The multiple bar charts for showing expenses and savings of Mr X is shown in Figure 6.6.

Here, the interrelated component parts are shown in adjoining bars, coloured or marked differently, thus allowing comparison between different parts.

These charts can be used if the overall total is not required. Some charts given earlier show totals also.

Two-Dimensional Diagrams

Two dimensional diagrams take two components of data for representation. These are also called area diagrams as it considers two dimensions. The types are rectangles, squares and pie. It can be best explained with the help of the following squares diagram example:
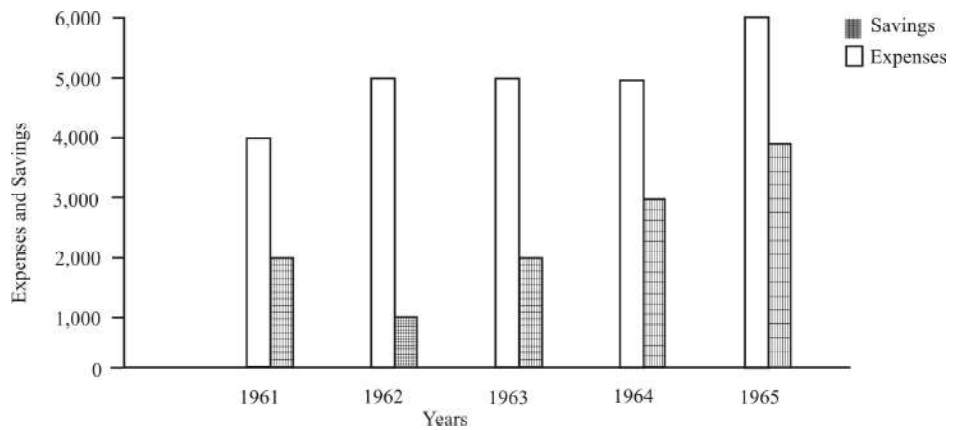
***Fig. 6.6** Multiple Bar Chart showing Expenses and Savings of Mr X*

Squares: The square diagram is easy and simple to draw. Take the square root of the values of various given items that are to be shown in the diagrams and then select a suitable scale to draw the squares.

Pie Diagrams

This type of diagram enables us to show the partitioning of a total into its component parts. The diagram is in the form of a circle and is also called a pie because the entire diagram looks like a pie and the components resemble slices cut from it. The size of the slice represents the proportion of the component out of the whole.

Example 6.3: The following figures relate to the cost of the construction of a house. The various components of cost that go into it are represented as percentages of the total cost.

Item	% Expenditure
Labour	25
Cement, bricks	30
Steel	15
Timber, glass	20
Miscellaneous	10

Construct a pie chart for the above data.

Solution: The pie chart for this data is presented as follows:
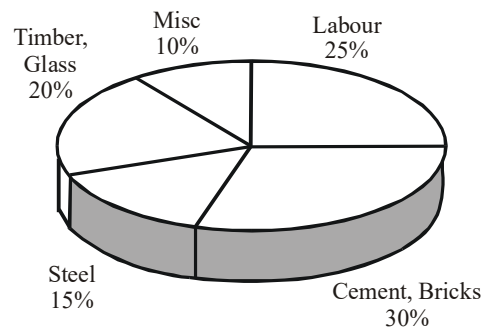
Pie charts are very useful for comparison purposes, especially when there are only a few components. If there are too many components, it may become confusing to differentiate the relative values in the pie.

Three-Dimensional Diagrams

Three dimensional diagrams are also termed as volume diagram and consist of cubes, cylinders, spheres, etc. In these diagrams, three dimensions namely length, width and height are taken into account. Cubes are used to represent where side of a cube is drawn in proportion to the cube root of the magnitude of data.

Example 6.4: Represent the following data using volume diagram.

Category	Number of Students
Under graduate	64000
Post graduate	27000
Professionals	8000

Solution: The sides of cubes are calculated as follows:

Category	Number of Students	Cube Root	Side of Cube
Undergraduate	64000	40	4 cm
Postgraduate	27000	30	3 cm
Professional	8000	20	2 cm
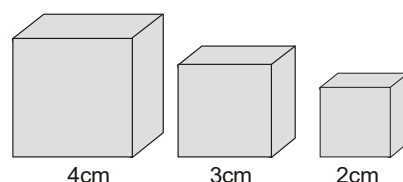
6.5 USE OF STATISTICS

In the preceding section, you learnt the techniques of data collection, summarizing and condensing data into various tables, diagrams and charts. Now, let us discuss arithmetic procedures that can be used for analysing and interpreting quantitative data. These measures and procedures relate to some properties and characteristics of data which include measures of central location of data, other measures of non-central location, measures of dispersion of data in itself and around the mean and the shape of the data.

6.5.1 Mean

This section pertains to arithmetic mean.

Arithmetic Mean

There are several commonly used measures such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data but also for the purpose of making comparisons among two or more sets of data.

As an example, questions like 'How hot is the month of June in Delhi?' can be answered generally by a single figure of the average for that month. Similarly, suppose we want to find out if boys and girls of age 10 years differ in height for the purpose of making comparisons. Then, by taking the average height of boys of that age and the average height of girls of the same age, we can compare and record the differences.

While arithmetic mean is the most commonly used measure of central tendency, mode and median are more suitable measures under certain set of conditions and for certain types of data. However, each measure of central tendency should meet the following requisites.

- It should be easy to calculate and understand.

- It should be rigidly defined. It should have only one interpretation so that the personal prejudice or the bias of the investigator does not affect its usefulness.

- It should be representative of the data. If it is calculated from a sample, the sample should be random enough to be accurately representing the population.

- It should have a sampling stability. It should not be affected by sampling fluctuations. This means that if we pick ten different groups of college students at random and compute the average of each group, then we should expect to get approximately the same value from each of these groups.

- It should not be affected much by extreme values. If few, very small or very large items are present in the data, they will unduly influence the value of the average by shifting it to one side or other, so that the average would not be really typical of the entire series. Hence, the average chosen should be such that it is not unduly affected by such extreme values.

Arithmetic mean is also commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost everyone in daily communication. We speak of an individual being an average student or of average intelligence. We always talk about average family size or average family income or grade point average (GPA) for students, and so on.

For discussion purposes, let us assume a variable X which stands for some scores such as the ages of students. Let the ages of 5 students be 19, 20, 22, 22 and 17 years. Then variable X would represent these ages as follows:

X: 19, 20, 22, 22, 17

Placing the Greek symbol Σ(Sigma) before X would indicate a command that all values of X are to be added together. Thus:

$\Sigma X = 19 + 20 + 22 + 22 + 17$

The mean is computed by adding all the data values and dividing it by the number of such values. The symbol used for sample average is $\overline{X}$ so that:

$$\overline{X} = \frac{19 + 20 + 22 + 22 + 17}{5}$$

In general, if there are n values in the sample, then

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

In other words,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad i = 1, \ 2 \ ... \ n$$

According to this formula, the mean can be obtained by adding up all values of X_i, where the value of i starts at 1 and ends at n with unit increments so that $i = 1, 2, 3, ... n$.

If instead of taking a sample, we take the entire population in our calculations of the mean, then the symbol for the mean of the population is μ(mu) and the size of the population is N, so that:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad i = 1, \ 2 \ ...N$$

If we have the data in grouped discrete form with frequencies, then the sample mean is given by:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

Here, Σf = Summation of all frequencies
= n

$\Sigma f(X)$ = Summation of each value of X multiplied by its corresponding frequency (f)

Example 6.5: Let us take the ages of 10 students as follows:

19, 20, 22, 22, 17, 22, 20, 23, 17, 18

Solution: This data can be arranged in a frequency distribution as follows:

(X)	(f)	$f(X)$
17	2	34
18	1	18
19	1	19
20	2	40
22	3	66
23	1	23
Total = 10		200

In this case, we have $\Sigma f = 10$ and $\Sigma f(X) = 200$, so that:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

$$= 200/10 = 20$$

Example 6.6: Calculate the mean of the marks of 46 students given in the following table.

Frequency of Marks of 46 Students

Marks (X)	Frequency (f)
9	1
10	2
11	3
12	6
13	10
14	11
15	7
16	3
17	2
18	1
Total	46

Solution: This is a discrete frequency distribution, and is calculated using the equation $\overline{x} = \frac{\Sigma f(x)}{\Sigma f}$. The following table shows the method of obtianing $\Sigma f(X)$.

Marks (X)	Frequency (f)	f(X)
9	1	9
10	2	20
11	3	33
12	6	72
13	10	130
14	11	154
15	7	105
16	3	48
17	2	34
18	1	18
	$\Sigma f = 46$	$\Sigma f(X) = 623$

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f} = \frac{623}{46} = 13.54$$

Characteristics of the mean

The arithmetic mean has three interesting properties. These are as follows:

(i) The sum of the deviations of individual values of X from the mean will always add up to zero. This means that if we subtract all the individual values from their mean, then some values will be negative and some will be positive, but if all these differences are added together then the total sum will be zero. In other words, the positive deviations must balance the negative deviations. Or symbolically:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0, i = 1, 2, \dots n$$

(ii) The second important characteristic of the mean is that it is very sensitive to extreme values. Since the computation of the mean is based upon inclusion of all values in the data, an extreme value in the data would shift the mean towards it, thus making the mean unrepresentative of the data.

(iii) The third property of the mean is that the sum of squares of the deviations about the mean is minimum. This means that if we take differences between individual values and the mean and square these differences individually and then add these squared differences, then the final figure will be less than the sum of the squared deviations around any other number other than the mean. Symbolically, it means that:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 \ = \ \text{Minimum}, \ i = 1, 2, \dots n$$

The following examples will make the concept clear about properties of mean.

(iv) The product of the arithmetic mean and the number of values on which the mean is based is equal to the sum of all given values. In other words, if we replace each item in series by the mean, then the sum of the these substitutions will equal the sum of individual items. Thus, in the figures 3, 5, 7, 9, if we substitute the mean for each item 6, 6, 6, 6 then the total is 24, both in the original series and in the substitution series.

This can be shown like this.

Since, $$\overline{X} \ = \ \frac{\Sigma X}{N}$$

$\therefore$ $$N\overline{X} \ = \ \Sigma X$$

For example, if we have a series of values 3, 5, 7, 9, the mean is 6. The squared deviations are:

X	$X - \overline{X} = X'$	X'^2
3	$3 - 6 = -3$	9
5	$5 - 6 = -1$	1
7	$7 - 6 = 1$	1
9	$9 - 6 = 3$	9
		$\Sigma X'^2 = 20$

This property provides a test to check if the computed value is the correct arithmetic mean.

Example 6.7: The mean age of a group of 100 persons (grouped in intervals 10–, 12–,..., etc.) was found to be 32.02. Later, it was discovered that age 57 was misread as 27. Find the corrected mean.

Solution: Let the mean be denoted by X. So, putting the given values in the formula of arithmetic mean, we have,

$$32.02 \ = \ \frac{\Sigma X}{100}, \text{ i.e., } \sum X = 3202$$

$$\text{Correct } \sum X \ = 3202 - 27 + 57 = 3232$$

$\therefore$ $$\text{Correct } \text{AM} = \frac{3232}{100} = 32.32$$

Example 6.8: The mean monthly salary paid to all employees in a company is ₹500. The monthly salaries paid to male and female employees average ₹520 and ₹ 420, respectively. Determine the percentage of males and females employed by the company.

Solution: Let N_1 be the number of males and N_2 be the number of females employed by the company. Also, let x_1 and x_2 be the monthly average salaries paid to male and female employees and $\bar{x}$ be the mean monthly salary paid to all the employees.

$$\bar{x} = \frac{N_1 x_1 + N_2 x_2}{N_1 + N_2}$$

or $$500 = \frac{520 N_1 + 420 N_2}{N_1 + N_2}$$ or $20 N_1 = 80 N_2$

or $$\frac{N_1}{N_2} = \frac{80}{20} = \frac{4}{1}$$

Hence, the males and females are in the ratio of 4 : 1 or 80 per cent are males and 20 per cent are females in those employed by the company.

Short-cut methods for calculating mean

We can simplify the calculations of mean by noticing that if we subtract a constant amount A from each item X to define a new variable $X' = X - A$, the mean $\bar{X}'$ of X' differs from $\bar{X}$ by A. This generally simplifies the calculations and we can then add back the constant A, termed as the *assumed mean:*

$$\bar{X} = A + \bar{X}' = A + \frac{\Sigma f(X')}{\Sigma f}$$

Table 6.6 illustrates the procedure of calculation by short-cut method using the data given in Example 6.8. The choice of A is made in such a manner as to simplify calculation the most, and is generally in the region of the concentration of data.

Table 6.6 *Short-Cut Method of Calculating Mean*

X	(f)	Deviation from Assumed Mean (13) X'	f(X')
9	1	–4	–4
10	2	–3	–6
11	3	–2	–6
12	6	–1	–6
13	10	0	–22
14	11	+1	+11
15	7	+2	+14
16	3	+3	+9
17	2	+4	+8
18	1	+5	+ 5
			+47
			–22
	$\Sigma f = 46$		$\Sigma fX' = 25$

The mean,

$$\bar{X} = A + \frac{\Sigma f(X')}{\Sigma f} = 13 + \frac{25}{46} = 13.54$$

This mean is same as calculated in Example 6.9.

In the case of grouped frequency data, the variable X is replaced by midvalue m, and in the short-cut technique; we subtract a constant value A from each m, so that the formula becomes:

$$\bar{X} = A + \frac{\Sigma f(m - A)}{\Sigma f}$$

In cases where the *class intervals are equal,* we may further simplify calculation by taking the factor i from the variable m–A defining,

$$X' = \frac{m - A}{i}$$

where i is the class width. It can be verified that when X' is defined, then, the mean of the distribution is given by:

$$\bar{X} = A + \frac{\Sigma f(X')}{\Sigma f} \times i$$

The following examples will illustrate the use of the short-cut method.

Example 6.9: The ages of twenty husbands and wives are given in the following table. Form frequency tables showing the relationship between the ages of husbands and wives with class intervals 20–24; 25–29; etc.

Calculate the arithmetic mean of the two groups after the classification.

S.No.	Age of Husband	Age of Wife
1	28	23
2	37	30
3	42	40
4	25	26
5	29	25
6	47	41
7	37	35
8	35	25
9	23	21
10	41	38
11	27	24
12	39	34
13	23	20
14	33	31
15	36	29
16	32	35
17	22	23
18	29	27
19	38	34
20	48	47

Solution:

Calculation of Arithmetic Mean of Husbands' Age

Class Intervals	Midvalues m	Husband Frequency (f_1)	$x_1' = \dfrac{m-37}{5}$	$f_1 x_1'$
20–24	22	3	–3	–9
25–29	27	5	–2	–10
30–34	32	2	–1	–2
				−21
35–39	37	6	0	0
40–44	42	2	1	2
45–49	47	2	2	4
				6
		$\Sigma f_1 = 20$		$\Sigma f_1 x_1' = -15$

Husband age, arithmetic mean:

$$\bar{x} = \frac{\Sigma f_1 x_1'}{N} \times i + A = \frac{-15}{20} \times 5 + 37 = 33.25$$

Calculation of Arithmetic Mean of Wives' Age

Class Intervals	Midvalues m	Wife Frequency (f_2)	$x_2' = \dfrac{m-37}{5}$	$f_2 x_2'$
20–24	22	5	–3	–15
25–29	27	5	–2	–10
30–34	32	4	–1	–4
35–39	37	3	0	0
40–44	42	2	1	2
45–49	47	1	2	2
		$\Sigma f_2 = 20$		$\Sigma f_2 x_2' = -25$

Wife age, arithmetic mean:

$$\bar{x} = \frac{\Sigma f_2 x_2'}{N} \times i + A = \frac{-25}{20} \times 5 + 37 = 30.75$$

Weighted arithmetic mean

In the computation of arithmetic mean we had given equal importance to each observation in the series. This equal importance may be misleading if the individual values constituting the series have different importance as in the following example:

The Raja Toy shop sells

Toy cars at	₹3 each
Toy locomotives at	₹5 each
Toy aeroplanes at	₹7 each
Toy double decker at	₹9 each

What shall be the average price of the toys sold, if the shop sells 4 toys, one of each kind?

Mean price, i.e., $\bar{x} = \dfrac{\sum x}{4} = Rs \dfrac{24}{4} = ₹6$

In this case, the importance of each observation (price quotation) is equal in as much as one toy of each variety has been sold. In the above computation of the arithmetic mean, this fact has been taken care of by including 'once only' the price of each toy.

But if the shop sells 100 toys: 50 cars, 25 locomotives, 15 aeroplanes and 10 double deckers, the importance of the four price quotations to the dealer is **not equal** as a source of earning revenue. In fact, their respective importance is equal to the number of units of each toy sold, i.e.,

The importance of toy car   50

The importance of locomotive   25

The importance of aeroplane   15

The importance of double decker   10

It may be noted that 50, 25, 15, 10 are the quantities of the various classes of toys sold. It is for these quantities that the term 'weights' is used in statistical language. Weight is represented by symbol 'w', and $\sum w$ represents the sum of weights.

While determining the 'average price of toy sold', these weights are of great importance and are taken into account in the manner illustrated as follows:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4}{w_1 + w_2 + w_3 + w_4} = \frac{\sum wx}{\sum w}$$

When w_1, w_2, w_3, w_4 are the respective weights of x_1, x_2, x_3, x_4 which in turn represent the price of four varieties of toys, viz., car, locomotive, aeroplane and double decker, respectively.

$$\bar{x} = \frac{(50 \times 3) + (25 \times 5) + (15 \times 7) + (10 \times 9)}{50 + 25 + 15 + 10}$$

$$= \frac{(150) + (125) + (105) + (90)}{100} = \frac{470}{100} = ₹4.70$$

The table below summarizes the steps taken in the computation of the weighted arithmetic mean.

Table 6.7 Weighted Arithmetic Mean of Toys Sold by the Raja Toy Shop

Toys	Price per Toy ₹x	Number Sold w	Price × Weight xw
Car	3	50	150
Locomotive	5	25	125
Aeroplane	7	15	105
Double Decker	9	10	90
		$\sum w = 100$	$\sum xw = 470$

$\sum w = 100;$ $\sum wx = 470$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{470}{100} = 4.70$$

The weighted arithmetic mean is particularly useful where we have to compute the *mean of means*. If we are given two arithmetic means, one for each of two different

series, in respect of the *same variable*, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination.

Example 6.10: The arithmetic mean of daily wages of two manufacturing concerns A Ltd. and B Ltd. is ₹5 and ₹7, respectively. Determine the average daily wages of both concerns if the number of workers employed were 2,000 and 4,000 respectively.

Solution: (*i*) Multiply each average (viz. 5 and 7), by the number of workers in the concern it represents.

(*ii*) Add up the two products obtained in (*i*) above.

(*iii*) Divide the total obtained in (*ii*) by the total number of workers.

Weighted Mean of Mean Wages of A Ltd. and B Ltd.

Manufacturing Concern	Mean Wages x	Workers Employed w	Mean Wages × Workers Employed wx
A Ltd.	5	2,000	10,000
B Ltd.	7	4,000	28,000
		$\sum w = 6,000$	$\sum wx = 38,000$

$$\bar{x} = \frac{\sum wx}{\sum w}$$

$$= \frac{38,000}{6,000}$$

$$= ₹6.33$$

The above mentioned examples explain that 'Arithmetic Means and Percentage' are not original data. They are derived figures and their importance is relative to the original data from which they are obtained. This relative importance must be taken into account by weighting while averaging them (means and percentage).

Advantages of mean

- Its concept is familiar to most people and is intuitively clear.

- Every data set has a mean, which is unique and describes the entire data to some degree. For example, when we say that the average salary of a professor is ₹ 25,000 per month, it gives us a reasonable idea about the salaries of professors.

- It is a measure that can be easily calculated.

- It includes all values of the data set in its calculation.

- Its value varies very little from sample to sample taken from the same population.

- It is useful for performing statistical procedures such as computing and comparing the means of several data sets.

Disadvantages of mean

- It is affected by extreme values, and hence, are not very reliable when the data set has extreme values especially when these extreme values are on one side of the ordered data. Thus, a mean of such data is not truly a representative of such data. For example, the average age of three persons of ages 4, 6 and 80 years gives us an average of 30.

- It is tedious to compute for a large data set as every point in the data set is to be used in computations.
- We are unable to compute the mean for a data set that has open-ended classes either at the high or at the low end of the scale.
- The mean cannot be calculated for qualitative characteristics such as beauty or intelligence, unless these can be converted into quantitative figures such as intelligence into IQs.

Geometric mean

The geometric mean (G) is the nth root of the product of n values.

$$G = \sqrt[n]{x_1 \times x_2 \times ... \times x_n}$$

The G.M. of 2, 4, 8 is the cube root of their product.

$$G = \sqrt[3]{2 \cdot 4 \cdot 8} = \sqrt[3]{64} = 4$$

If the frequencies of $x_1, x_2, ..., x_k$ are respectively $f_1, f_2, ..., f_k (\Sigma f = n)$

$$G = \sqrt[n]{x_1^{f_1} \cdot x_z^{f_2} .... x_k^{f_k}}$$

Logarithms may be used in the calculation of G.M.

$$\text{Log } G = \frac{1}{n}\left[f_1 \log x_1 + f_2 \log x_2 + ... + f_k \log x_k\right] = \frac{\Sigma f \log x}{n}$$

$$G = \text{Antilog } \frac{1}{n}\Sigma f \log x$$

If there are no frequencies, $G = (x_1 x_2 ... x_n)^{\frac{1}{n}}$ and $\log G = \frac{1}{n}\Sigma \log x$

Merits and uses of geometric mean

Most of the properties and merits of G.M. resemble those of A.M.

- The GM takes into account all the items in the data and condenses them into one respectative value.
- It has a downward bias. It gives more weight to smaller values than to larger values.
- It is determinate. For the same data there cannot be two geometric means.
- It balances the ratios of the values on either side of the data. It is ideally suited to average rates of change such as index numbers and ratios between measures and percentages.
- It is amenable to algebraic manipulations like the A.M.

Demerits of geometric mean

- It is difficult to use and to compute.
- It is determined for positive values and cannot be used for negative values of zero. A zero will convert the whole product into zero.

6.5.2 Median

The second measure of central tendency that has a wide usage in statistical works is the median. Median is that *value* of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it. Half the total number of observations lie below the median, and half above it. The median is thus a positional average.

The median of ungrouped data is found easily if the items are first arranged in order of the magnitude. The median may then be located simply by counting, and its value can be obtained by reading the value of the middle observations. If we have five observations whose values are 8, 10, 1, 3 and 5, the values are first arrayed: 1, 3, 5, 8 and 10. It is now apparent that the value of the median is 5, since two observations are below that value and two observations are above it. When there is an even number of cases, there is no actual middle item and the median is taken to be the average of the values of the items lying on either side of $(N+1)/2$, where N is the total number of items. Thus, if the values of six items of a series are 1, 2, 3, 5, 8 and 10, then the median is the value of item number $(6+1)/2 = 3.5$, which is approximated as the average of the third and the fourth items, i.e., $(3+5)/2 = 4$.

Thus, the steps required for obtaining median are:

1. Arrange the data as an array of increasing magnitude.

2. Obtain the value of the $(N+1)/2$th item.

Even in the case of grouped data, the procedure for obtaining median is straightforward as long as the variable is discrete or non-continuous as is clear from the following example.

Example 6.11: Obtain the median size of shoes sold from the following data.

Number of Shoes Sold by Size in One Year

Size	Number of Pairs	Cumulative Total
5	30	30
$5\frac{1}{2}$	40	70
6	50	120
$6\frac{1}{2}$	150	270
7	300	570
$7\frac{1}{2}$	600	1170
8	950	2120
$8\frac{1}{2}$	820	2940
9	750	3690
$9\frac{1}{2}$	440	4130
10	250	4380
$10\frac{1}{2}$	150	4530
11	40	4570
$11\frac{1}{2}$	39	4609
	Total	4609

Solution: Median, is the value of $\dfrac{(N+1)}{2}$th $= \dfrac{4609+1}{2}$th $= 2305$th item. Since the items are already arranged in ascending order (size-wise), the size of 2305th item is easily determined by constructing the cumulative frequency. Thus, the median size of shoes sold is 8½, the size of 2305th item.

In the case of grouped data with continuous variable, the determination of median is a bit more involved. Consider the following table where the data relating to the distribution of male workers by average monthly earnings is given. Clearly the median of 6291 is the earnings of $(6291+1)/2 = 3146$th worker arranged in ascending order of earnings.

From the cumulative frequency, it is clear that this worker has his income in the class interval 67.5–72.5. But, it is impossible to determine his exact income. We therefore, resort to approximation by assuming that the 795 workers of this class are distributed *uniformly* across the interval 67.5 to 72.5. The median worker is $(3146-2713) = 433$rd of these 795, and hence, the value corresponding to him can be approximated as,

$$67.5 + \frac{433}{795} \times (72.5 - 67.5) = 67.5 + 2.73 = 70.23$$

Distribution of Male Workers by Average Monthly Earnings

Group No.	Monthly Earnings (₹)	No. of Workers	Cumulative No. of Workers
1	27.5–32.5	120	120
2	32.5–37.5	152	272
3	37.5–42.5	170	442
4	42.5–47.5	214	656
5	47.5–52.5	410	1066
6	52.5–57.5	429	1495
7	57.5–62.5	568	2063
8	62.5–67.5	650	2713
9	67.5–72.5	795	3508
10	72.5–77.5	915	4423
11	77.5–82.5	745	5168
12	82.5–87.5	530	5698
13	87.5–92.5	259	5957
14	92.5–97.5	152	6109
15	97.5–102.5	107	6216
16	102.5–107.5	50	6266
17	107.5–112.5	25	6291
		Total	6291

The value of the median can thus be put in the form of the formula,

$$Me = l + \frac{\frac{N+1}{2} - C}{f} \times i$$

Where l is the lower limit of the median class, i its width, f its frequency, C the cumulative frequency upto (but not including) the median class, and N is the total number of cases.

Finding median by graphical analysis

The median can quite conveniently be determined by reference to the ogive which plots the cumulative frequency against the variable. The value of the item below which half the items lie, can easily be read from the ogive.

Example 6.12: Obtain the median of data given in the following table.

Monthly Earnings	Frequency	Less Than	More Than
27.5	—	0	6291
32.5	120	120	6171
37.5	152	272	6019
42.5	170	442	5849
47.5	214	656	5635
52.5	410	1066	5225
57.5	429	1495	4796
62.5	568	2063	4228
67.5	650	2713	3578
72.5	795	3508	2783
77.5	915	4423	1868
82.5	745	5168	1123
87.5	530	5698	593
92.5	259	5957	334
97.5	152	6109	182
102.5	107	6216	75
107.5	50	6266	25
112.5	25	6291	0

Solution: It is clear that this is grouped data. The first class is 27.5–32.5, whose frequency is 120, and the last class is 107.5–112.5, whose frequency is 25. Figure 6.7 shows the ogive of less than cumulative frequency. The median is the value below which $N/2$ items lie, is $6291/2 = 3145.5$ items lie, which is read of from Figure 6.8 as about 70. More accuracy than this is unobtainable because of the space limitation on the earning scale.
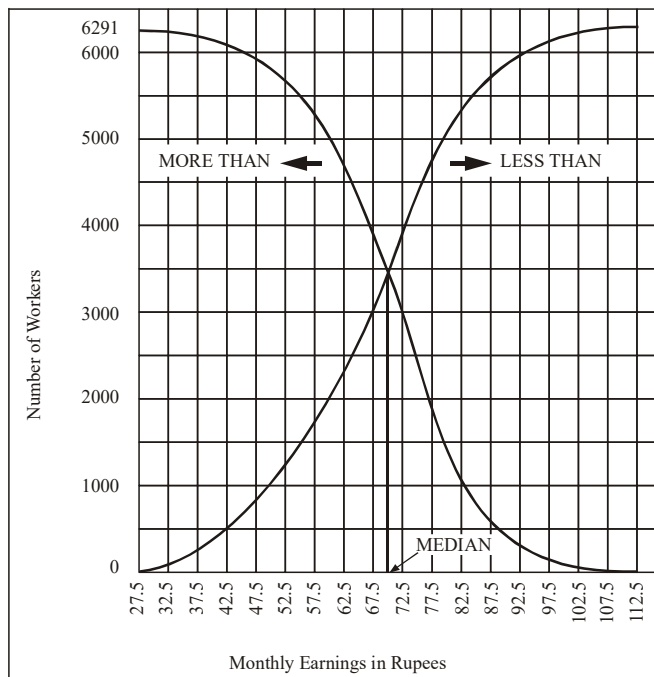
Fig. 6.7 *Median Determination by Plotting Less than and More than Cumulative Frequency*

The median can also be determined by plotting both 'less than' and 'more than' cumulative frequency as shown in Figure 6.7. It should be obvious that the two curves should intersect at the median of the data.
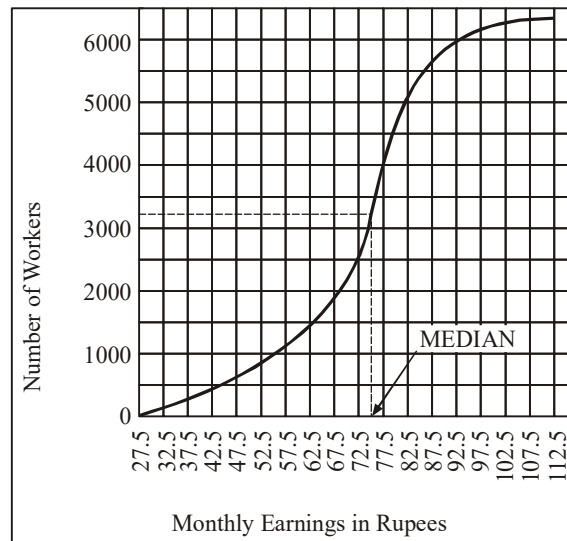
***Fig. 6.8** Median*

Advantages of median

- Median is a positional average and hence the extreme values in the data set do not affect it as much as they do to the mean.
- Median is easy to understand and can be calculated from any kind of data, even from grouped data with open-ended classes.
- We can find the median even when our data set is qualitative and can be arranged in the ascending or the descending order, such as average beauty or average intelligence.
- Similar to mean, median is also unique, meaning that, there is only one median in a given set of data.
- Median can be located visually when the data is in the form of ordered data.
- The sum of absolute differences of all values in the data set from the median value is minimum. This means that, it is less than any other value of central tendency in the data set, which makes it more central in certain situations.

Disadvantages of median

- The data must be arranged in order to find the median. This can be very time consuming for a large number of elements in the data set.
- The value of the median is affected more by sampling variations. Different samples from the same population may give significantly different values of the median.
- The calculation of median in case of grouped data is based on the assumption that the values of observations are evenly spaced over the entire class interval and this is usually not so.
- Median is comparatively less stable than mean, particularly for small samples, due to fluctuations in sampling.
- Median is not suitable for further mathematical treatment. For example, we cannot compute the median of the combined group from the median values of different groups.

6.5.3 Mode

The mode is that value of the variable which occurs or repeats itself the greatest number of times. The mode is the most 'fashionable' size in the sense that it is the most common and typical, and is defined by Zizek as 'the value occurring most frequently in a series (or group of items) and around which the other items are distributed most densely'.

The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It is the most frequent or the most common value, provided that a sufficiently large number of items are available, to give a smooth distribution. It will correspond to the value of the maximum point (ordinate), of a frequency distribution if it is an 'ideal' or smooth distribution. It may be regarded as the most typical of a series of values. The modal wage, for example, is the wage received by more individuals than any other wage. The modal 'hat' size is that, which is worn by more persons than any other single size.

It may be noted that the occurrence of one or a few extremely high or low values has no effect upon the mode. If a series of data are unclassified, not have been either arrayed or put into a frequency distribution, the mode cannot be readily located.

Taking first an extremely simple example, if seven men are receiving daily wages of ₹5, 6, 7, 7, 7, 8 and 10, it is clear that the modal wage is ₹7 per day. If we have a series such as 2, 3, 5, 6, 7, 10 and 11, it is apparent that there is no mode.

There are several methods of estimating the value of the mode. But, it is seldom that the different methods of ascertaining the mode give us identical results. Consequently, it becomes necessary to decide as to which method would be most suitable for the purpose in hand. In order that a choice of the method may be made, we should understand each of the methods and the differences that exist among them.

The four important methods of estimating mode of a series are: (*i*) Locating the most frequently repeated value in the array; (*ii*) Estimating the mode by interpolation; (*iii*) Locating the mode by graphic method; and (*iv*) Estimating the mode from the mean and the median. Only the last three methods are discussed in this unit.

Estimating the mode by interpolation

In the case of continuous frequency distributions, the problem of determining the value of the mode is not so simple as it might have appeared from the foregoing description. Having located the modal class of the data, the next problem in the case of continuous series is to interpolate the value of the mode within this 'modal' class.

The interpolation is made by the use of any one of the following formulae:

$$(i)\ Mo = l_1 + \frac{f_2}{f_0 + f_2} \times i; \quad (ii)\ Mo = l_2 - \frac{f_0}{f_0 + f_2} \times i$$

$$(iii)\ Mo = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times i$$

Where l_1 is the lower limit of the modal class, l_2 is the upper limit of the modal class, f_0 equals the frequency of the preceding class in value, f_1 equals the frequency of the modal class in value, f_2 equals the frequency of the following class (class next to modal class) in value, and i equals the interval of the modal class.

Example 6.13: Determine the mode for the data given in the following table.

Wage Group	Frequency (f)
14 — 18	6
18 — 22	18
22 — 26	19
26 — 30	12
30 — 34	5
34 — 38	4
38 — 42	3
42 — 46	2
46 — 50	1
50 — 54	0
54 — 58	1

Solution: In the given data, 22 – 26 is the modal class since it has the largest frequency. The lower limit of the modal class is 22, its upper limit is 26, its frequency is 19, the frequency of the preceding class is 18, and of the following class is 12. The class interval is 4. Using the various methods of determining mode, we have,

(i) $Mo = 22 + \dfrac{12}{18+12} \times 4$ (ii) $Mo = 26 - \dfrac{18}{18+12} \times 4$

$= 22 + \dfrac{8}{5}$ $= 26 - \dfrac{12}{5}$

$= 23.6$ $= 23.6$

(iii) $Mo = 22 + \dfrac{19-18}{(19-18)+(19-12)} \times 4 = 22 + \dfrac{4}{8} = 22.5$

In formulae (i) and (ii), the frequency of the classes adjoining the modal class is used to pull the estimate of the mode away from the midpoint towards either the upper or lower class limit. In this particular case, the frequency of the class preceding the modal class is more than the frequency of the class following and therefore, the estimated mode is less than the midvalue of the modal class. This seems quite logical. If the frequencies are more on one side of the modal class than on the other it can be reasonably concluded that the items in the modal class are concentrated more towards the class limit of the adjoining class with the larger frequency.

The formula (iii) is also based on a logic similar to that of (i) and (ii). In this case, to interpolate the value of the mode within the modal class, the differences between the frequency of the modal class, and the respective frequencies of the classes adjoining it are used. This formula usually gives results better than the values obtained by the other and exactly equal to the results obtained by graphic method. The formulae (i) and (ii) give values which are different from the value obtained by formula (iii) and are more close to the central point of modal class. If the frequencies of the class adjoining the modal are equal, the mode is expected to be located at the midvalue of the modal class, but if the frequency on one of the sides is greater, the mode will be pulled away from the central point. It will be pulled more and more if the difference between the frequencies of the classes adjoining the modal class is higher and higher. In Example 6.13, the frequency of the modal class is 19 and that of preceding class is 18. So, the mode should be quite close to the lower limit of the modal class. The midpoint of the modal class is 24 and lower limit of the modal class is 22.

Locating the mode by the graphic method

The upper corners of the rectangle over the modal class have been joined by straight lines to those of the adjoining rectangles as shown in the diagram; the right corner to the corresponding one of the adjoining rectangle on the left, etc. If a perpendicular is drawn from the point of intersection of these lines, we have a value for the mode indicated on the base line. The graphic approach is, in principle, similar to the arithmetic interpolation explained earlier.

The mode may also be determined graphically from an ogive or cumulative frequency curve. It is found by drawing a perpendicular to the base from that point on the curve where the curve is most nearly vertical, i.e., steepest (in other words, where it passes through the greatest distance vertically and smallest distance horizontal). The point where it cuts the base gives us the value of the mode. How accurately this method determines the mode is governed by: (*i*) The shape of the ogive, (*ii*) The scale on which the curve is drawn.

Estimating the mode from the mean and the median

There usually exists a relationship among the mean, median and mode for moderately asymmetrical distributions. If the distribution is symmetrical, the mean, median and mode will have identical values, but if the distribution is skewed (moderately) the mean, median and mode will pull apart. If the distribution tails off towards higher values, the mean and the median will be greater than the mode. If it tails off towards lower values, the mode will be greater than either of the other two measures. In either case, the median will be about one-third as far away from the mean as the mode is. This means that,

$$\text{Mode} = \text{Mean} - 3\ (\text{Mean} - \text{Median})$$
$$= 3\ \text{Median} - 2\ \text{Mean}$$
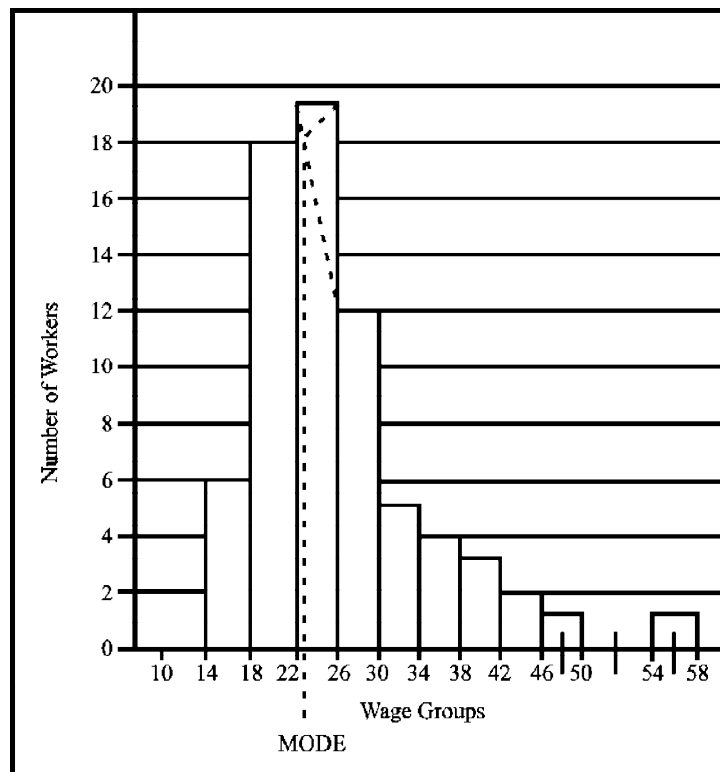
Fig. 6.9 *Method of Mode Determination by Graphic Interpolation*

In the case of the average monthly earnings, the mean is 68.53 and the median is 70.2. If these values are substituted in the above formula, we get,

$$\text{Mode} = 68.5 - 3(68.5 - 70.2)$$
$$= 68.5 + 5.1 = 73.6$$

According to the formula used earlier,

$$\text{Mode} = l_1 + \frac{f_2}{f_0 + f_2} \times i$$
$$= 72.5 + \frac{745}{795 + 745} \times 5$$
$$= 72.5 + 2.4 = 74.9$$

OR

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$
$$= 72.5 + \frac{915 - 795}{2 \times 915 - 795 - 745} \times 5$$
$$= 72.5 + \frac{120}{290} \times 5 = 74.57$$

The difference between the two estimates is due to the fact that the assumption of relationship between the mean, median and mode may not always be true which is obviously not valid in this case.

Example 6.14: (*i*) In a moderately symmetrical distribution, the mode and mean are 32.1 and 35.4 respectively. Calculate the median.

(*ii*) If the mode and median of moderately asymmetrical series are respectively 16" and 15.7", what would be its most probable median?

(*iii*) In a moderately skewed distribution, the mean and the median are respectively 25.6 and 26.1 inches. What is the mode of the distribution?

Solution: (*i*) We know,

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

or $\qquad$ 3 Median $= \text{Mode} + 2\,\text{Mean}$

or $\qquad \text{Median} = \dfrac{32.1 + 2 \times 35.4}{3}$
$$= \frac{102.9}{3}$$
$$= 34.3$$

(*ii*) $\qquad$ 2 Mean $= 3\,\text{Median} - \text{Mode}$

or $\qquad \text{Mean} = \dfrac{1}{2}(3 \times 15.7 - 16.0) = \dfrac{31.1}{2} = 15.55$

(*iii*) $\qquad \text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$
$$= 3 \times 26.1 - 2 \times 25.6 = 78.3 - 51.2 = 27.1$$

Advantages of mode

- Similar to median, the mode is not affected by extreme values in the data.
- Its value can be obtained in open-ended distributions without ascertaining the class limits.

- It can be easily used to describe qualitative phenomenon. For example, if most people prefer a certain brand of tea, then this will become the modal point.
- Mode is easy to calculate and understand. In some cases, it can be located simply by observation or inspection.

Disadvantages of mode

- Quite often, there is no modal value.
- It can be bi-modal or multi-modal, or it can have all modal values making its significance more difficult to measure.
- If there is more than one modal value, the data is difficult to interpret.
- A mode is not suitable for algebraic manipulations.
- Since the mode is the value of maximum frequency in the data set, it cannot be rigidly defined if such frequency occurs at the beginning or at the end of the distribution.
- It does not include all observations in the data set, and hence, less reliable in most of the situations.

6.5.4 Quartiles, Deciles and Percentiles

Some measures other than measures of central tendency are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. These are positional measures and are called quantiles and consist of quartiles, deciles and percentiles. The quartiles divide the data into four equal parts. The deciles divide the total ordered data into ten equal parts and percentiles divide the data into 100 equal parts. Consequently, there are three quartiles, nine deciles and 99 percentiles. The quartiles are denoted by the symbol Q so that Q_1 will be such point in the ordered data which has 25 per cent of the data below and 75 per cent of the data above it. In other words Q_1 is the value corresponding to $\left(\dfrac{n+1}{4}\right)$ th ordered observation. Similarly, Q_2 divides the data in the middle, and is also equal to the median and its value Q_2 is given by:

$$Q_2 = \text{The value of } 2\left(\frac{n+1}{4}\right)\text{th ordered observation in the data.}$$

Similarly, we can calculate the values of various deciles. For instance,

$$D_1 = \left(\frac{n+1}{10}\right)\text{th observaton in the data, and}$$

$$D_7 = 7\left(\frac{n+1}{10}\right)\text{th observation in the ordered data.}$$

Percentiles are generally used in the research area of education where people are given standard tests and it is desirable to compare the relative position of the subject's performance on the test. Percentiles are similarly calculated as:

$$P_7 = 7\left(\frac{n+1}{100}\right)\text{th observation in the ordered data.}$$

and,

$$P_{69} = 69\left(\frac{n+1}{100}\right)\text{th observation in the ordered data.}$$

Quartiles

The formula for calculating the values of quartiles for grouped data is given as follows.

$$Q = L + (j/f)C$$

where,

Q = The quartile under consideration.

L = Lower limit of the class interval which contains the value of Q.

j = The number of units we lack from the class interval which contains the value of Q, in reaching the value of Q.

f = Frequency of the class interval containing Q.

C = Size of the class interval.

Let us assume we took the data of the ages of 100 students and a frequency distribution for this data has been constructed as shown.

The frequency distribution is as follows:

Ages (CI)	Mid-point (X)	(f)	f(X)	f(X)²
16 and upto 17	16.5	4	66	1089.0
17 and upto 18	17.5	14	245	4287.5
18 and upto 19	18.5	18	333	6160.5
19 and upto 20	19.5	28	546	10647.0
20 and upto 21	20.5	20	410	8405.0
21 and upto 22	21.5	12	258	5547.0
22 and upto 23	22.5	4	90	2025.0
	Totals = 100		1948	38161

In our case, in order to find Q_1, where Q_1 is the cut off point so that 25 per cent of the data is below this point and 75 per cent of the data is above, we see that the first group has 4 students and the second group has 14 students making a total of 18 students. Since Q_1 cuts off at 25 students, it is the third class interval which contains Q_1. This means that the value of L in our formula is 18.

Since we already have 18 students in the first two groups, we need 7 more students from the third group to make it a total of 25 students, which is the value of Q_1. Hence, the value of (j) is 7. Also, since the frequency of this third class interval which contains Q_1 is 18, the value of (f) in our formula is 18. The size of the class interval C is given as 1. Substituting these values in the formula for Q, we get

$$Q_1 = 18 + (7/18)1$$
$$= 18 + .38 = 18.38$$

This means that 25 per cent of the students are below 18.38 years of age and 75 per cent are above this age.

Similarly, we can calculate the value of Q_2, using the same formula. Hence,

$$Q_2 = L + (j/f)C$$
$$= 19 + (14/28)1 = 19.5$$

This also happens to be the median.

By using the same formula and same logic we can calculate the values of all deciles as well as percentiles.

We have defined the median as the value of the item which is located at the centre of the array. We can define other measures which are located at other specified points. Thus, the Nth percentile of an array is the value of the item such that N per cent items lie below it. Clearly then the N_{th} percentile P_n of grouped data is given by

$$P_n = l + \frac{\frac{nN}{100} - C}{f} \times i$$

where l is the lower limit of the class in which $nN/100$th item lies, i its width, f its frequency, C the cumulative frequency upto (but not including) this class, and N is the total number of items.

We similarly define the Nth decile as the value of the item below which $(nN/10)$ items of the array lie. Clearly,

$$D_n = P_{10n} = l + \frac{\frac{nN}{10} - C}{f} \times i$$

where the symbols have the obvious meanings.

The other most commonly referred to measures of location are the quartiles. Thus, nth quartile is the value of the item which lies at the $n(N/5)$th item. Clearly Q_2, the second quartile is the median, for grouped data.

$$Q_n = P_{25n} = l + \frac{\frac{nN}{4} - C}{f} \times i$$

6.5.5 Measures of Dispersion

A measure of dispersion, or simply dispersion may be defined as statistics signifying the extent of the scatteredness of items around a measure of central tendency.

A measure of dispersion may be expressed in an 'absolute form', or in a 'relative form'. It is said to be in an absolute form when it states the actual amount by which the value of an item on an average deviates from a measure of central tendency. Absolute measures are expressed in concrete units, i.e., units in terms of which the data have been expressed, e.g., rupees, centimetres, kilograms, etc., and are used to describe frequency distribution.

A relative measure of dispersion computed is a quotient by dividing the absolute measures by a quantity in respect to which absolute deviation has been computed. It is as such a pure number and is usually expressed in a percentage form. Relative measures are used for making comparisons between two or more distributions.

A measure of dispersion should possess all those characteristics which are considered essential for a measure of central tendency, viz.

- It should be based on all observations.
- It should be readily comprehensible.
- It should be fairly easily calculated.
- It should be affected as little as possible by fluctuations of sampling.
- It should be amenable to algebraic treatment.

The following are some common measures of dispersion:

(*i*) The range, (*ii*) the semi-interquartile range or the quartile deviation, (*iii*) the mean deviation, and (*iv*) the standard deviation. Of these, the standard deviation is the best measure. We describe these measures in the following sections.

Range

The crudest measure of dispersion is the range of the distribution. The range of any series is the difference between the highest and the lowest values in the series. If the marks received in an examination taken by 248 students are arranged in ascending order, then the range will be equal to the difference between the highest and the lowest marks.

In a frequency distribution, the range is taken to be the difference between the lower limit of the class at the lower extreme of the distribution and the upper limit of the class at the upper extreme.

Table 6.8 *Weekly Earnings of Labourers in Four Workshops of the Same Type*

Weekly earnings ₹	No. of workers			
	Workshop A	*Workshop B*	*Workshop C*	*Workshop D*
15–16	...	...	2	...
17–18	...	2	4	...
19–20	...	4	4	4
21–22	10	10	10	14
23–24	22	14	16	16
25–26	20	18	14	16
27–28	14	16	12	12
29–30	14	10	6	12
31–32	...	6	6	4
33–34	...	...	2	2
35–36	...	...	...	...
37–38	...	...	4	...
Total	80	80	80	80
Mean	25.5	25.5	25.5	25.5

Consider the data on weekly earning of worker on four workshops given in the Table 6.8. We note the following:

Workshop	Range
A	9
B	15
C	23
D	15

From these figures, it is clear that the greater the range, the greater is the variation of the values in the group.

The range is a measure of absolute dispersion and as such cannot be usefully employed for comparing the variability of two distributions expressed in different units. The amount of dispersion measured, say, in pounds, is not comparable with dispersion measured in inches. So the need of measuring relative dispersion arises.

An absolute measure can be converted into a relative measure if we divide it by some other value regarded as standard for the purpose. We may use the mean of the distribution or any other positional average as the standard.

For Table 6.8, the relative dispersion would be:

$$\text{Workshop } A = \frac{9}{25.5} \qquad \text{Workshop } C = \frac{23}{25.5}$$

$$\text{Workshop } B = \frac{15}{25.5} \qquad \text{Workshop } D = \frac{15}{25.5}$$

An alternate method of converting an absolute variation into a relative one would be to use the total of the extremes as the standard. This will be equal to dividing the difference of the extreme items by the total of the extreme items. Thus,

$$\text{Relative Dispersion} = \frac{\text{Difference of extreme items, i.e, Range}}{\text{Sum of extreme items}}$$

The relative dispersion of the series is called the coefficient or ratio of dispersion. In our example of weekly earnings of workers considered earlier, the coefficients would be:

$$\text{Workshop } A = \frac{9}{21+30} = \frac{9}{51} \qquad \text{Workshop } B = \frac{15}{17+32} = \frac{15}{49}$$

$$\text{Workshop } C = \frac{23}{15+38} = \frac{23}{53} \qquad \text{Workshop } D = \frac{15}{19+34} = \frac{15}{53}$$

Merits and limitations of range

Merits

Of the various characteristics that a good measure of dispersion should possess, the range has only two, viz (i) it is easy to understand, and (ii) its computation is simple.

Limitations

Besides the aforesaid two qualities, the range does not satisfy the other test of a good measure and hence it is often termed as a crude measure of dispersion.

The following are the limitations that are inherent in the range as a concept of variability:

(i) Since it is based upon two extreme cases in the entire distribution, the range may be considerably changed if either of the extreme cases happens to drop out, while the removal of any other case would not affect it at all.

(ii) It does not tell anything about the distribution of values in the series relative to a measure of central tendency.

(iii) It cannot be computed when distribution has open-end classes.

(iv) It does not take into account the entire data. These can be illustrated by the following illustration. Consider the data given in Table 6.9.

Table 6.9 *Distribution with the Same Number of Cases,
but Different Variability*

Class	No. of students		
	Section A	Section B	Section C
0–10	...	...	...
10–20	1	...	...
20–30	12	12	19
30–40	17	20	18
40–50	29	35	16
50–60	18	25	18
60–70	16	10	18
70–80	6	8	21
80–90	11	...	...
90–100	...	...	...
Total	110	110	110
Range	80	60	60

The table is designed to illustrate three distributions with the same number of cases but different variability. The removal of two extreme students from section *A* would make its range equal to that of *B* or *C*.

The greater range of *A* is not a description of the entire group of 110 students, but of the two most extreme students only. Further, though sections *B* and *C* have the same range, the students in section B cluster more closely around the central tendency of the group than they do in section *C*. Thus, the range fails to reveal the greater homogeneity of *B* or the greater dispersion of *C*. Due to this defect, it is seldom used as a measure of dispersion.

Specific uses of range

In spite of the numerous limitations of the range as a measure of dispersion, there are the following circumstances when it is the most appropriate one:

(i) In situations where the extremes involve some hazard for which preparation should be made, it may be more important to know the most extreme cases to be encountered than to know anything else about the distribution. For example, an explorer, would like to know the lowest and the highest temperatures on record in the region he is about to enter; or an engineer would like to know the maximum rainfall during 24 hours for the construction of a storem water drain.

(ii) In the study of prices of securities, range has a special field of activity. Thus to highlight fluctuations in the prices of shares or bullion it is a common practice to indicate the range over which the prices have moved during a certain period of time. This information, besides being of use to the operators, gives an indication of the stability of the bullion market, or that of the investment climate.

(iii) In statistical quality control the range is used as a measure of variation. We, e.g., determine the range over which variations in quality are due to random causes, which is made the basis for the fixation of control limits.

Quartile Deviation

Another measure of dispersion, much better than the range, is the semi-interquartile range, usually termed as 'quartile deviation'. As stated in the previous unit, quartiles are the points which divide the array in four equal parts. More precisely, Q_1 gives the value of the item 1/4th the way up the distribution and Q_3 the value of the item 3/4th the way up the distribution. Between Q_1 and Q_3 are included half the total number of items. The difference between Q_1 and Q_3 includes only the central items but excludes the extremes. Since under most circumstances, the central half of the series tends to be fairly typical of all the items, the interquartile range $(Q_3 - Q_1)$ affords a convenient and often a good indicator of the absolute variability. The larger the interquartile range, the larger the variability.

Usually, one-half of the difference between Q_3 and Q_1 is used and to it is given the name of quartile deviation or semi-interquartile range. The interquartile range is divided by two for the reason that half of the interquartile range will, in a normal distribution, be equal to the difference between the median and any quartile. This means that 50 per cent items of a normal distribution will lie within the interval defined by the median plus and minus the semi-interquartile range.

Symbolically:

$$Q.D. = \frac{Q_3 - Q_1}{2} \qquad \qquad ...(6.1)$$

Let us find quartile deviations for the weekly earnings of labour in the four workshop whose data is given in Table 6.8. The computations are as shown in Table 4.5.

As shown in the table, Q.D. of workshop A is ₹2.12 and median value in 25.3. This means that if the distribution is symmetrical the number of workers whose wages vary between (25.3–2.1) = ₹23.2 and (25.3 + 2.1) = ₹27.4, shall be just half of the total cases. The other half of the workers will be more than ₹2.1 removed from the median wage. As this distribution is not symmetrical, the distance between Q_1 and the median Q_2 is not the same as between Q_3 and the median. Hence the interval defined by median plus and minus semi inter-quartile range will not be exactly the same as given by the value of the two quartiles. Under such conditions the range between ₹23.2 and ₹27.4 will not include precisely 50 per cent of the workers.

If quartile deviation is to be used for comparing the variability of any two series, it is necessary to convert the absolute measure to a coefficient of quartile deviation. To do this the absolute measure is divided by the average size of the two quartile.

Symbolically:

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad \qquad ...(6.2)$$

Applying this to our illustration of four workshops, the coefficients of Q.D. are as given below.

Table 6.10 *Calculation of Quartile Deviation*

		Workshop A	*Workshop B*	*Workshop C*	*Workshop D*
Location of Q_2	$\frac{N}{2}$	$\frac{80}{2} = 40$	$\frac{80}{2} = 40$	$\frac{80}{2} = 40$	$\frac{80}{2} = 40$
	Q_2	$24.5 + \frac{40-30}{22} \times 2$	$24.5 + \frac{40-30}{18} \times 2$	$24.5 + \frac{40-30}{16} \times 2$	$24.5 + \frac{40-30}{16} \times 2$
		$= 24.5 + 0.9$	$= 24.5 + 1.1$	$= 24.5 + 0.75$	$= 24.5 + 0.75$
		$= 25.4$	$= 25.61$	$= 25.25$	$= 25.25$

Location of Q_1	$\dfrac{N}{4}$	$\dfrac{80}{4} = 20$	$\dfrac{80}{4} = 20$	$\dfrac{80}{4} = 20$	$\dfrac{80}{4} = 20$
	Q_1	$22.5 + \dfrac{20-10}{22} \times 2$ $= 22.5 + .91$ $= 23.41$	$22.5 + \dfrac{20-16}{14} \times 2$ $= 22.5 + .57$ $= 23.07$	$20.5 + \dfrac{20-10}{10} \times 2$ $= 20.5 + 2$ $= 22.5$	$22.5 + \dfrac{20-18}{16} \times 2$ $= 22.5 + .25$ $= 22.75$
Location of Q_3	$\dfrac{3N}{4}$	$3 \times \dfrac{80}{4} = 60$	60	60	60
	Q_3	$26.5 + \dfrac{60-52}{14} \times 2$ $= 26.5 + 1.14$ $= 27.64$	$26.5 + \dfrac{60-48}{16} \times 2$ $= 26.5 + 1.5$ $= 28.0$	$26.5 + \dfrac{60-50}{12} \times 2$ $= 26.5 + 1.67$ $= 28.17$	$26.5 + \dfrac{60-50}{12} \times 2$ $= 26.5 + 1.67$ $= 28.17$
Quartile Deviation	$\dfrac{Q_3 - Q_1}{2}$	$\dfrac{27.64 - 23.41}{2}$ $= \dfrac{4.23}{2} = ₹2.12$	$\dfrac{28 - 23.07}{2}$ $= \dfrac{4.93}{2} = ₹2.46$	$\dfrac{28.17 - 22.5}{2}$ $= \dfrac{5.67}{2} = ₹2.83$	$\dfrac{28.17 - 22.75}{2}$ $= \dfrac{5.42}{2} = ₹2.71$
Coefficient of quartile deviation $\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$	$= \dfrac{27.64 - 23.41}{27.64 + 23.41}$ $= 0.083$		$\dfrac{28 - 23.07}{28 + 23.07}$ $= 0.097$	$\dfrac{28.17 - 22.5}{28.17 + 22.5}$ $= 0.112$	$\dfrac{28.17 - 22.75}{28.17 + 22.75}$ $= 0.106$

Characteristics of quartile deviation

- The size of the quartile deviation gives an indication about the uniformity or otherwise of the size of the items of a distribution. If the quartile deviation is small it denotes large uniformity. Thus, a coefficient of quartile deviation may be used for comparing uniformity or variation in different distributions.

- Quartile deviation is not a measure of dispersion in the sense that it does not show the scatter around an average, but only a distance on scale. Consequently, quartile deviation is regarded as a measure of partition.

- It can be computed when the distribution has open-end classes.

Limitations of quartile deviation

Except for the fact that its computation is simple and it is easy to understand, a quartile deviation does not satisfy any other test of a good measure of variation.

Mean Deviation

A weakness of the measures of dispersion discussed earlier, based upon the range or a portion thereof, is that the precise size of most of the variants has no effect on the result. As an illustration, the quartile deviation will be the same whether the variates between Q_1 and Q_3 are concentrated just above Q_1 or they are spread uniformly from Q_1 to Q_3. This is an important defect from the viewpoint of measuring the divergence of the distribution from its typical value. The mean deviation is employed to answer the objection.

Mean deviation also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.

We can measure the deviations from any measure of central tendency, but the most commonly employed ones are the median and the mean. The median is preferred because it has the important property that the average deviation from it is the least.

Calculation of the mean deviation then involves the following steps:

(a) Calculate the median (or the mean) Me (or $\overline{X}$).

(b) Record the deviations $|d| = |x - Me|$ of each of the items, ignoring the sign.

(c) Find the average value of deviations.

$$\text{Mean Deviation} = \frac{\Sigma |d|}{N} \qquad \qquad \text{...(6.3)}$$

Example 6.15: Calculate the mean deviation from the following data giving marks obtained by 11 students in a class test.

14, 15, 23, 20, 10, 30, 19, 18, 16, 25, 12.

Solution: $\text{Median} = \text{Size of } \dfrac{11+1}{2} \text{ th item}$

$= \text{size of 6th item} = 18.$

Serial No.	Marks	$\lvert x - Median \rvert$ $\lvert d \rvert$
1	10	8
2	12	6
3	14	4
4	15	3
5	16	2
6	18	0
7	19	1
8	20	2
9	23	5
10	25	7
11	30	12
		$\Sigma \lvert d \rvert = 50$

$$\text{Mean deviation from median} \quad = \frac{\Sigma |d|}{N}$$

$$= \frac{50}{11} = 4.5 \text{ marks.}$$

For grouped data, it is easy to see that the mean deviation is given by

$$\text{Mean deviation, M.D.} = \frac{\Sigma f |d|}{\Sigma f} \qquad \qquad \text{...(6.4)}$$

where $|d| = |x - \text{median}|$ for grouped discrete data, and $|d| = M - \text{median}|$ for grouped continuous data with M as the mid-value of a particular group. The following examples illustrate the use of this formula.

Example 6.16: Calculate the mean deviation from the following data

Size of item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

Solution:

Size	Frequency f	Cumulative frequency	Deviations from median (9) $|d|$	$f|d|$
6	3	3	3	9
7	6	9	2	12
8	9	18	1	9
9	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
	48			60

Median = the size of $\dfrac{48+1}{2}$ = 24.5th item which is 9.

Therefore, deviations d are calculated from 9, i.e., $|d| = |x - 9|$.

Mean deviation $= \dfrac{\Sigma f |d|}{\Sigma f} = \dfrac{60}{48} = 1.25$

Example 6.17: Calculate the mean deviation from the following data:

x	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
f	18	16	15	12	10	5	2	2

Solution:

This is a frequency distribution with continuous variable. Thus, deviations are calculated from mid-values.

x	Mid-value	f	Less than c.f.	Deviation from median $|d|$	$f|d|$
0–10	5	18	18	19	342
10–20	15	16	34	9	144
20–30	25	15	49	1	15
30–40	35	12	61	11	132
40–50	45	10	71	21	210
50–60	55	5	76	31	155
60–70	65	2	78	41	82
70–80	75	2	80	51	102
		80			1182

$$\text{Median} = \text{the size of } \frac{80}{2} \text{ th item}$$

$$= 20 + \frac{6}{15} \times 10 = 24$$

and then, mean deviation $\quad = \dfrac{\Sigma f |d|}{\Sigma f}$

$$= \frac{1182}{80} = 14.775.$$

Merits and demerits of the mean deviation

Merits

- It is easy to understand.
- As compared to standard deviation (discussed later), its computation is simple.
- As compared to standard deviation, it is less affected by extreme values.
- Since it is based on all values in the distribution, it is better than range or quartile deviation.

Demerits

- It lacks those algebraic properties which would facilitate its computation and establish its relation to other measures.
- Due to this, it is not suitable for further mathematical processing.

Coefficient of mean deviation

The coefficient or relative dispersion is found by dividing the mean deviations recorded. Thus,

$$\text{Coefficient of M.D.} \quad = \frac{\text{Mean Deviation}}{\text{Mean}} \qquad \qquad \text{...(6.5)}$$

(when deviations were recorded from the mean)

$$= \frac{\text{M.D.}}{\text{Median}} \qquad \qquad \text{...(6.6)}$$

(when deviations were recorded from the median)

Applying the above formula to Example 6.17.

$$\text{Coefficient of Mean deviation} \quad = \frac{14.775}{24}$$

$$= 0.616$$

6.5.6 Standard Deviation

By far the most universally used and the most useful measure of dispersion is the standard deviation or root mean square deviation about the mean. We have seen that all the methods of measuring dispersion so far discussed are not universally adopted for want of adequacy and accuracy. The range is not satisfactory as its magnitude is determined by most extreme cases in the entire group. Further, the range is notable because it is dependent on the item whose size is largely matter of chance. Mean deviation method is also an unsatisfactory measure of scatter, as it ignores the algebraic signs of deviation.

We desire a measure of scatter which is free from these shortcomings. To some extent standard deviation is one such measure.

The calculation of standard deviation differs in the following respects from that of mean deviation. First, in calculating standard deviation, the deviations are squared. This is done so as to get rid of negative signs without committing algebraic violence. Further, the squaring of deviations provides added weight to the extreme items, a desirable feature for certain types of series.

Secondly, the deviations are always recorded from the arithmetic mean, because although the sum of deviations is the minimum from the median, the sum of squares of deviations is minimum when deviations are measured from the arithmetic average. The deviation from $\bar{x}$ is represented by d.

Thus, standard deviation, σ (sigma) is defined as the square root of the mean of the squares of the deviations of individual items from their arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}} \qquad \qquad ...(6.7)$$

For grouped data (discrete variables)

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}} \qquad \qquad ...(6.8)$$

and, for grouped data (continuous variables)

$$\sigma = \sqrt{\frac{\Sigma f(M - \bar{x})}{\Sigma f}} \qquad \qquad ...(6.9)$$

where M is the mid-value of the group.

The use of these formulae is illustrated by the following examples.

Example 6.18: Compute the standard deviation for the following data:

       11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

Solution:

Here formula (6.7) is appropriate. We first calculate the mean as $\bar{x} = \Sigma x/N = 176/11 = 16$, and then calculate the deviation as follows:

x	$(x - \bar{x})$	$(x - \bar{x})^2$
11	−5	25
12	−4	16
13	−3	9
14	−2	4
15	−1	1
16	0	0
17	+1	1
18	+2	4
19	+3	9
20	+4	16
21	+5	25
176		10

Thus by formula (7).

$$\sigma = \sqrt{\frac{110}{11}} = \sqrt{10} = 3.16$$

Example 6.19: Find the standard deviation of the data in the following distributions:

x	12	13	14	15	16	17	18	20
f	4	11	32	21	15	8	6	4

Solution:

For this discrete variable grouped data, we use formula 6.8. Since for calculation of $\bar{x}$, we need $\sum fx$ and then for σ we need $\sum f(x-\bar{x})^2$, the calculations are conveniently made in the following format.

x	f	fx	$d = x - \bar{x}$	d^2	fd^2
12	4	48	−3	9	36
13	11	143	−2	4	44
14	32	448	−1	1	32
15	21	315	0	0	0
16	15	240	1	1	15
17	8	136	2	4	32
18	5	90	3	9	45
20	4	80	5	25	100
	100	1500			304

Here $\bar{x} = \sum fx/\sum f = 1500/100 = 15$

and $\sigma = \sqrt{\dfrac{\sum fd^2}{\sum f}}$

$= \sqrt{\dfrac{304}{100}} = \sqrt{3.04} = 1.74$

Example 6.20: Calculate the standard deviation of the following data.

Class	1–3	3–5	5–7	7–19	9–11	11–13	13–15
frequency	1	9	25	35	17	10	3

Solution: This is an example of continuous frequency series and formula 6.9 seems appropriate.

Class	Mid-point x	Frequency f	fx	Deviation of mid-point x from mean (8)	Squared deviation d^2	Squared deviation times frequency d^2
1–3	2	1	2	−6	36	36
3–5	4	9	36	−4	16	144
5–7	6	25	150	−2	4	100
7–9	8	35	280	0	0	0
9–11	10	17	170	2	4	68
11–13	12	10	120	4	16	160
13–15	14	3	42	6	36	108
		100	800			616

First the mean is calculated as

$\bar{x} = \sum fx/\sum x = 800/100 = 8.0$

Then the deviations are obtained from 8.0. The standard deviation

$$\sigma = \sqrt{\frac{\sum f(M - \bar{x})^2}{\sum f}}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} = \sqrt{\frac{616}{100}}$$

$$= 2.48$$

Calculation of Standard Deviation by Short-cut Method

The three examples worked out above have one common simplifying feature, namely $\bar{x}$ in each, turned out to be an integer, thus, simplifying calculations. In most cases, it is very unlikely that it will turn out to be so. In such cases, the calculation of d and d^2 becomes quite time-consuming. Short-cut methods have consequently been developed. These are on the same lines as those for calculation of mean itself.

In the short-cut method, we calculate deviations x' from an assumed mean A. Then,

for ungrouped data

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2} \qquad \qquad ...(6.10)$$

and for grouped data

$$\sigma = \sqrt{\frac{\sum fx'^2}{\sum f} - \left(\frac{fx'}{\sum f}\right)^2} \qquad \qquad ...(6.11)$$

This formula is valid for both discrete and continuous variables. In case of continuous variables, x in the equation $x' = x - A$ stands for the mid-value of the class in question.

Note that the second term in each of the formulae is a correction term because of the difference in the values of A and $\bar{x}$. When A is taken as $\bar{x}$ itself, this correction is automatically reduced to zero. Examples 6.7 to 6.11 explain the use of these formulae.

Example 6.21: Compute the standard deviation by the short-cut method for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

Solution: Let us assume that $A = 15$.

	$x' = (x - 15)$	x'^2
11	−4	16
12	−3	9
13	−2	4
14	−1	1
15	0	0
16	1	1
17	2	4
18	3	9
19	4	16
20	5	25
21	6	36
$N = 11$	$\sum x' = 11$	$\sum x'^2 = 121$

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

$$= \sqrt{\frac{121}{11} - \left(\frac{11}{11}\right)^2}$$

$$= \sqrt{11 - 1}$$

$$= \sqrt{10}$$

$$= 3.16.$$

Another method

If we assumed A as zero, then the deviation of each item from the assumed mean is the same as the value of item itself. Thus, 11 deviates from the assumed mean of zero by 11, 12 deviates by 12, and so on. As such, we work with deviations without having to compute them, and the formula takes the following shape:

x	x^2
11	121
12	144
13	169
14	196
15	225
16	256
17	289
18	324
19	361
20	400
21	441
176	2,926

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

$$= \sqrt{\frac{2926}{11} - \left(\frac{176}{11}\right)^2} = \sqrt{266 - 256} = 3.16$$

Example 6.22: Calculate the standard deviation of the following data by short method.

Person	1	2	3	4	5	6	7
Monthly income (Rupees)	300	400	420	440	460	480	580

Solution: In this data, the values of the variable are very large making calculations cumbersome. It is advantageous to take a common factor out. Thus, we use $x' = \frac{x - A}{20}$.

The standard deviation is calculated using x' and then the true value of σ is obtained by multiplying back by 20. The effective formula then is

$$\sigma = C \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

where C represents the common factor.

Using $x' = (x - 420)/20$.

x	Deviation from Assumed mean $x' = (x - 420)$	x'	x'^2
300	−120	−6	36
400	−20	−1	1
420	0	0	0
		−7	
440	20	1	1
460	40	2	4
480	60	3	9
580	160	8	64
		+ 14	
$N = 7$		7	115

$$\sigma = 20 \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

$$= 20\sqrt{\frac{115}{7} - \left(\frac{7}{7}\right)^2}$$

$$= 78.56$$

Example 6.23: Calculate the standard deviation from the following data:

Size	6	9	12	15	18
Frequency	7	12	19	10	2

Solution:

x	Frequency f	Deviation from assumed mean 12	Deviation divided by common factor 3 x'	x' times frequency fx'	x'^2 times frequency fx'^2
6	7	−6	−2	−14	28
9	12	−3	−1	−12	12
12	19	0	0	0	0
15	10	3	1	10	10
18	2	6	2	4	8
	$N = 50$			$\sum fx' = -12$	$\sum fx'^2 = 58$

Since deviations have been divided by a common factor, we use

$$\sigma = C\sqrt{\frac{\sum fx'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

$$= 3\sqrt{\frac{58}{50} - \left(\frac{-12}{50}\right)^2}$$

$$= 3\sqrt{1.1600 - .0576} = 3 \times 1.05 = 3.15.$$

Example 6.24: Obtain the mean and standard deviation of the first N natural numbers, i.e., of 1, 2, 3, ..., N – 1, N.

Solution: Let x denote the variable which assumes the values of the first N natural numbers.

Then

$$\bar{x} = \frac{\sum_{1}^{N} x}{N} = \frac{\frac{N(N+1)}{2}}{N} = \frac{N+1}{2}$$

because $\sum_{1}^{N} x = 1 + 2 + 3 + ... + (N-1) + N$

$$= \frac{N(N+1)}{2}$$

To calculate the standard deviation σ, we use 0 as the assumed mean A. Then

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

But $\sum x^2 = 1^2 + 2^2 + 3^2 + ... (N-1)^2 + N^2 = \frac{N(N+1)(2N+1)}{6}$

Therefore

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{6N} - \frac{N^2(N+1)^2}{4N^2}}$$

$$= \sqrt{\frac{(N+1)}{2}\left[\frac{2N+1}{3} - \frac{N+1}{2}\right]} = \sqrt{\frac{(N+1)(N-1)}{12}}$$

Thus for first 11 natural numbers

$$\bar{x} = \frac{11+1}{2} = 6$$

and $\sigma = \sqrt{\frac{(11+1)(11-1)}{12}} = \sqrt{10} = 3.16$

Example 6.25:

	Mid-point x	Frequency f	Deviation from class of assumed mean x'	Deviation time frequency fx'	Squared deviation times frequency fx'^2
0–10	5	18	–2	–36	72
10–20	15	16	–1	–16	16
				–52	
20–30	25	15	0	0	0
30–40	35	12	1	12	12
40–50	45	10	2	20	40
50–60	55	5	3	15	45
60–70	65	2	4	8	32
70–80	75	1	5	5	25
				– 60	
		79		60	242
				–52	
				$\sum fx' = 8$	

Solution: Since the deviations are from assumed mean and expressed in terms of class-interval units,

$$\sigma = i \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

$$= 10 \times \sqrt{\frac{242}{79} - \left(\frac{8}{79}\right)^2}$$

$$= 10 \times 1.75 = 17.5.$$

Combining Standard Deviations of Two Distributions

If we were given two sets of data of N_1 and N_2 items with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively, we can obtain the mean and standard deviation $\bar{x}$ and σ of the combined distribution by the following formulae:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} \qquad \qquad ...(6.12)$$

and $\quad \sigma = \sqrt{\dfrac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}} \qquad ...(6.13)$

Example 6.26: The mean and standard deviations of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the standard deviation of all taken together.

Solution: Combined mean

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{100 \times 50 + 150 \times 40}{100 + 150}$$

$$= 44$$

Combined standard deviation

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{100 \times (5)^2 + 150(6)^2 + 100(44 - 50)^2 + 150(44 - 40)^2}{100 + 150}}$$

$$= 7.46.$$

Example 6.27: A distribution consists of three components with 200, 250, 300 items having mean 25, 10 and 15 and standard deviation 3, 4 and 5, respectively. Find the standard deviation of the combined distribution.

Solution: In the usual notations, we are given here

$$N_1 = 200, N_2 = 250, N_3 = 300$$
$$\bar{x}_1 = 25, \bar{x}_2 = 10, \bar{x}_3 = 15$$

The formulae (12) and (13) can easily be extended for combination of three series as

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$$

$$= \frac{200 \times 25 + 250 \times 10 + 300 \times 15}{200 + 250 + 300}$$

$$= \frac{12000}{750} = 16$$

and

$$\sigma = \sqrt{\frac{\begin{array}{c} N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1(\bar{x} - \bar{x}_1)^2 \\ + N_2(\bar{x} - \bar{x}_2)^2 + N_3(\bar{x} - \bar{x}_3)^2 \end{array}}{N_1 + N_2 + N_3}}$$

$$= \sqrt{\frac{200 \times 9 + 250 \times 16 + 300 \times 25 + 200 \times 81 + 250 \times 36 + 300 \times 1}{200 + 250 + 300}}$$

$$= \sqrt{51.73} = 7.19.$$

Comparison of Various Measures of Dispersion

The range is the easiest to calculate the measure of dispersion, but since it depends on extreme values, it is extremely sensitive to the size of the sample, and to the sample variability. In fact, as the sample size increases the range increases dramatically, because the more the items one considers, the more likely it is that some item will turn up which is larger than the previous maximum or smaller than the previous minimum. So, it is, in general, impossible to interpret properly the significance of a given range unless the sample size is constant. It is for this reason that there appears to be only one valid application of the range, namely in statistical quality control where the same sample size is repeatedly used, so that comparison of ranges are not distorted by differences in sample size.

The quartile deviations and other such positional measures of dispersions are also easy to calculate but suffer from the disadvantage that they are not amenable to algebraic treatment. Similarly, the mean deviation is not suitable because we cannot obtain the mean deviation of a combined series from the deviations of component series. However, it is easy to interpret and easier to calculate than the standard deviation.

The standard deviation of a set of data, on the other hand, is one of the most important statistics describing it. It lends itself to rigorous algebraic treatment, is rigidly defined and is based on all observations. It is, therefore, quite insensitive to sample size (provided the size is 'large enough') and is least affected by sampling variations.

It is used extensively in testing of hypothesis about population parameters based on sampling statistics.

In fact, the standard deviations has such stable mathematical properties that it is used as a standard scale for measuring deviations from the mean. If we are told that the performance of an individual is 10 points better than the mean, it really does not tell us enough, for 10 points may or may not be a large enough difference to be of significance. But if we know that the *s* for the score is only 4 points, so that on this scale, the performance is 2.5*s* better than the mean, the statement becomes meaningful. This indicates an extremely good performance. This sigma scale is a very commonly used scale for measuring and specifying deviations which immediately suggest the significance of the deviation.

The only disadvantages of the standard deviation lies in the amount of work involved in its calculation, and the large weight it attaches to extreme values because of the process of squaring involved in its calculations.

6.6 SUMMARY

- Data processing involves the analysis and manipulation of the collected data by performing various functions. The functions that can be performed on data are: editing, coding, tabulation and classification.

- Editing of data involves the testing of data collection instruments in order to ensure maximum accuracy.

- Coding of data can be defined as representing the data symbolically using some predefined rules.

- Classification of data involves arrangement of data in groups or classes on the basis of some common characteristics.

- Tabulation means placing the data collected and results from research in a tabular form.

- Analysis of data is the process of transforming data for the purpose of extracting useful information, which in turn facilitates the discovery of some useful conclusions.

- The various types of analyses are: multiple regression analysis, multiple discriminant analysis and canonical correlation analysis.

- Analysis on the basis of the variance that exists in the data can be: univariate analysis, bivariate analysis and multivariate analysis.

- Data interpretation refers to the identification of trends in different variables. The researcher uses statistics for this purpose. The level of measurement used can be nominal, ordinal, interval or ratio.

- Data in raw form can be cumbersome to deal with. It must be presented in a suitable and summarized form without any loss of relevant information so that it can be efficiently used for decision-making. The common forms of representation of data are: graphs, tables, diagrams, etc.

- Tabulation is a systematic arrangement of data in columns and rows. The analysis of the data is done so by arranging the columns and rows to facilitate analysis and comparisons. A table is constructed depending on the type of information to be presented and the requirements of the statistical analysis.

- Diagrams and graphs give visual indications of magnitudes, groupings, trends and patterns in the data. These important features are more simply presented in the form of graphs. Also, diagrams facilitate comparisons between two or more sets of data.

- While arithmetic mean is the most commonly used measure of central tendency, mode and median are more suitable measures under certain set of conditions and for certain types of data.

- Arithmetic mean is commonly known as the mean. Even though average, in general, means measure of central tendency, when we use the word average in our daily routine, we always mean the arithmetic average.

- Median is that value of a variable which divides the series in such a manner that the number of items below it is equal to the number of items above it.

- The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It is the most frequent or the most common value, provided that a sufficiently large number of items are available, to give a smooth distribution.

- Some measures than other measures of central tendency are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. The quartiles divide the data into four equal parts, the deciles divide the total ordered data into ten equal parts and the percentile divide the data into 100 equal parts.

- A measure of dispersion, or simply dispersion may be defined as statistics signifying the extent of the scatteredness of items around a measure of central tendency. These can be in the form of mean deviation, quartile deviation or standard deviation.

6.7 KEY TERMS

- **Data processing:** It refers to the analysis and manipulation of the collected data by performing various functions.

- **Coding of data:** It is defined as representing the data symbolically using some predefined rules.

- **Analysis of data:** It is the process of transforming data for the purpose of extracting useful information, which in turn facilitates the discovery of some useful conclusions.

- **Mean:** It refers to the arithmetic average and measure of central location.

- **Mode:** It is a form of average that can be defined as the most frequently occurring value in the data.

- **Median:** It refers to a measure of central tendency that appears in the centre of an ordered data.

- **Standard deviation:** The square root of the average of the squared deviations from their mean of a set of observations.

6.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The functions that can be performed on data are: editing, coding, tabulation and classification.

2. Coding decisions are usually taken at the designing state of the questionnaire.

3. Class intervals refer to a range of values of a variable. This interval is used to calibrate the scale of a variable in order to tabulate the frequency distribution of a sample.

4. The following are the different methods used in hand tabulation: direct tally method, list and tally method and card sort method.

5. Bivariate analysis is a type of analysis that examines the relationship between two variables. It tries to find the extent of association that exists among these variables.

6. The ordinal scale of measurement is used to calculate and derive data pertaining to the median, percentage, rank order, correlations and percentile.

7. A frequency polygon is a line chart of frequency distribution in which, either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined tighter by straight lines.

8. In histograms, the given data is plotted in the form of a series of rectangles. Class intervals are marked along the X-axis and the frequencies along the Y-axis according to a suitable scale.

9. The following are the characteristics of mean: the sum of the deviation of individual values of X from the mean will always add up to zero, it is very sensitive to extreme values, and the sum of the squares of the deviations about the mean is minimum.

10. When there is an even number of cases, there is no actual middle item and the median is taken to be the average of the values of the items lying on either side of (N+1)/2, where N is the total number of items.

11. The four important methods of estimating mode of a series are: (i) locating the most frequently repeated value in the array; (ii) estimating the mode by interpolation; (iii) locating the mode by graphic method; and (iv) estimating the mode from the mean and the median.

12. Range is the crudest measure of dispersion. It is the difference between the highest and lowest values in the series.

13. The median deviation is preferred over mean because it has the important property that the average deviation from it is the least.

6.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Distinguish between the terms data classification according to attributes and class intervals.

2. State any three advantages of tabulation.

3. What is an Ogive curve?

4. Differentiate between a mean and a mode.

5. Write three characteristics of mean.

6. What is the importance of arithmetic mean in statistics?

7. Define quartiles, deciles and percentiles with suitable examples.

8. What is geometric mean? How is it calculated?

9. Write the definition and formula of quartile deviation.

10. How will you calculate the mean deviation of a given data?

11. Explain standard deviation. Why is it used in statistical evaluation of data?

Long-Answer Questions

1. Briefly describe the process and significance of coding of data.

2. Enumerate and elaborate on the methods of tabulation.

3. Elaborate on any three ways of representation of the collected data.

4. Discuss the various types of diagrams used for data representation.

5. Explain the term descriptive statistics with the help of examples.

6. Discuss the various measures of central tendency.

7. Discuss the significance of diagrammatic representation of data.

8. Explain the common techniques of diagrammatic representation.

9. The following table gives the heights (in inches) of 100 boys of a class. Calculate mean, mode and median of the height.

Height (inches)	No. of Students
60–62	5
62–64	18
64–66	42
66–68	20
68–70	8
70–72	7
	100

Solution: 65.58

10. The daily profits in rupees of 100 shops are distributed as follows. Draw a histogram of the data and then find the modal value. Check this value by direct calculation.

Profits per shop	Number of shops
0–100	12
100–200	18
200–300	27
300–400	20
400–500	17
500–600	6

Since class 200–300 has the highest frequency, i.e., 27, mode lies in this class.

Mode = 256.25

6.10 FURTHER READING

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House.

Gupta, S. C. 2006. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.

Gupta, S. P., 2005. *Statistical Methods*. New Delhi: Sultan Chand and Sons.

Hooda, R. P. 2002. *Statistics for Business and Economics*. New Delhi: Macmillan India.

Kothari, C. R., 1984. *Quantitative Techniques*. New Delhi: Vikas Publishing House.

Monga, G. S. 2000. *Mathematics and Statistics for Economics*. New Delhi: Vikas Publishing House.

UNIT 7 REPORT WRITING

Structure

7.0 INTRODUCTION

The final unit will discuss the writing of research reports. A research study is a tedious task and calls for exhaustive investigation on the part of the researcher. This quite often leads to accumulation of bulk data obtained from the research study. Even if the concerned study results in brilliant hypotheses or a generalized theory, it is the responsibility of the researcher to format this bulk study into an easy-to-understand pattern or format. This is where a research report comes in.

One cannot overemphasize the significance of a well-documented and structured research report. This step is often taken as extremely rudimentary and is, thus, ignored. However, just like all the other steps in the research process, this requires careful and sequential progression. The unit will discuss in detail the formation and presentation of the research study. The format and the steps might be moderately adjusted and altered based on the reader's requirement. Thus, it might be for an academic and theoretical purpose or might need to be clearly spelt and linked with the business manager's decision dilemma.

7.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss the importance of report writing
- Describe the types of research reports
- Explain the report preparation and presentation
- Identify the components of report
- Assess the formulation of report writing and guidelines

7.2 NEED FOR EFFECTIVE DOCUMENTATION

On completion of the research study and after obtaining the research results, the real skill of the researcher lies in terms of analysing and interpreting the findings and linking them with the propositions formulated in the form of research hypotheses at the beginning of the study. The statistical or qualitative summary of results would be little more than numbers or conclusions unless one is able to present the documented version of the research endeavour.

7.2.1 Importance of Report Writing

Depending on the business researcher's orientation, the intention might be different and would be reflected in the form of the presentation but the significance is critical to both. Essentially, this is so because of the following reasons:

- The research report fulfills the historical task of serving as a concrete proof of the study that was undertaken. This serves the purpose of providing a framework for any work that can be conducted in the same or related areas.

- It is the complete detailed report of the research study undertaken by the researcher, thus it needs to be presented in a comprehensive and objective manner. This is a one-way communication of the researcher's study and analysis to the reader/manager, and thus needs to be all-inclusive and yet neutral in its reporting.

- For academic purposes, the recorded document presents a knowledge base on the topic under study and for the business manager seeking help in taking more informed decisions, the report provides the necessary guidance for taking appropriate action.

- As the report documents all the steps followed and the analysis carried out, it also serves to authenticate the quality of the work carried out and establishes the strength of the findings obtained.

Thus, effective recording and communicating of the results of the study becomes an extremely critical step of the research process. Based on the nature of the research study and the researcher's orientation, the report can take different forms.

7.2.2 Types of Research Reports

The form and structure of the research report might change according to the purpose for which it has been designed. Based on the size of the report, it is possible to divide the report into the following types:

Brief Reports

These kinds of reports are not formally structured and are generally short, sometimes not running more than four to five pages. The information provided is of a limited scope and is prepared either for immediate consumption or as a prelude to the formal structured report that would subsequently follow. These reports could be designed in several ways.

- *Working papers* or *basic reports* are written for the purpose of collating the process carried out in terms of scope and framework of the study, the methodology followed and instrument designed. The results and findings would also be recorded

here. However, the interpretation of the findings and study background might be missing, as the focus is more on the present study rather than past literature. These reports are significant as they serve as a reference point when writing the final report or when the researcher wants to revisit the detailed steps followed in collecting the study-related information.

- *Survey reports* might or might not have an academic orientation. The focus here is to present findings in an easy-to-comprehend format that includes figures and tables. The reader can then study the patterns in the findings to arrive at appropriate conclusions, essential for resolving the research problem. The advantage of these reports is that they are simple and easy to understand and present the findings in a clear and usable format.

Detailed Reports

These are more formal and pedantic in their structure and are essentially either academic, technical or business reports. Sometimes, the researcher may prepare both kinds—for an academic as well as for a business purpose. The language, presentation and format of the two kinds of reports would be vastly different as they would need to be prepared for the understanding of the reader's capabilities and intentions.

Technical Reports

These are major documents and would include all elements of the basic report, as well as the interpretations and conclusions, as related to the obtained results. This would have a complete problem background and any additional past data/records that are essential for comprehending and interpreting the present study output. All sources of data, sampling plan, data collection instrument(s), data analysis outputs would be formally and sequentially documented.

Business Reports

These reports would not have the technical rigour and details of the technical report and would be in the language and include conclusions as understood and required by the business manager. The tables, figures and numbers of the first report would now be pictorially shown as bars and graphs and the reporting tone would be more in business terms rather than in conceptual or theoretical terms. If needed, the tabular data might be attached in the appendix.

7.3 COMPONENTS OF REPORT

Whatever the type of report, the reporting and dissemination of the study and its findings require a structured format and by and large, the process is standardized. As stated above, the major difference amongst the types of reports is that all the elements that essentially constitute a research report would be present only in a detailed technical report.

The entire research project needs to be recorded either as a single written report or into several reports, depending on the need of the readers. The researcher would need to assist the business manager in deciphering the report, executing the findings, and in case of need, to revise the report to suit the specific actionable requirements of the manager.
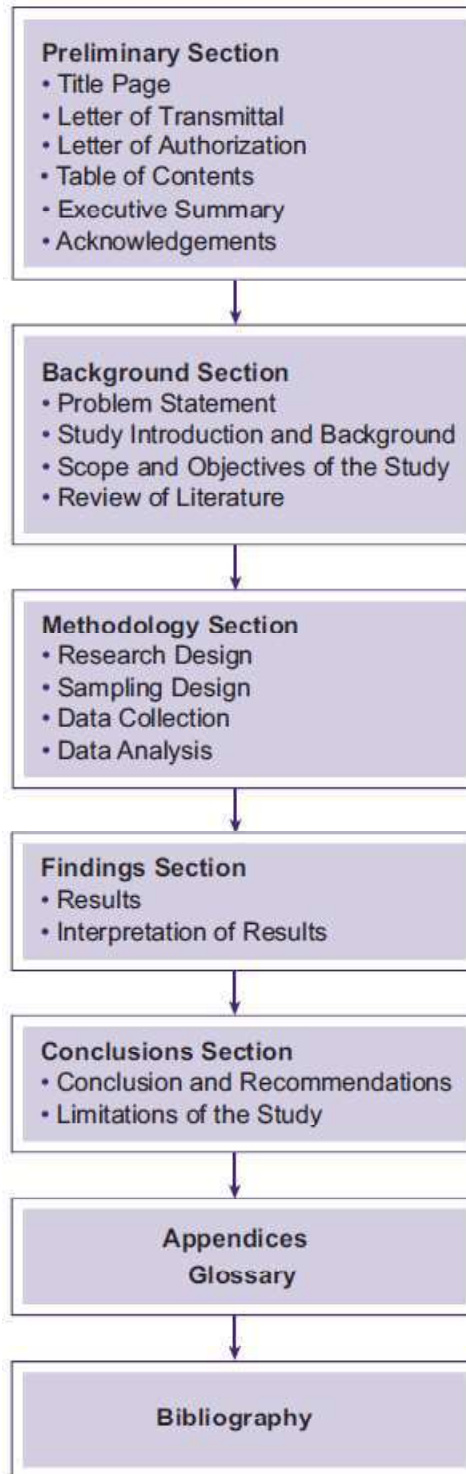
As presented in Figure 7.1, most research reports include the following sections:

Preliminary Section
• Title Page
• Letter of Transmittal
• Letter of Authorization
• Table of Contents
• Executive Summary
• Acknowledgements

Background Section
• Problem Statement
• Study Introduction and Background
• Scope and Objectives of the Study
• Review of Literature

Methodology Section
• Research Design
• Sampling Design
• Data Collection
• Data Analysis

Findings Section
• Results
• Interpretation of Results

Conclusions Section
• Conclusion and Recommendations
• Limitations of the Study

Appendices
Glossary

Bibliography

Fig. 7.1 *The Process of Report Formulation and Writing*

1. Preliminary Pages

This section mainly consists of identification information for the study conducted. It has the following individual elements:

Title page: This includes classification data about:

- The target audience, or the intended reader of the report.
- The report author(s), including their name, affiliation and address.
- The title of the study presented in a manner to clearly indicate the study variables; the relationship or status of the variables studied and the population to which the results apply. The title should be crisp and indicative of the nature of the project, as illustrated in the following examples.
  - o Comparative analysis of BPO workers and schoolteachers with reference to their work–life balance
  - o Segmentation analysis of luxury apartment buyers in the National Capital Region (NCR).
  - o An assessment of behavioural factors impacting consumer financial investment decisions.

Letter of transmittal: This is the letter that goes alongside the formalized copy of the final report. It broadly refers to the purpose behind the study. The tone in this note can be slightly informal and indicative of the rapport between the client-reader and the researcher. A sample letter of transmittal is presented in Exhibit 7.1. The letter broadly refers to three issues: It indicates the term of the study or objectives; next it goes on to broadly give an indication of the process carried out to conduct the study and the implications of the findings. The conclusions are generally indicative of the researcher's interest/learning from the study and in some cases may be laying the foundation for future research opportunities.

Letter of authorization: Sometimes the letter of authorization may be redundant as indications of the formal approval for conducting the study might be included in the letter of transmittal. The author of this letter is the business manager or corporate representative who formally gives the permission for executing the project. The tone of this letter, unlike the above document, is very precise and formal, leaving no room for speculation or interpretation.

Exhibit 7.1 Sample Letter of Transmittal

To:	Mr Prem Parashar	**From:**	Nayan Navre
Company:	Just Bondas Corporation (JBC)	**Company:**	Jigyasa Associates
Location:	Mumbai 116879	**Location:**	Sabarmati Dham, Mumbai
Telephone:	48786767; 4876768	**Telephone:**	41765888
Fax:	48786799	**Fax:**	41765899

Addendums: Highlight of findings (pages: 20)

15 January 2011

Dear Prem,

Please find the enclosed document which covers a summary of the findings of the November- December 2010 study of the new product offering and its acceptibility. I would be sending three hard copies of the same tomorrow.

> Once the core group has discussed the direction of the expected results I would request you to kindly get back with your comments/queries/suggestions, so that they can be incorporated in the preparation of the final report document.
>
> The major findings of the study were that the response of the non-vegetarians consuming the new *keema bonda pav* at Just Bondas was positive. As you can observe, however, the introduction of the non-vegetarian *bonda* has not been well received by the regular customers who visit the outlets for their regular *alloo bonda*. These findings, though on a small respondent base, are significant as they could be an indication of a deflecting loyal customer base.
>
> Best regards,
>
> Nayan

As explained, this letter is not critical to submission, in case reference to the same has been made in the transmittal letter. However, in case it is to be included in the report, it is advisable to reproduce the exact prototype of the original letter.

Table of contents: All reports should have a section that clearly indicates the division of the report based on the formal areas of the study as indicated in the research structure. The major divisions and subdivisions of the study, along with their starting page numbers, should be presented. The subheadings and the smaller sections of a topic need not be indicated here as then the presentation of the content seems cluttered.

Once the major sections of the report are listed, the list of tables come next, followed by the list of figures and graphs, exhibits (if any) and finally the list of appendices.

Executive summary: This is the last and the most critical element of the preliminary section. The summary of the entire report, starting from the scope and objectives of the study to the methodology employed and the results obtained, have to be presented in a brief and concise manner. In case the research requirement was to provide recommended changes based on the findings, it is advisable to provide short pointers here. Interestingly, it has been observed that in most instances the business managers read only the executive summary in its complete detail and most often just glance through the rest of the report. Thus, it becomes extremely critical to present a Gestaltan view of the entire report in a suitable condensed form.

The executive summary essentially can be divided into four or five sections. It begins with the study background, scope and objectives of the study, followed by the execution, including the sample details and methodology of the study. Next comes the findings and results obtained. The fourth section covers the conclusions which are more or less based on the opinion of the researcher. Finally, as stated earlier, in case the study objectives necessitates implications, the last section would include recommendations and suggestions.

Acknowledgements: A small note acknowledging the contribution of the respondents, the corporates and the experts who provided inputs for accomplishing the study is to be included here.

Though the executive summary comes before the main body of the report, it is always prepared after the entire report has been finalized and is ready in its final form. The length of this section is one or two pages only and the researcher needs to effectively present the most significant parts of the study in a succinct form. It has been observed that the executive summary is a standalone document that is often circulated independently to the interested managers who might be directly or indirectly related to the study.

2. Main Text

This is the most significant and academically robust part of the report. The sections of this division follow the essential pattern of a typical research study.

Problem definition: This section begins with the formal definition of the research problem. The problem statement is the research intention and is more or less similar to what was stated earlier as the title of the research study.

Study background: Study background presents details of the preliminary conceptualization of the management decision problem and all the groundwork done in terms of secondary data analysis, industry experts' perspectives and any other earlier reporting of similar approaches undertaken. Thus, essentially, the section begins by presenting the decision-makers' problem and then moves on to a description of the theoretical and contemporary market data that laid the foundation that guided the research.

In case the study is an academic research, there is a separate section devoted to the review of related literature, which presents a detailed reporting of work done on the same or related topic of interest.

Study scope and objectives: The logical arguments then conclude in the form of definite statements related to the purpose of the study. A clear definition of the scope and objective of the study is presented usually after the study background; in case the study is causal in nature, the formulated hypotheses are presented here as well.

Methodology of research: This section would not be sequentially placed here, for short reports or for a business report. In such reports, a short description of the methodology followed would be documented in the appendix. However, for a technical and academic report, this is a significant and primary contribution of the research study. The section would essentially have five to six sections specifying the details of how the research was conducted. These would essentially be:

- *Research framework or design:* The variables and concepts being investigated are clearly defined, with a clear reference to the relationship being studied. The justification for using a particular design has to be presented in a sequential and step-wise manner enlisting the experimental and control conditions, in case of a causal study. The researcher must take care to keep the technical details of the execution in the appendix and present the execution details in simple language, in the main body.

- *Sampling design:* The entire sampling plan in terms of the population being studied, along with the reasons for collecting the study-related information from the given group is given here. The execution details, in terms of sample size calculations, sampling frame considered and field work details can be recorded in the appendix rather than in the main body of the report. However, the sample profile and identification details are included in the main section. As stated earlier, the report needs to be reader-friendly, and too much technical information might not be required by the decision-maker.

- *Data collection methods:* In this section, the researcher should clearly list the information needed for the study as drawn from the study objectives stated earlier. The secondary data sources considered and the primary instrument designed for the specific study are discussed here. However, the final draft of the measuring instrument can be included in the appendix, which includes the execution details in terms of how the information was collected; how the open ended or opinion-

based questions were handled; and how irregularities were handled and accounted for in the study. These and similar information enable a clear insight into the standardization of procedures maintained.

- *Data analysis:* Here, the researcher again needs to revisit the research objectives and the study design in order to justify the analytical tools and techniques used in the study. The assumptions and constraints of the analysis need to be explained here in simple, non-technical terms. There is no need to give a detailed description of the statistical calculations here.

- *Study results and findings:* This is the most critical chapter of the report and requires special care; it is probably also one of the longest chapters in the document. The researcher could, thus, consider either breaking this into subchapters or at least clear subheadings.

Researchers commonly divide the chapter on the basis of the data collection plan, i.e., there is a section on interview analysis, another one on focus group discussion and the third referring to the questionnaire analysis. This, however, does not serve any purpose as the results would then seem repetitive and disjointed. Instead, the result should be organized according to the information areas on which the data was collected or on the basis of the research objectives. There are also times when the data would be presented for the whole sample and then will be split and presented for the sub-population studied. For example, in the study on work-life balance, the findings were presented for the whole sample and then at the micro level for the BPO sector and separately for the school teacher segment. For each group, first the sample profile in terms of the demographic details of age, education, income (individual and family), years of experience, marital status, family size and other details was presented. Next, the descriptive data was made available on the seven sub-scales studied—and lastly—the predictive data–based on a multiple regression analysis with work-life balance as the dependent variable and the seven variables as independent, was presented. There was only one open-ended question related to the individual's suggestion as to what support was required from one's place of work to achieve work-life balance. This was presented last in the form of a bar chart showing variability in the responses given. Again as advised earlier, it is essential to present the findings in the form of simplified tables, graphs and figures, with the same being explained in simple text subsequently.

Interpretations of Results and Suggested Recommendations

The section study results and findings, i.e., the main report, presents a bird's eye view of the information as it exists in a summarized and numerical form. This kind of information might become difficult to understand and convert into actionable steps, thus the real skill of the researcher lies in simplifying the data in a reader-friendly language. Here, it is recommended that this section should be more analytical and opinion based. The results could be supported by the data that was presented earlier, for example, industry forecasts or the expert opinion. In case the report had an earlier section on literature review, the researcher could demonstrate the similarity of findings with past studies done on the topic. For example, in a study conducted on analysing the antecedents of turnover intention, the results obtained were explained as follows:

The results of the logit regression indicate that organizational commitment, age and martial status are significant at 5 per cent and 10 per cent levels respectively. The results indicate that as organizational commitment increases, the log of odd ratios in the favour of high turnover intention reduces, which is very logical. This is in accordance with the results obtained by Mobley, et al. (1978), Cotton and Tuttle (1986), Igbaria and Greenhaus (1992), Ahuja, et al. (2007). Thus, when employees feel committed to an organization, they are more likely to stay with the organization.

Sometimes, the research results obtained may not be in the direction as found by earlier researchers. Here, the skill of the researcher in justifying the obtained direction is based on his/her individual opinion and expertise in the area of study. For example, in the same study on turnover intentions, contrary findings were explained as follows:

> ...the results indicate that the log of odd ratios in favour of high turnover intention is more in the case of older respondents; this is contrary to the findings of Zeffane and Gul (1995) and Finegold, et al. (2002). However, this has to be understood in the light of the profession, as in India, most people take the BPO sector as a stop-gap career and use the time at the BPO employment as an opportunity to enhance their academic qualification and then move on, which is also one of the reasons why this sector is a young sector.

Subsequent to the subsection on the interpretation of results, sometimes, the study requirement might be to formulate indicative recommendations to the decision-makers as well. Thus, in case the report includes recommendations, they should be realistic, workable and topically related to the industry studied. For example, to the business manager of organic food products, the following recommendation was made to build awareness amongst potential customers about the benefits of organic products:

Organic food study: An illustration: The power of the print media in promoting a high-involvement product is unsurpassed. Thus, articles by leading nutritionists and doctors (88 per cent of consumers are influenced by others in consuming health alternatives) on any aspect of organic food would work well. The organic players need to take care that they do not advertise only their product offerings and price alone but they also need to educate the consumer on the health benefits of the products in their advertisements.

The article/advertisement could be placed in the Sunday supplements of newspapers so that people would read them at leisure. The major decision-makers for groceries are women thus magazines like Femina, Health and Savvy would be likely choices (the magazines suggested are English fortnightlies and have a reader profile similar to our sample profile). This is also because the product is a premium and niche product and thus requires selective exposure.

Limitations of the Study

The last in this section is a brief discussion of the problems encountered during the study and the constraints in terms of time, financial or human resources. There could also have been constraints in obtaining the required information, either because the data about the topic of interest has not been collected or because it is not readily available to all. These clear revelations about the drawbacks are thus kept in mind by the reader when analysing the results and the implications of the study.

3. End Text

The final section of the report provides all the supportive material in the study. Some of the common details presented in this section are as follows:

Appendices: The appendix section follows the main body of the report and essentially consists of two kinds of information:

1. Secondary information like long articles or in case the study uses/is based on/ refers to some technical information that needs to be understood by the reader. Or long tables or articles or legal or policy documents.

2. Primary data that can be compressed and presented in the main body of the report. This includes: Original questionnaire, discussion guides, formula used for the study, sample details, original data, long tables and graphs which can be described in statement form in the text.

Bibliography: This is an important part of the final section as it provides the complete details of the information sources and papers cited in a standardized format. It is recommended to follow the publication manuals from the American Psychological Association (APA) or the Harvard method of citation for preparing this section. In fact, with the advancement in computer technology the latest version of Microsoft office Word 2007 can automatically generate a bibliography based on any of these formats, based on the source information provided in the document.

The reporting content of the bibliography could also be in terms of:

- **Selected bibliography:** Selective references are cited in terms of relevance and reader requirement. Thus, the books or journals, that are technical and not really needed to understand the study outcomes are not reported.

- **Complete bibliography:** All the items that have been referred to, even when not cited in the text, are given here.

- **Annotated bibliography:** Along with the complete details of the cited work, some brief information about the nature of information sought from the article is given. This could run into three or four lines or a brief paragraph.

At this juncture we would like to refer to another method of citation that an author might wish to use during report writing. This could be in the form of a footnote. To explain the difference we would first like to explain what a typical footnote is:

Footnote: A typical footnote, as the name indicates, is part of the main report and comes at the bottom of a page or at the end of the main text. This could refer to a source that the author has referred to or it may be an explanation of a particular concept referred to in the text.

The referencing protocol of a footnote and bibliography is different. In a footnote, one gives the first name of the person first and the surname next. However, this order is reversed in the bibliography. Here we start first with the surname and then the first name. In a bibliography, we generally mention the page numbers of the article or the total pages in the book. However, in a footnote, the specific page from which the information is cited is mentioned. A bibliography is generally arranged alphabetically depending on the author's name, but in the footnote the reporting is based on the sequence in which they occur in the text.

Glossary of terms: In case there are specific terms and technical jargon used in the report, the researcher should consider putting a glossary in the form of a word list of terms used in the study. This section is usually the last section of the report.

7.4 REPORT WRITING: REPORT FORMULATION

An important point to remember in report writing is that the document compiled is meant for specific readers. Thus, one needs to design the same according to the needs of the reader. Listed below are some features of a good research study that should be kept in mind while documenting and preparing the report.

Clear report mandate: While writing the research problem statement and study background, the writer needs to be focused, precise and very explicit in terms of the problem under study, the background that provided the impetus to conduct the research and the study domain. This is prepared on the assumption that the writer at no point in time needs to be physically present in order to clarify the research mandate. One cannot make an assumption that the reader has earlier insights into the problem situation. The writer needs to be absolutely clear on the need for lucidity of thought and dissemination of this knowledge to the reader.

Clearly designed methodology: Any research study has its unique orientation and scope and thus has a specific and customized research design, sampling and data collection plan. The writer, thus, needs to be explicit in terms of the logical justification for having used the study methods and techniques. However, as stated earlier, the language should be non-technical and reader friendly and any technical explanations or details must be provided in the appendix. In researches, that are not completely transparent on the set of procedures, one cannot be absolutely confident of the findings and resulting conclusions.

Clear representation of findings: The sample size for each analysis, any special conditions or data treatment must be clearly mentioned either as a footnote or as an endnote, so that the reader takes this into account while interpreting and understanding the study results. The sample base is very important in justifying a trend or taking a strategic decision; for example, if amongst a sample of bachelors we say that 100 per cent young bachelors want to buy grocery online or on the telephone and the recommended strategy is to suggest this as the delivery channel, one might be making an error if the size of the bachelors was four out of a total sample of 100 grocery buyers considered. Thus, complete honesty and transparency in stating the treatment and editing of missing or contrary data is extremely critical.

Representativeness of study finding: A good research report is also explicit in terms of extent and scope of the results obtained, and in terms of the applicability of findings. This is also dependent on whether the assumptions and preconditions made for formulating the conclusions and recommendations of the study have been explicitly stated.

In order to ensure that one has been able to achieve the above stated objective, the reader must ensure a standardization of procedures in writing the document as well as follow standard protocols for preparing graphs and tables. In the following section we will briefly discuss some simple rules that the researcher can use as guidelines for this.

Check Your Progress

3. List the three issues addressed by the letter of transmittal.
4. What is an executive summary?
5. List the sections which specify how the methodology of research.
6. Give examples of primary data which is included in the appendices.

7.4.1 Guidelines for Effective Documentation

The following are certain guidelines that are needed for effective documentation.

Command over the medium: Even though one may have done an extremely rigorous and significant research study, the fundamental test still remains as to how the learning has been disseminated. Regardless of how effective the graphs and figures are in showcasing the findings, the verbal description and explanation—in terms of why it was done, how it was done, and what was the outcome, still remain the acid test.

Thus, a correct and effective language of communication is critical in putting ideas and objectives in the vernacular of the reader/decision-maker. The writer may, thus, be advised to read professionally written reports and, if necessary, seek assistance from those proficient in preparing business reports.

Phrasing protocol: There is a debate about whether or not one makes use of personal pronoun while reporting. To understand this, one needs to revisit the responsibility of the researcher, which is to present the findings of his/her study, with complete objectivity and precision. The use of personal pronoun such as 'I think…..' or 'in my opinion…..' lends a subjectivity and personalization of judgement. Thus, the tone of the reporting should be neutral. For example:

'Given the nature of the forecasted growth and the opinion of the respondents, it is likely that the……'

Whenever the writer is reproducing the verbatim information from another document or comment of an expert or published source, it must be in inverted commas or italics and the author or source should be duly acknowledged.

For example:

Sarah Churchman, Head of Diversity, PricewaterhouseCoopers, states 'At PricewaterhouseCoopers we firmly believe that promoting work–life balance is a 'business-critical' issue and not simply the 'right thing to do'. Profitable growth and sustainable business depends on attracting and retaining top talent and we know, from our own research and experience that work–life policies are an essential ingredient of successful recruitment and retention strategies.'

The writer should avoid long sentences and break up the information in clear chunks, so that the reader can process it with ease. Similar is the case in structuring of the chapters or sections of the report that can be logically broken down into smaller sections that are comprehensive and complete and yet maintain a strong but logical link with the flow of reporting.

With the onset of the use of abbreviated communications in SMS and emails, most people tend to use shortened form as 'cd.' for could and 'u' for you, etc. Also the use of colloquial language and slangs must be avoided, as this is a formal document and one must maintain the sanctity of the formal documentation required in a research report.

Simplicity of approach: Along with grammatically and structurally correct language, care must be taken to avoid technical jargon as far as possible. The business manager, might have been a business student who had prepared a research report in his academic pursuits but now understands simple common terms and does not have the time or inclination to juggle the dictionary and the report together. In case it is imperative to use certain terminology, then, as stated earlier, the definition of these terms can be provided in the glossary of terms at the end of the report.

Sometimes the writer may prepare different research reports for the same study to suit the need of diverse readers, for example, the business report needs to be crisp and simple with definable and workable recommendations. On the other hand, an academic report could discuss extensively the literature review section, as well as the statistical analysis and interpretation.

Report formatting and presentation: In terms of paper quality, page margins and font style and size, a professional standard should be maintained. The font style must be uniform throughout the report. The topics, subtopics, headings and subheadings must be construed in the same manner throughout the report. Sometimes certain academic reports have a mandated format for presentation which the writers need to follow, in which case there is no choice in presentation.

However, when this is not clear, it is advisable that the writer creates his/her own formatting rules and saves it on a notepad so that they can be implemented in a standardized and professional manner.

The researcher can provide data relief and variation by adequately supplementing the text with graphs and figures. Pictorial representations are simple to comprehend and also break the monotony and fatigue of reading. They should be used effectively whenever possible in the report.

Guidelines for Presenting Tabular Data

We have discussed this topic in detail in the previous unit. In this section, we will recall some of the concepts again along with some new information.

Most research studies involve some form of numerical data, and even though one can discuss this in text, it is best represented in tabular form. The advantage of doing this is that statistical tables present the data in a concise and numeral form, which makes quantitative analysis and comparisons easier. Tables formulated could be general tables following a statistical format for a particular kind of analysis. These are best put in the appendix, as they are complex and detailed in nature. The other kind is simple summary tables, which only contain limited information and yet, are, essentially critical to the report text.

The mechanics of creating a summary table are very simple and are illustrated below with an example (Table 7.1). The illustration has been labelled with numbers which relate to the relevant section.

Table identification details: The table must have a title (1a) and an identification number (1b). The table title should be short and usually would not include any verbs or articles. It only refers to the population or parameter being studied. The title should be briefly yet clearly descriptive of the information provided. The numbering of tables is usually in a series and generally one makes use of Arabic numbers to identify them.

Data arrays: The arrangement of data in a table is usually done in an ascending manner. This could either be in terms of time, as shown in Table 7.1 (column-wise) or according to sectors or categories (row-wise) or locations, e.g., north, south, east, west and central. Sometimes, when the data is voluminous, it is recommended that one goes alphabetically, e.g., country or state data. Sometimes there may be subcategories to the main categories, for example, under the total sales data—a column-wise component of the revenue statement—there could be subcategories of department store, chemists and druggists, mass merchandisers and others. Then these have to be displayed under the sales data head, after giving a tab command as follows:

Table 7.1 *Automobile Domestic Sales Trends*

Category	Year-wise data (number of cars)				
	2002-2003	2003-2004	2004-2005	2006-2007	2007-2008
Passenger vehicles......	707,198	902,096	1,061,572	1,143,076	1,379,979
Commercial Vehicles......	190,682	260,114	318,430	351,041	467,765
Three-wheelers......	231,529	284,078	307,862	359,920	403,910
Two-wheelers......	4,812,126	5,364,249	6,209,765	7,052,391	7,872,334
Grand Total*	5,941,535	6,810,537	7,897,629	8,906,428	10,123,988

*Does not include second hand car sales.

Source: SIAM

Total sales

Mass market

Department store

Drug stores

Others (including paan beedi outlets)

Measurement unit: The unit in which the parameter or information is presented should be clearly mentioned.

Spaces, Leaders and Rulings (SLR): For limited data, the table need not be divided using grid lines or rulings. Simple white spaces add to the clarity of information presented and processed. In case the number of parameters are too many and the data seems to be bulky to be simply separated by space, it is advisable to use vertical ruling. Horizontal lines are drawn to separate the headings from the main data, as can be seen in Table 7.1. When there are a number of subheadings as in the sales data example, one may consider using leaders (.......) to assist the eye movement in absorbing and processing the information.

Total sales

Mass market………

Department store………

Drug stores………

Others (including *paan beedi* outlets)………

Assumptions, details and comments: Any clarification or assumption made, or a special definition required to understand the data, or formula used to arrive at a particular figure, e.g., total market sale or total market size can be given after the main tabled data in the form of footnotes.

Data sources: In case the information documented and tabled is secondary in nature, complete reference of the source must be cited after the footnote, if any.

Special mention: In case some figure or information is significant and the reader should pay special attention to it, the number or figure can be bold or can be highlighted to increase focus.

Guidelines for Visual Representations: Graphs

Similar to the summarized and succinct data in the form of tables, the data can also be presented through visual representations in the form of graphs. As we have seen the previous unit, the visual representation of the findings in the form of lines or boxes and bars relative to a number line is easy to comprehend and interpret. There are some standard rules and procedures available to the researcher for this; also there are computer programs like MS Excel and SPSS, where the numbered data can be converted with ease into graphical form.

Line and curve graphs: Usually, when the objective is to demonstrate trends and some sort of pattern in the data, a line chart is the best option available to the researcher as the line is able to clearly portray any change in pattern during a particular time period. On the same chart, it is also possible to show patterns of growth of different sectors or industries in the same time period or to compare the change in the studied variable across different organizations or brands in the same industry. Certain points to be kept in mind while formulating line charts include:

- The time units or the causal variable being studied are to be put on the X-axis, or the horizontal axis.

- If the intention is to compare different series on the same chart, the lines should be of different colours or forms (Figure 7.2).

- Too many lines are not advisable on the same chart as then the data becomes too cluttered; an ideal number would be five or less than five lines on the chart.
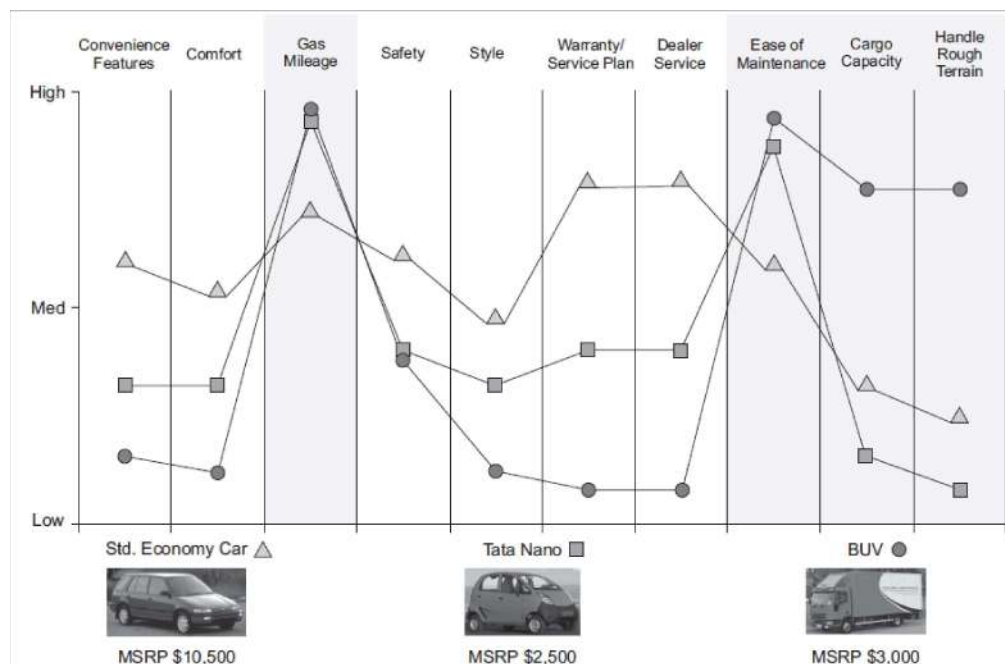
Fig. 7.2 *Comparative Analysis of Vehicles (including Nano) on Features Desired by*

Source: vytrak.com

- The researcher also must take care to formulate the zero baseline in the chart as otherwise, the data would seem to be misleading. For example, in Figure 7.3(a), in case the zero baseline is (as shown in the chart) the expected change in the number of hearing aids units to be sold over the time period 2002–03 to 2007–08,

it can be accurately perceived. However, in Figure 7.3(b), where the zero is at 1,50,000 units, the rate of growth can be misjudged to be more swift.
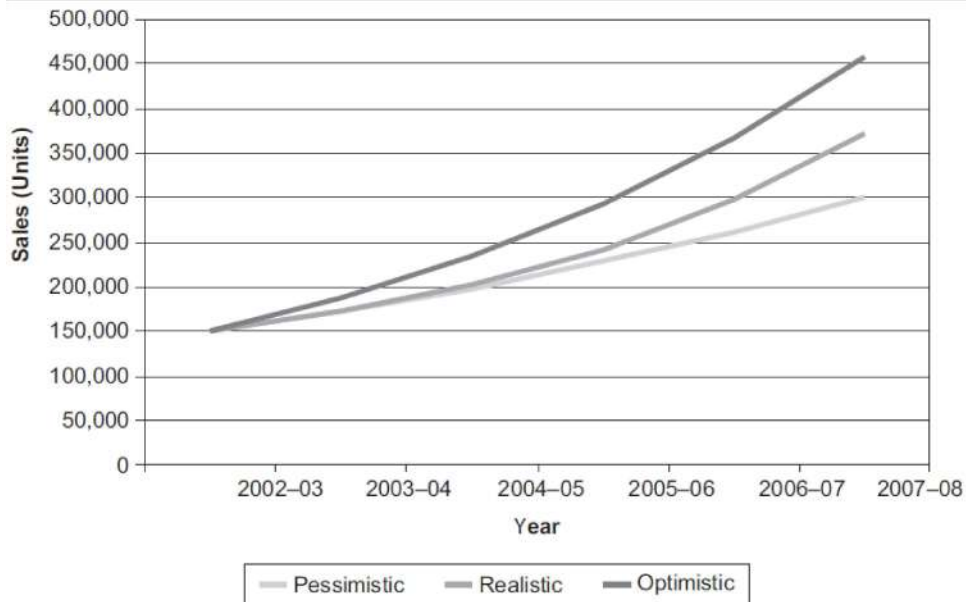

Fig. 7.3(a) *Expected Growth in the Number of Hearing Aids Units to be Sold in North India (Three Perspectives)*
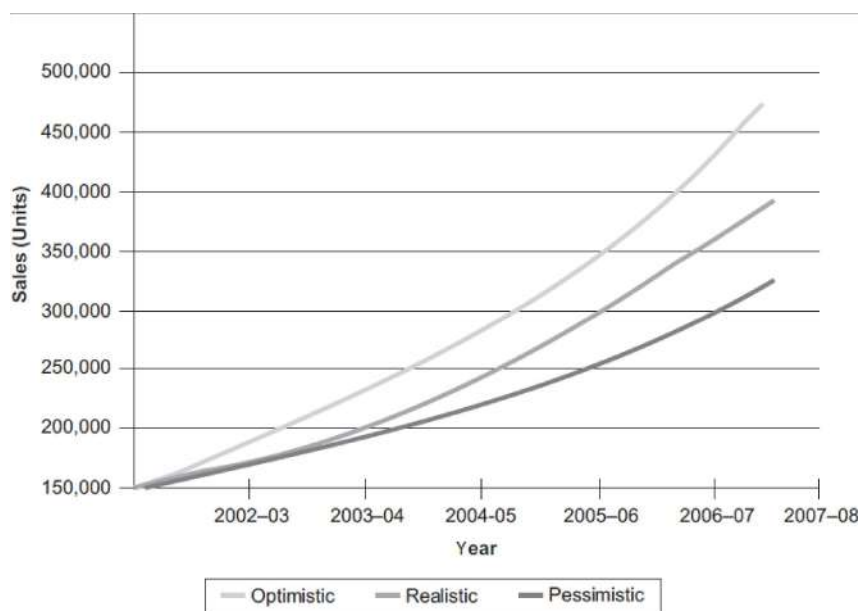
Fig. 7.3(b) *Expected Growth in the Number of Hearing Aids Units to be Sold in North India (Three Perspectives)*

Area or stratum charts: Area charts are like the line charts, usually used to demonstrate changes in a pattern over a period of time. However, here there are multiple lines that are essentially components of the original composite data. What is done is that the change in each of the components is individually shown on the same chart and each of them is stacked one on top of the other. The areas between the various lines indicate the scale or volume of the relevant factors/categories (Figure 7.4).
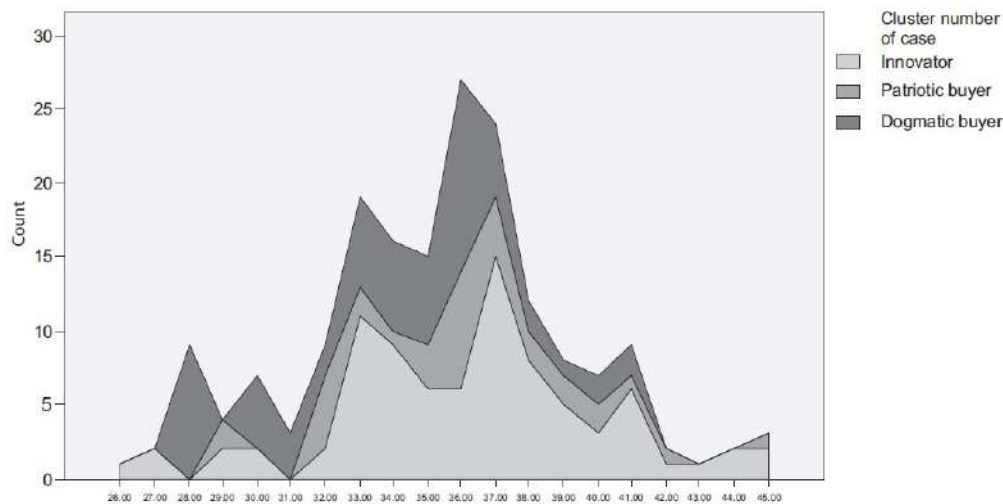
Fig. 7.4 *Perception of Nano by Three Psychographic*
Segments of Two-wheeler Owners

Pie charts: Another way of demonstrating the area or stratum or sectional representation is through the pie charts. The critical difference between a line and pie chart is that the pie chart cannot show changes over time. It simply shows the cross-section of a single time period. The sections or slices of the pie indicate the ratio of that section to the total area of the parameter being displayed. There are certain rules that the researcher should keep in mind while creating pie charts.

- The complete data must be shown as a 100 per cent area of the subject being graphed.

- It is a good idea to have the percentages displayed within or above the pie rather than in the legend as then it is easier to understand the magnitude of the section in comparison to the total. For example, Figure 7.5 shows the brand-wise sales in units for the existing brands of hearing aids in the North Indian market.
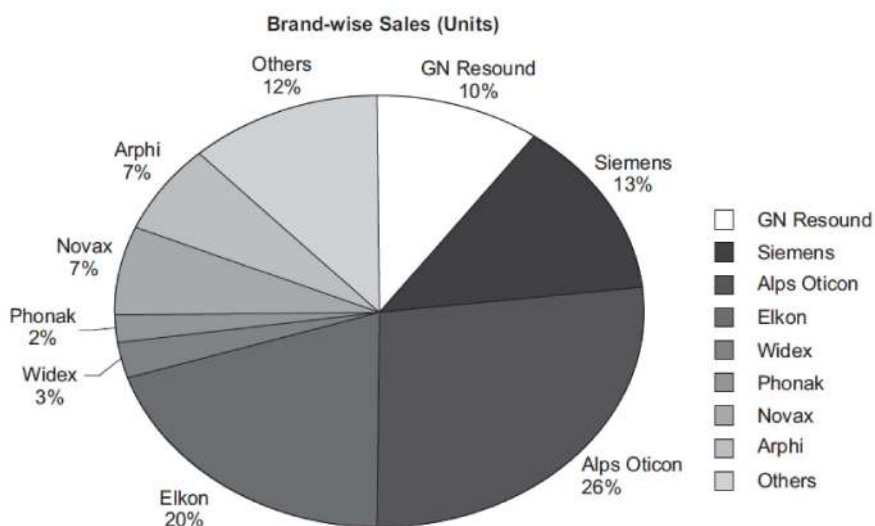
Fig. 7.5 *Brand-wise Sales (units) of Hearing Aids in the*
*North Indian Market (2002–03)*

- Showing changes over time is difficult through a pie chart, as stated earlier. However, the change in the components at different time periods could be demonstrated as in Figure 7.6, showing share of the car market in India in 2009 and the expected market composition of 2015.
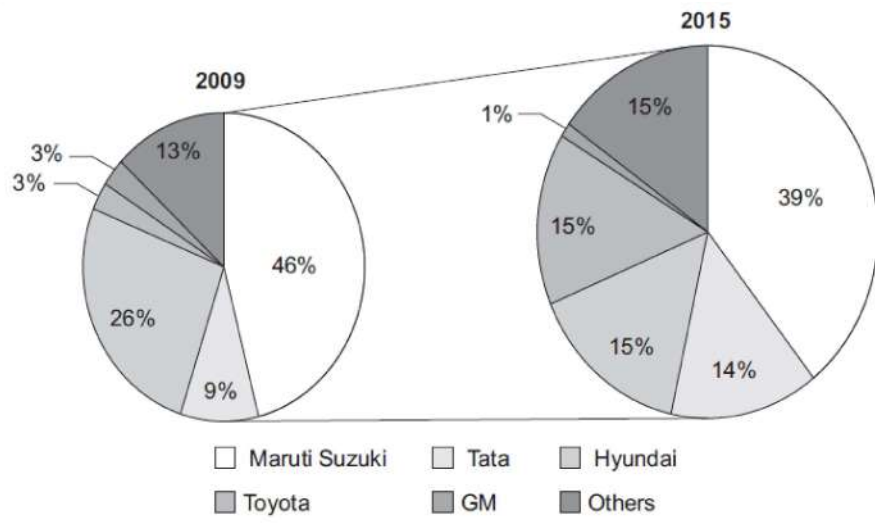

Fig. 7.6 *Current Structure of the Indian Car Market (2009) and the Forecasted Structure for 2015*

Bar charts and histograms: A very useful representation of quantum or magnitude of different objects on the same parameter are bar diagrams. The comparative position of objects becomes very clear. The usual practice is to formulate vertical bars; however, it is possible to use horizontal bars as well if none of the variable is time related [Figure 7.7(a)]. Horizontal bars are especially useful when one is showing both positive and negative patterns on the same graph [Figure 7.7(b)]. These are called bilateral bar charts and are especially useful to highlight the objects or sectors showing a varied pattern on the studied parameter. It is possible to generate bar graphs with relative ease with computer programs today and the distance between the bars can be extremely precise as compared to those created by hand.
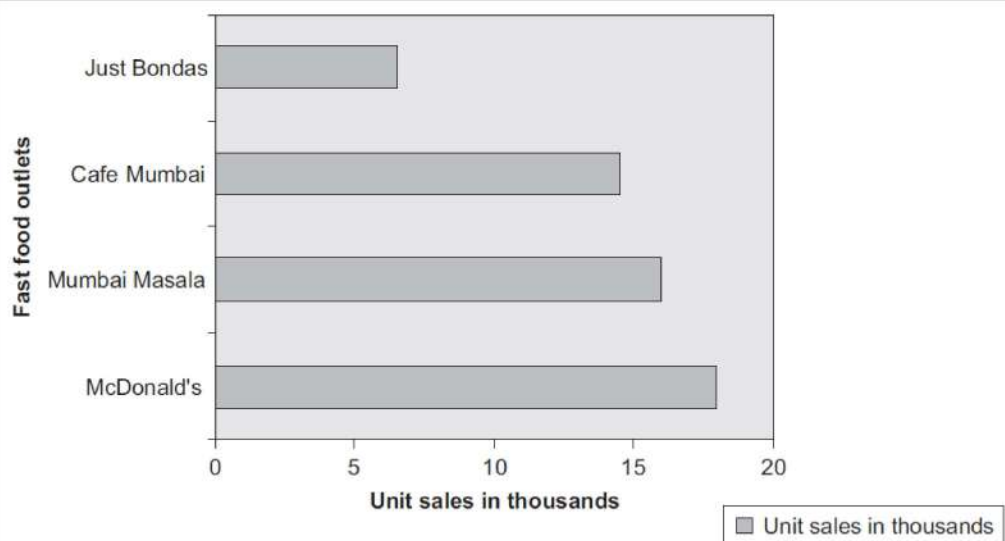

Fig. 7.7(a) *Bar Chart Per day, Unit Sales (Thousands) at Fast Food Outlets in Mumbai*
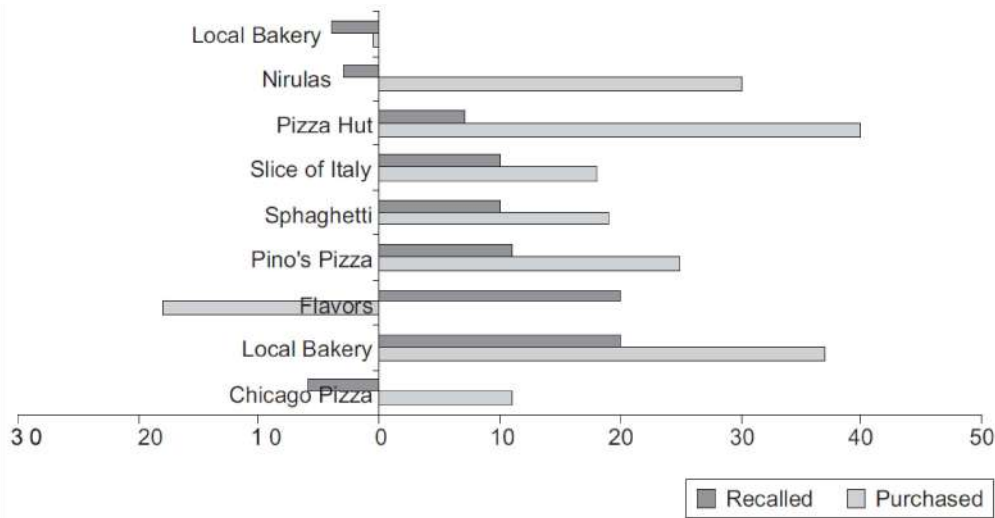
Fig. 7.7(b) *Bilateral Bar Chart—the Brand Recall and Brand*
Purchase Response for Pizza Joints in the NCR

Another variation of the bar chart is the histogram (Figure 7.8) here the bars are vertical and the height of each bar reflects the relative or cumulative frequency of that particular variable.


Std. Dev = 6.24
Mean = 61.2
N = 37.00
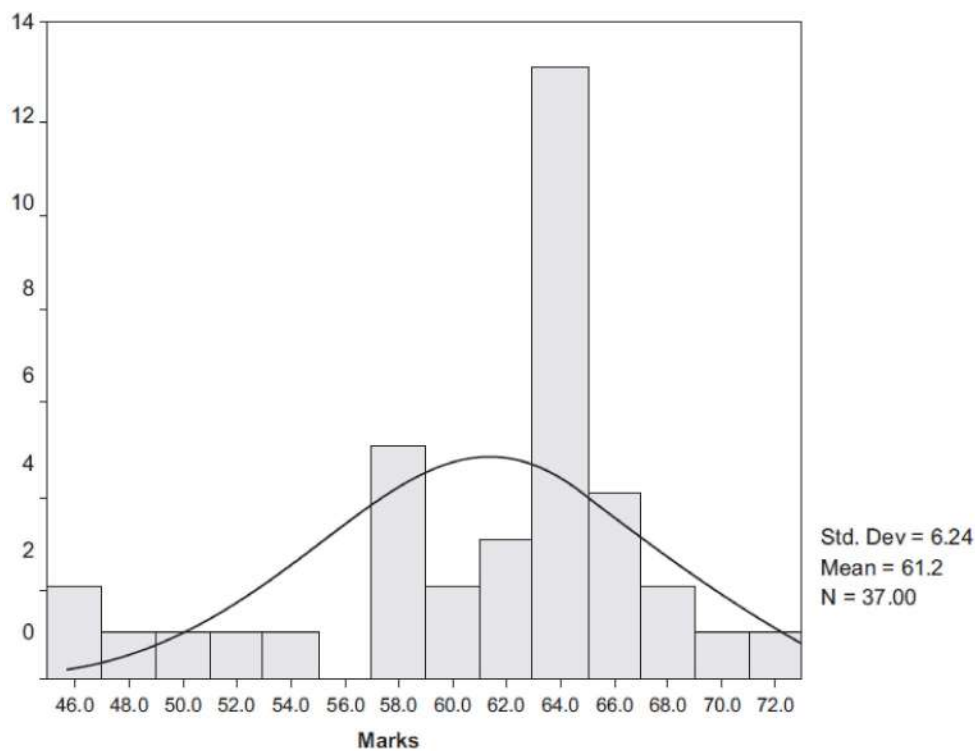
Fig. 7.8 *Histogram (With Normal Curve) Displaying Marks in a*
Course on Research Methods for Management

Pictogram: A pictogram shows graphical representation of data. Pictograms are most often used in popular and general read such as in magazines and newspapers, as they are eye-catching and easy to comprehend by one and all. They are not a very accurate or scientific representation of the actual data and, thus, should be used with caution in an
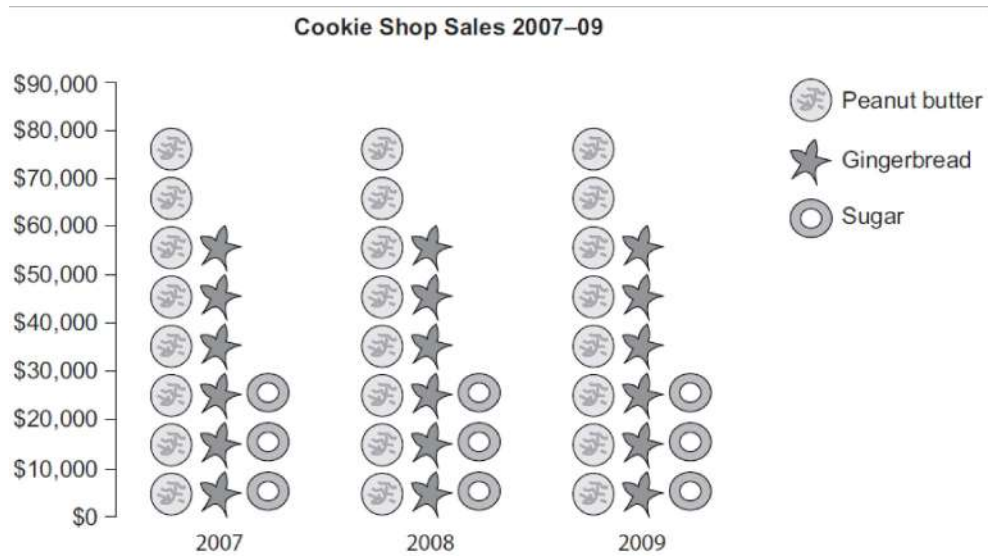
academic or technical report. Examples of pictograph are given in Figures 7.9(a) and 7.9(b).
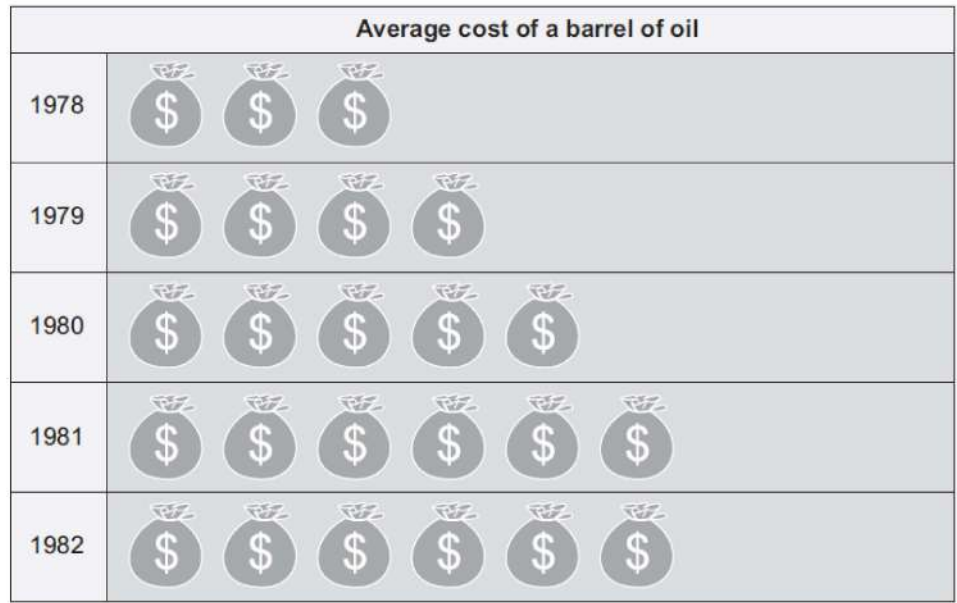
Fig. 7.9(a) *Pictogram Displaying Change in the Cost of Oil Vver a Five-Year Block (1978–1982)*

Source: tutorvista.com

Fig. 7.9(b) *Pictogram Displaying Sales for Cookie Shop over Three Years (2007–09)*

Source: 4spreadsheets.pbworks.com

Geographic representation: Geographic or regional maps related to countries, states, districts, territories can be used as a base to show occurrence of the studied variable in various regions or to show comparative analysis about major brands or industries or minerals. In case of comparative data, the researcher must provide the legend in the displayed map, for example any map of the location may be given.

7.4.2 Research Briefings: Oral Presentation

Once the final draft of the research report is prepared and documented, the last stage is sharing the findings and research implications with the client or interested audience. This is usually done orally and with the support of visual aids. The presentation that the researcher might be making could be detailed for his team members or for an academic audience. However, in case the presentation is for the client or for a business audience, brevity and focus of the presentation is critical. A thumb rule for this is not to go beyond 20 minutes with more time for question and answers and interactive discussion on the findings.

Regardless of the audience for the presentation, the most critical aspect of the presentation is two-fold:

(a) Who is the listener? What does he/she seek from the presentation?

(b) What is the core of the briefing—is it background, or methodology, key findings or decision directions that the findings are indicating?

Once the researcher is clear on this, he needs to need to focus on three key aspects:

Study background: This should be essentially 10–15 per cent of the entire presentation. It should explain the impetus behind the study as briefly and with suitable emphasis as possible.

Study findings: The major conclusions of the study need to be shared in simple words and with appropriate supportive visuals or material. The researcher must be able to demonstrate clearly the link between the study objectives and the findings.

Study implications: In case this was agreed upon between the researcher and the client or was specified as a study objective by the researcher, this section would be the last section of the presentation. The link between what was found and what is suggested must be clear to the audience. The researcher may vary the discussion time between the earlier section and this as 45 per cent each or 30–70 or 70–30, depending on the study objective, i.e., more findings or more implication oriented. As supportive material the researcher can make use of:

Handouts: These could be in the form of the primary questionnaire designed for the study or company brochures and other related secondary material. They should be distributed to the audience when the presenter is referring to them.

Slides: These are created today with the help of computer programmes. There are endless possibilities enhancing the material be presented and for engaging the listener. The designing and creation of the material requires considerable skill and care to ensure that the presentation style should be the supportive aid for an effective delivery and not a showcase of the computer graphics that the researcher is well versed with. Too much clutter and a random mix of text and graphics should be avoided. Animation of the data in synchronization with the vocal delivery makes the presentation more forceful.

Chalkboards and flipcharts: These are additional visual aids that could be kept as standby for the question-and-answer session when an idea might have to be highlighted or demonstrated in the response of some query raised by the listeners. However, use of these means during an active presentation should be avoided as they necessitate the presenter to be engaged with the medium at the cost of losing contact with the listener.

Video and audio tapes: Again, these are supportive materials that can be used to emphasize a point.

The world has become smaller as a consequence of technological innovations that make dissemination of knowledge seem like child's play. Thus, the significance of communication and presentation of this learning cannot be overemphasized.

7.5 SUMMARY

- On completion of the research study and after obtaining the research results, the real skill of the researcher lies in terms of analysing and interpreting the findings and linking them with the propositions formulated in the form of research hypotheses at the beginning of the study.

- The following is the significance of report: it fulfills the historical task of serving as a concrete proof, it ideally presents a comprehensive and objective study of the research problem, it presents a knowledge based on the topic under study, it documents all the steps followed and the analysis carried out.

- Research can be divided into brief reports, detailed reports, technical reports, and business reports.

- Whatever the type of report, the reporting and dissemination of the study and its findings require a structured format and by and large, the process is standardized.

- The following are the steps involved in the report formulation and presentation: a preliminary section with rudimentary parts, the background section with the scope, objectives, background etc; the methodology section and the conclusions.

- The preliminary pages include the letter of transmittal, letter of authorization, table of contents, executive summary and acknowledgements.

- The main text of the report includes problem definition, study background, study scope and objectives and the methodology of research.

- The end text of the report includes appendices, bibliography, footnote and glossary of terms.

- The features of a good research study: clear report mandate, clearly designed methodology, clear representation of findings, representativeness of study finding, etc.

- Guidelines for effective documentation include pointers related to command over the medium, phrasing protocol, simplicity of approach and report formatting and presentation.

- Guidelines for presenting tabular data includes discussion related to the table identification details, data arrays, measurement unit, space, leaders and rulings, assumptions, details and comments, data sources and special mention. There are also certain guidelines related to graphic representation and the use of different graphs.

- Once the final draft of the research report is prepared and documented, the last stage is sharing the findings and research implications with the client or interested audience. This is usually done orally with the support of visual aids.

7.6   KEY TERMS

- **Preliminary pages:** It is the initial section of the report which should carry a 'title' and a 'date,' followed by the acknowledgements in the form of Preface or Foreword.

- **Main text:** It is the main text of the report which comprises of the complete outline of the research report with all the details.

- **Letter of transmittal:** It is the letter that goes alongside the formalized copy of the final report containing the purpose behind the study.

- **Executive summary:** It includes the summary of the report starting from the scope and objectives of the study to the methodology employed and the results obtained in a brief and concise manner.

7.7   ANSWERS TO 'CHECK YOUR PROGRESS'

1. Effective recording and communicating of the results of the study is the extremely critical step of the research process.

2. Working papers or basic reports are written for the purpose of collating the process carried out in terms of scope and framework of the study, the methodology followed and instrument designed.

3. The following are the three issues addressed by the letter of transmittal: it indicates the term of the study or objectives, next it goes on to broadly give an indication of the process carried out to conduct the study and the implications of the findings.

4. An executive summary is the summary of the entire report, starting from the scope and objectives of the study to the methodology employed and the results obtained, have to be presented in a brief and concise manner.

5. The following sections specify the methodology of research: research framework, sampling design, data collection methods, data analysis and study results and findings.

6. Some examples of the primary data which are included in the appendices are: original questionnaire, discussion guides, formula used for the study, sample details, original data, and long tables and graphs which can be described in statement form in the text.

7. The research problem statement is prepared on the assumption that the writer at no point in time needs to be physically present in order to clarify the research mandate.

8. The advantage of presenting information in a tabular form is that the statistical tables present the data in a concise and numeral form, which makes quantitative analysis and comparisons easier.

9. The critical difference between a line and pie chart is that the pie chart cannot show changes over time. It simply shows the cross-section of a single time period. The sections or slices of the pie indicate the ratio of that section to the total area of the parameter being displayed.

10. The thumb rule to be followed to ensure the brevity and focus of the presentation is for the presentation to not go beyond 20 minutes with more time for question and answers and interactive discussion on the findings.

7.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is the significance of report writing?
2. List the classification of data included in the title page of a report.
3. Briefly explain the different types of graphs and their uses.
4. Explain the research briefings for oral presentation.
5. What is the difference between geographic representations and pictograms?

Long-Answer Questions

1. Explain the different types of report.
2. What are the guidelines for effective report writing? Illustrate with suitable examples.
3. Discuss the concept of methodology of report mentioned in the main text of the report.
4. Critically examine the interpretations of results and suggested recommendations
5. Discuss the features of the report writing.
6. 'Visual representations of results are best understood by a reader; thus special care must be taken for this formulation.' Examine the truth of this statement by giving suitable examples.

7.9 FURTHER READING

Kothari, C.R. 2009. *Research Methodology: Methods and Techniques*. New Delhi: New Age International Pvt. Ltd. Publishers.

Chawla, Deepak and Neena Sondhi. 2016. *Research Methodology Concepts and Cases*, 2nd edition. New Delhi: Vikas Publishing House Pvt. Ltd.

Bajpai, Naval. 2011. *Business Research Methods*. 1st edition. New Jersey: Pearson Education.

David, Matthew and Carole Sutton. 2004. *Social Research: The Basics*. United States (California): SAGE Publications Ltd.

Krishnaswami, O R. 2013. *Methodology of Research in Social Sciences*. New Delhi: Himalaya Publishing.