RESEARCH METHODOLOGY

M A [Political Science]

902 C

Third Semester

[ENGLISH EDITION]

Directorate of Distance Education

TRIPURA UNIVERSITY

Reviewer

Dr Hari Shankar Shyam

Assistant Prof., School of Business Studies, Sharda University, Greater Noida

Author: Lalit Kumar Sharma, Assistant Professor (Marketing), Institute of Technology and Science, Ghaziabad

Books are developed, printed and published on behalf of Directorate of Distance Education, Tripura University by Vikas Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 7361, Ravindra Mansion, Ram Nagar, New Delhi - 110 055
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Research Methodology

Syllabi	Mapping in Book
Unit I **Basics of Research:** Meaning and Aims, Steps in Research Process, Research Design, Methods and Types of Research **Survey and Research:** Kinds of Survey, Stages in Survey Method **Hypothesis:** Meaning, Criteria for Hypothesis Construction, Nature of Hypothesis, Difference between Proposition, A Hypothesis and a Theory	**Unit-1:** Research: An Overview **(Pages 3-46)**
Unit - II **Sampling:** Probability and Non-Probability Sampling, Their Applications **Data Collection:** Types of Data-Primary and Secondary Data, Qualitative and Quantitative Data, Talcott Parsons (on Social System and Pattern Variables)	**Unit-2:** Sampling and Data Collection **(Pages 47-69)**
Unit - III Measures of Central Tendency, Measures of Dispersion, Correlation	**Unit-3:** Data Processing, Representation and Analysis **(Pages 71-122)**
Unit - IV Regression and Prediction, Testing Hypothesis, Basic Concepts of Computer	**Unit-4:** Measurement and Scaling Techniques, Regression and Forecasting **(Pages 123-173)**

CONTENTS

INTRODUCTION

NOTES

Research is the quest for knowledge or a systematic investigation in order to establish facts. It helps to solve problems and increase knowledge. The basic aim of research is to discover, interpret and develop methods and systems to advance human knowledge on diverse scientific matters. Thus, research is a process of enquiry and investigation. It helps to solve problems and increase knowledge. One of the main purposes of research is to review the existing knowledge and provide solutions to problems. There are different types of research, such as exploratory, descriptive, experimental and analytical. Exploratory research is done when few or no previous studies of the subject exist. Descriptive research is used to classify and identify the characteristics of a subject. Experimental research suggests or explains why or how something happens. Analytical research suggests or explains why or how something happens. Thus, one of the primary aims of research is to explain new phenomena and generate new knowledge.

Before conducting any research, a specific approach should be decided upon, called research methodology. Research methodology refers to the way research can be conducted. It is also known as the process of collecting data for various research projects. 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 the research report.

Research methodology is a very important function in today's business environment. There are many new trends in research methodology through which an organization can function in this dynamic environment. There are two basic types of research approaches, namely quantitative and qualitative. The main emphasis of quantitative research is on collecting numerical data. It also concentrates on measuring the scale, range and frequency of a phenomenon. Qualitative research is more subjective in nature than quantitative research and involves analysis of data. Quantitative research involves examining the tangible aspects of research, such as values, attitudes and perceptions.

A researcher, before beginning the research, must first formulate an effective research design. Research design is a systematic plan for collecting and utilizing data so that the desired information can be obtained with sufficient accuracy. Therefore, research design is the means of obtaining reliable, objective and generalized data. A few important factors in research methodology include the validity and reliability of research data and the level of ethics. A job is considered half done if the data analysis is conducted improperly. Formulation of appropriate research questions and sampling probable or non-probable factors are followed by measurement using survey and scaling techniques. This is followed by research design that may be experimental. A research design is a systematic plan for collecting

NOTES

and utilizing data so that the desired information can be obtained with sufficient accuracy. Therefore, research design is the means of obtaining reliable, objective and generalized data. Research methodology is a very important function in today's business environment. There are many new trends in research methodology through which an organization can function in this dynamic environment.

In research, formulating a research report is very important for an organization. Engineers, scientists and managers write research reports in order to communicate the results of research, fieldwork or other activities. Most often, a research report is the only concrete evidence of your research and the quality of the research may be judged directly by the quality of writing and how well you convey the importance of your findings to the management of an organization. Most research reports contain the same major sections, although the names of the sections vary widely and sometimes it is appropriate to omit certain sections or add others. If you are submitting a research report to an organization, check for specific requirements and guidelines before beginning to write your report.

This book, *Research Methodology*, is divided into four units and has been written in a simple and easy-to-understand manner. In accordance with the self-instructional mode or SIM format, each unit begins with an 'Introduction' to the topic and is followed by an outline of the 'Unit Objective'. The detailed content is then presented in a simple and structured form, interspersed with 'Check Your Progress' questions to test the student's understanding. A 'Summary' of the content, along with a list of 'Key Terms' and a set of 'Questions and Exercises' is provided at the end of each unit for effective recapitulation.

UNIT 1 RESEARCH: AN OVERVIEW

NOTES

Structure

1.0 INTRODUCTION

Human knowledge, as it exists today, broadly consists of facts and theories. New facts, new concepts and new ways of doing things increase their quantum with the passage of time. This knowledge enables us to understand, comprehend, explain, control, predict, or cope with a given situation. The sources from which we obtain knowledge range from those that are highly reliable to those that are either less reliable or completely unreliable. The knowledge obtained from the less reliable or completely unreliable sources is based on assumptions, beliefs, and untested generalizations. Such generalizations are usually accepted on faith, tradition, or authority and no effort is made to verify their validity. In contrast, reliable knowledge is based on objective verification of generalizations.

The acquisition and expansion of knowledge is not an automatic and self-perpetuating process. On the contrary, it requires constant and planned effort by intelligent and highly trained people or researchers. The present level of knowledge is an outcome of the various methods, such as research method adopted by man over a period of several centuries. Research may be defined as the application of

the scientific method in the study of problems. In this unit, you will get acquainted with the meaning, nature, types and scope of research; the kinds of survey and the stages in the survey method; and the meaning, criteria and nature of hypothesis.

NOTES

1.1 UNIT OBJECTIVES

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

- Discuss the meaning, aims, types and methods of research
- Describe the steps in the research process and the concept of a research design
- Identify the types of survey and the stages in the survey method
- Discuss the meaning and nature of hypothesis, and the criteria of hypothesis construction
- Differentiate between proposition, hypothesis and theory

1.2 BASICS OF RESEARCH

This section will deal with the meaning, aims and steps in research; research design; and the methods and types of research.

1.2.1 Meaning and Aims of Research

Research is a systematic approach to a purposeful investigation. In the words of renowned researcher Clifford Woody, research involves defining and redefining problems; formulating suggested solutions or hypotheses; collecting, evaluating and organizing data, reaching conclusions and making deductions and carefully testing the conclusions to find out if they fit the formulating hypothesis or not.

Research means a systematic and objective study to find facts which can be answers to questions and solutions to problems. Research explains unexplained phenomenon to clarify all doubts and correct misconceived facts. Research unfolds many facts to society and business organizations, which contribute to their progress.

The facts can be searched with the help of the following methods:

- **Arbitrary (or unscientific) method:** This method is very useful for getting answers to question related to imagination, opinion, blind faith or impression. This method is useful for the issues for which data or facts are not available. It is also very useful when analysing the reasons for which no logic is established.
- **Scientific method:** This method is a systematic, logical approach for getting facts. It overcomes the weaknesses of the arbitrary method. It is objective and brings result on the basis of verifiable evidences. Under this method all

available data for the purpose of research is evaluated to check its reliability and efficiency.

Therefore, scientific method is more reliable than the arbitrary method.

NOTES

Research and Scientific Method

Research is scientific by nature as it involves many scientific methods. The meaning of scientific method is 'gaining knowledge which is acceptable universally'. It does not refer to any particular body of knowledge.

The scientific method is, thus based on certain basic postulates. These are as follows:

- **Reliance on empirical evidence:** Truth is based on evidence. Confusion is drawn only when there is no evidence. The solution to a problem is not decided by imagination.
- **Use of relevant concepts:** A lot of facts are experienced by us with the help of our senses. Concepts with specific meanings are used in order to deal with facts.
- **Commitment of objectivity:** It is committed to only objective considerations. These considerations are based on facts and are unbiased.
- **Ethical neutrality:** Science cannot be manipulated to favour the researcher's interest. According to Schrödinger, 'science never imposes anything, science states.' Science focusses on truth based on the facts and makes logical statements about its objects.
- **Generalization:** Science can be generalized as it can be applied to any phenomenon and the results will be the same irrespective of the given conditions.
- **Verifiability:** The conclusion drawn by a researcher or scientist should be verifiable. He should explain the process of arriving at the conclusion for critical examination or scrutiny.
- **Logical reasoning:** It focuses on formulating the most general axioms or what can be termed as scientific theories.

Characteristics and Purpose of Research

The following are the characteristics of research:

- It is a systematic and critical investigation into a phenomenon.
- It uses scientific methods.
- It is objective and logical.
- It requires empirical evidence.
- It focusses on finding facts to questions and solution to problems.

NOTES

The following points will help in understanding the purpose of research:

- Research helps in extending the knowledge of human beings, the environment and natural phenomenon to others.
- It brings the information which is not developed fully during ordinary course of life.
- It verifies the existing facts and identifies the changes in these existing facts.
- It helps in developing facts for critical evaluation.
- It analyses the interrelationship between variables and derives causal explanations.
- It develops new tools and techniques for those who study unknown phenomenon.
- It helps in planning and development.

1.2.2 Types of Research

Although research is a vast subject and is difficult to categorize, it can be classified according to its intent or as per the methods of study.

On the basis of intent, research can be classified as follows:

- **Pure research:** It is done only for the sake of knowledge. The intention is not to apply it in regular practice. Pure research is also called basic or fundamental research. It is not focussed on specific problems, but instead it focusses on the extension of knowledge. New theory or refinements of an existing theory are developed with the help of pure research. It lays the foundation for applied research. It helps in finding the critical factors in a problem. It helps in generating alternative solutions and choosing the best one amongst them.
- **Applied research:** When real-life problems require some solution and decision-making, applied research is carried out. This means that applied research is problem oriented and action directed. It brings immediate and practical results; for example, marketing research carried on for identifying customer habits to purchase something. Though it is problem oriented and action directed it can contribute to the development of theoretical knowledge by leading to the discovery of new facts.
- **Exploratory research:** It is also called formulative research. When a researcher has no knowledge or little knowledge about an unfamiliar problem, they do a preliminary study. The objective of this research is to generate new ideas, gather new facts, precise formulation of problem and increasing familiarity of the researcher to the unfamiliar problem. Katz conceptualizes two levels of exploratory research. At the first level is the discovery of significant variables in particular situations; at the second, the discovery of relationship between variables.

NOTES

- **Descriptive research:** In this research, facts are analysed in detail for clear understanding. This research is simple in nature and in its application. It is more specific than exploratory research. It focusses on the problem under study and also aims at a classification of the range of elements comprising the subject matter of study. Empirical observations are used to conceptualize the problems and facts. It highlights methods of data collection and interpretation.
- **Diagnostic research:** It is just like descriptive research but with a different focus. It is aimed towards in depth approaches to reach the basic causal relations of a problem and possible solutions for it. Prior knowledge of the problem is required for this type of research. Problem formulation, defining the population correctly for study purposes, proper methods for collecting accurate information, correct measurement of variables, statistical analysis and tests of significance are essential in diagnostic research.

The classification of research can be done as per methods of study in the following manner:

- **Fundamental:** This type of research is mainly concerned with identifying certain important principles in a specific field. It intends to find out information that has a broad base of application. Examples of fundamental research are John Robinson's imperfect competition theory in Economics and Maslow's hierarchy of needs theory of motivation, etc.
- **Applied:** This type of research aims at finding a solution to an immediate problem, faced by a society or an industrial organization. It is supposed to discover a solution to some basic practical problems. Applied research suggests corrective methods to minimize a social or business problem.
- **Historical:** Historical research studies the social effects of the past that may have given rise to current situations, i.e., past incidents are used to analyse the present as well as the future conditions. The study of the current state of Indian labour based on past labour union movements in the Indian economy to formulate the Indian Labour Policy is an example of this type of research.
- **Formulative or exploratory:** It helps examine a problem with suitable hypothesis. This research, on social science, is mainly significant for clarifying concepts and innovations for further researches. The researchers are mainly concerned with the principles of developing hypothesis and testing with statistical tools.
- **Experimental:** The experimental type of research enables a person to calculate the findings, employ the statistical and mathematical devices and measure the results thus quantified.
- **Ex post facto:** This type of research is the same as experimental research, which is conducted to deal with the situations that occur in or around an

NOTES

organization. Examples of such a research are market failure of an organization's product being researched later and research into the causes for a landslide in the country.

- **Case study:** This method undertakes intensive research that requires thorough study of a particular chapter.

Research approaches

Quantitative approach and qualitative approach are the two main approaches for data collection. When data is quantified for analysis, the quantitative approach is the best approach which includes collection of data for quantitative analysis. Subjective assessment of attitudes, behaviour and opinion related study requires a qualitative approach. The researcher's impressions are very important for study in this approach. The results achieved in qualitative research are in the form of non-quantitative measures. It can also be in a form which cannot be used for quantitative analysis. Techniques like depth interviews, focus group interviews and projective techniques are used for data collection.

1.2.3 Steps in the Research Process

Research is a systematic, objective and scientific study done to collect the research data related to current problems. Research enables a company to exploit the opportunities available in the environment. If the research has not been planned systematically, it is difficult for a firm to achieve the desired objectives. The research process can be described as follows:

1. Defining the Problem and Objectives

The defined objectives should be SMART.

S – Specific

M – Measurable

A – Attainable

R – Realistic

T – Time bound

The first step in research is definition of a problem. Selection of a problem is itself a difficult decision. The success of research depends on right selection of the problem. If the problem has not been identified in right manner, it is very difficult for the researcher to find the right solution to the issue.

The following sources can help a researcher identify the research problems:

- **Brainstorming:** A researcher can learn new dimensions of a problem by discussing ideas, thoughts, facts and data with other people who have knowledge of the subject.
- **Consultations:** By consulting others, the researcher identifies new dimensions of a problem.

NOTES

- **Daily experience:** Daily experience develops the evaluative thinking in a researcher.
- **Academic experience:** Academic experience helps the researcher develop critical thinking towards the happenings.
- **Field situations:** Research is done because every field today is developing and hence changing constantly.

Objectives of formulating the problem

'A well-defined problem is half solved.' This statement reveals the fact that how important it is to formulate or define a problem. The primary objective of a research is to collect relevant data and analyse this data to get answers to the research problem. This means that the success of research depends upon accuracy of data and information required for investigation. Right formulation solves this purpose. Proper definition of the problem, its analysis, identifying questions for data collection, formulation of hypothesis to be tested are key steps which are required for formulation of the problem. Once the exact and accurate data is known to the researcher, he can plan the other steps without wastage of resources. Thus, right formulation of the problem gives the right direction to the entire research and limits the approach towards pertinent facts out of the large variety of facts. It helps us in determining statistical methods to be used for research.

Criteria for formulation of the problem

Criteria for formulating one problem out of identified problems can be grouped into:

(i) Internal criteria

Internal criteria consist of the following:

- **Interest of the researcher:** The problem should be from the subject of interest of the researcher and can be challenging to him. Without interest in the problem, it becomes very difficult for the researcher to sustain continuity in the research. A researcher's interest depends on his experience, educational background, sensitivity, etc.
- **Own resources of the researcher:** Research requires a lot of money. If the researcher does not have enough money and he is unable to manage external finance, the researcher should not go in for research. Moreover, time resource is more important than money. Research requires more time and hence, it should be utilized properly.
- **Competence of the researcher:** A mere interest in research is not enough. The researcher must be competent enough to plan and carry out a study of the problem. He should have sufficient knowledge of the subject matter, relevant methodology and statistical procedures.

NOTES

(ii) External criteria

External criteria consist of the following:

- **Potential for research:** Very narrow or extremely vague problems should be avoided. In order to be researched, a problem must be one for which observation or other data collection in real world can provide the answer.
- **Importance and urgency:** Issues that require investigation are unlimited but available research efforts are very limited. Therefore, relative importance and significance of the problem is required. Important and urgent issues should be given priority over an unimportant one.
- **Novelty of the problem:** A problem on which a lot of research work has been done should not be considered for research as there are fewer chances of throwing light on any new factor.
- **Feasibility:** Novelty of the problem is not sufficient if it is not feasible to conduct the study on problem in real world, i.e., it should contain facts which can be analysed. Even if the problem is novel we should make a small feasibility study first and proceed only after this if study allows.
- **Facilities:** A well-equipped library, proper guidance in data analysis, etc., are basic facilities which are required to carry on any research.
- **Research personnel:** Availability of adequate research personnel like investigators and research officers is very important for data collection, which is a major issue in many developing countries like India.

Techniques involved in formulating the problem

Defining a research problem properly and clearly is a crucial part of the research study and must, in no case, be done hurriedly. The technique for this purpose involves undertaking the following steps, generally one after the other:

- **Statement of the problem in a general way:** The problem should be carefully worded. The problem statement should indicate nature of the problem and intention of researcher.
- **Understanding the nature of the problem:** The best way to understand the nature of the problem is to discuss with those who have prior experience in the same kind of research. This will ensure that the origination of problem and the objectives in view are correct. If the marketer has stated the problem himself, he should consider all the facts that induced him to make a general statement concerning the problem.
- **Developing ideas through discussions:** Many new ideas are developed by discussing them with others. This discussion provides useful information for research. Discussion is done with those people who have enough experience in the concerned field.
- **Rephrasing the research problem:** After going through the given four steps, the researcher gets a clear idea about the environment in which the

problem is to be studied. Now rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing, the researcher puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypothesis.

2. Developing Data Source

Finding the answers to questions for a research study is called data collection. Data is the collection of facts and other relevant materials, from which we can draw conclusions. Data source is developed for ensuring the availability of data for decision-making.

3. Data Collection

The following are the various methods of data collection:

(a) Interviewing

Interviews are generally classified into the following two categories:

(i) Structured or directive interview

In this type of interview, the investigator goes to the respondent with a detailed schedule. Some questions are asked from all the respondents in the same sequence. The style of asking every question is the same in all the interviews as it generates more reliability. It is used for large sample size.

Merits

The following are the advantages of structured or directive interviews:

- Data can be easily compared.
- Recording and coding of data is done easily.
- There is no wastage of time.

Demerits

The following are the limitations of structured or directive interviews:

- There is no natural conversation.
- Researcher's bias may be reflected in the answers.
- There is no scope for exploring the data.

(ii) Unstructured or non-directive interview

The respondent is encouraged to give his honest opinion on the given topic without or with minimum help from others. In this interview, the pre-planned schedule is not used. The investigator uses only major guidelines for the interview. No particular sequence of questions is followed in the interview. When the path of investigation in exploratory research is not clearly defined, unstructured interview is very useful to get the direction of research study.

NOTES

Merits

The following are the advantages of unstructured or non-directive interviews:

- There is natural conversation.
- Researcher's biasness is not reflected in the answers.
- There is a lot of scope for exploring the data.

Demerits

The following are the limitations of unstructured or non-directive interviews:

- It is not easy to compare the data.
- The suitability of the data for research study needs to be checked.
- There is excessive unproductive conversation.
- There excessive of time consumption.

(b) Observation

Observation may be defined as specific viewing with the purpose of gathering the data for a specific research study. Observation is a classical method of scientific study. It has high importance in any research study as it is an effective method for data collection. Following are the characteristics of the observation method of data collection:

- **Physical and mental activity:** Eyes observe many things in our surroundings but our focus or attention is only on data which is relevant to the research study.
- **Observation is selective:** It is very difficult for a researcher to observe everything in the surroundings. He only observes the data which is useful for his research study and meets the scope of the study. The researcher ignores all the data which is not relevant to the study.
- **Observation is purposive and not casual:** Observation is purposive as it is relevant to a particular study. The purpose of observation is to collect data for the research study. It focuses on human behaviour, which occurs in a social setup. It analyses the relationship of different variables in a specific context.
- **Accuracy and standardization:** Observation of pertinent data should be accurate and standardized for its applications.

Different concepts define the classification of observations.

With respect to the investigator's role, observation may be classified as:

- Participant observation
- Non-participant observation

With respect to observation, it can be classified into:

- Direct observation
- Indirect observation

With reference to the rigour of system adopted, observation can be classified into:

- Controlled observation
- Uncontrolled observation

NOTES

Prerequisites of observation

The following are some of the prerequisites of observation:

- Conditions of observation must provide accurate results. The observer should be in a position to observe the object clearly.
- The right number of respondents as sample size is essential to ensure that observation produces the desired results.
- There should be accurate and completed recording of an event.
- If it is possible, two separate observers and sets of instruments can be used in all or some observations. Then result can be compared to measure accuracy and completeness.

Advantages of observation

The following are the advantages of observation:

- It ensures that behaviour is studied in accordance with the occurrence of events. The observer does not ask anything from the representatives. He just watches what the sample does and says.
- The data collected by observation defines the observed phenomenon as they occur in their natural settings.
- When object itself is not able to define the meaning of its behaviour, observation is best method for analysis; for instance, animals, birds, children, etc.
- Observation covers the entire happenings of an event.
- Observation is less biased than questioning.
- It is easy to conduct disguised observation studies than disguised questioning.
- Use of mechanical devices can generate more accuracy and completeness.

Demerits of observation

The following are the limitations of observation:

- Past studies and events are of no use to observation as the researcher has to personally go through narrations, people and the related documents.
- Difficult to understand attitudes with the help of observation.

NOTES

- Observations cannot be performed by the choice of observer. He has to wait for an event to occur.
- It is difficult to predict when and where the event will occur. Thus, it may not be possible for the observer to reach every event.
- Observation requires more time and more money.

(c) Sampling procedure

Different types of sampling are also used to collect data. The term 'universe' refers to the complete population that is to be studied or measured. A part of the population is called sample. Selecting a part of the universe with a view to draw conclusions about the 'universe' or 'population' for a study is known as sampling. Researchers use sampling for saving time and cost, as the selected sample represents the whole population. The following are the different types of sampling used:

- Probability or random sampling
- Simple random sampling
- Stratified random sampling
- Systematic random sampling
- Non-probability or non-random sampling
- Purposive or (judgment) sampling

4. Data Analysis

Processing of data refers to the preparation of data for research analysis. On the basis of the results of this data processing, further selection of the tools for analysis would be done. Data processing is an intermediary stage of work between data collection and data interpretation. The data gathered in the form of questionnaires, interview, schedules, field notes and data sheets, is mostly in the form of large volume of research variables.

The following are some of the methods for data analysis:

Spearman's Rank Correlation Method

Charles Edward Spearman, a British psychologist, devised a method for measuring the correlation between two variables based on ranks given to the observations. This method is adopted when the variables are not capable of quantitative measurements like intelligence, beauty, etc., and in such cases, they are variables. It is in such cases that rank correlation is useful.

Concurrent Deviation Method

In this method, correlation is calculated between the direction of deviations and not their magnitudes. As such, only the direction of deviations is taken into account in the calculation of this coefficient and their magnitude is ignored.

5. Presenting the Findings

Format of the research report

NOTES

Every reader who is reading a research report should be made aware of the research study so that they can obtain the scientific knowledge and can judge the adequacy of its methods and thus can develop an opinion of how the findings are to be taken. For this purpose we need a proper format of the report. Format of a report refers to the contents of a research report. A format of the research report should comprise of the following:

- **Preliminary pages:** The preliminary pages of a report contain the following:
  - o The title
  - o Date
  - o Acknowledgement in the form of 'preface' or 'foreword'
  - o Table of contents
  - o List of tables and illustrations to help the readers find the information they require in the report
- **Main text:** A broad outline of the main content of the report is given in the main text. The main text of the research report contains the following sections:
  - o Introduction
  - o Statement of findings and recommendation
  - o The results
  - o Implication drawn from the results
  - o Summary

Final presentation of research report

Certain specific considerations must be kept in mind while making a written presentation. These are as follows:

- **Size and physical design:** The research report must be written on unruled paper of 8.5" × 11" size. Black or blue ink should be used if the report is to be written by hand. At least a 1 ½" margin on the left and 1" on the right side of the paper should be given. The page should also have 1" margins on the top and the bottom. In case the research report needs to be typed, then all typing should be double spaced on one side of the page only, except for the insertion of the long quotations.
- **Procedure:** The following are the steps in the preparation of a research report:
  - o **Format/layout:** According to the nature and objective of the research, the layout of the report should be decided and followed in a proper manner.
  - o **Quotations:** Quotations should be punctuated with quotation marks and double spaces. However, if a quotation is too lengthy, it should be

NOTES

single-spaced and indented at least half an inch to the right of the normal text margin.

- **Footnotes:** Footnotes are meant for cross-references. They are placed at the bottom of the page, separated from the textual material by a space of 1/2" and a line that is around 1 1/2" long. Footnotes are always typed in single space, though they are divided from one another by double space.
- **Documentation style:** While writing the first footnote reference for any given work, we should make sure that it is complete in its documentation, and contains all the necessary facts about the edition used. It is not compulsory to give the details of the references that follow. If the work is cited again without any other work intervening, it may be indicated as *ibid*, followed by a comma and the page number.
- **Punctuation and abbreviations in footnotes:** A researcher must ensure that correct punctuation is used while writing the footnotes, to make sure that the text is understood clearly. To remove or reduce repetition, certain English and Latin abbreviations are used in bibliographies and footnotes.
- **Use of statistics, charts and graphs:** The use of statistics is very common in research study, as it provides the scientific learning and understanding to both the researcher and the reader. Use of statistics simplifies the complex issues in research and gives logic to relationships of variables. Generally, statistics are represented as pictograms, line graphs, bars, charts and different types of tables.

- **Final draft:** While writing the final draft of the research report, we should make sure that the language is simple and easy to understand. Also, usage of jargons should be avoided.
- **Bibliography:** A bibliography is a list of writings with the time and place of publication. It is a list of the resources used by the author which a reader can refer to for further reading.
- **Index:** An index acts as a good guide to the reader. It can be prepared both as a subject index and author index giving names of subjects and names of authors, respectively. The names are followed by the page numbers of the report, where they have appeared or been discussed.

1.2.4 Research Design

It is not possible for any researcher to remember all the decisions he has taken. Even if he does remember these, he would have difficulty in understanding how these are interrelated. Therefore, he records his all decision on a paper or record disc by using relevant concepts or symbols. Such symbolic construction can be called the research design. A research design is a systematic, objective and scientific plan developed for directing a research study. It constitutes the overview for data

collection, measurement and analysis of data. Research design is the road map for the functioning of a researcher.

Need for research design

NOTES

There is a need for research design as it ensures a smooth flow of many research operations, thereby making research as efficient as possible, producing maximum information with minimum effort, time and cost. The ideal design is concerned with specifying the optimum research procedure that could be followed where there are no practical restrictions. To manage with the future changes, a researcher must have a flexible research design. This flexibility ensures the desired achievements in a research. A research design tells the researcher about the methodologies adopted for research work.

Features of a good research design

The following are the features of a good research design:

- Ensuring research progress in the right direction
- Minimizing time and cost of research
- Encouraging coordination and effective organization
- Minimizing bias and maximizing the reliability of the data collected and analysed

Types of Research Designs

Research designs can be categorized as:

- Research design in case of exploratory research studies
- Research design in case of descriptive and diagnostic research studies
- Research design in case of hypothesis-testing research studies

1. Research Design in Case of Exploratory Research Design

Formulative research is another term used for exploratory research. The main objective of such studies is problem formation with more precision for research and developing research hypothesis to get the results for operations. The key concern in such type of studies is to generate ideas and finding the insights. Thus relevant research designs for this type of studies must be flexible to provide an opportunity for various dimensions of the issues under the study. In such studies:

- The sample size is small.
- Non-probability sampling designs are used.
- Data requirements are vague.
- The objective is general rather than specific.
- No definite recommendations are made as a result of the analysis.

NOTES

The following are the methods of research design for such studies:

- **Survey concerning literature:** This is one of the most uncomplicated and easy methods to formulate the problem with more precision for research and developing research hypothesis to get the results for operations. Hypothesis formulated by previous researchers can be assessed and an evaluation of their importance is done for further research. Many a times the work of intellectual researchers provides the framework for formulating hypothesis for operations.
- **Experience survey:** It refers to a survey of the respondents who are familiar with the research problem (to be studied). This means that they have already experienced similar problems in past. The main objective of such a survey is to know the relationship between the variable and new ideas related to research problems. In this survey, it is important to select competent people to share their new ideas about the same problem with the researcher.
- **Researcher's interpretation:** These are fruitful methods for selecting the hypothesis for research. This method is suitable in areas where small experience serves as a guide to research study. The detailed study of choice phenomenon in which the researcher wants to research is required. Investigator's attitude, the concentration of the study and the availability of the investigator to draw together diverse information into a united interpretation are the main features of this method.

2. Research Design in Case of Descriptive and Diagnostic Research Studies

Studies describing the individuality of a particular person or group are called descriptive research, whereas research studies defining the occurrence of any happening or association of one happening with others are called diagnostic research. In such studies:

- The study describes the phenomenon under study.
- The collected data may relate to the demographic or the behaviour variables of the respondents under study.
- The research has got a very specific objective, clear cut data requirements and uses a large sample which is drawn through a probability sampling designs.
- The recommendation/findings in descriptive research are definite.

3. Research Design in Case of Hypothesis Testing Research Studies

Hypothesis testing research studies (also known as experimental studies) are the research studies where the hypothesis is tested to define the causal relationship between variables in an operation.

Principles of Experimental Design

The three principles enumerated by Prof. Fisher for experimental design are as follows:

NOTES

(i) Principle of replication

According to the principle of replication, the same experiment is repeated more than once. Every time the same experiment is repeated in different experimental units instead of one. By doing so, the numerical precision of the experiments is improved. For instance, let us consider that we have to examine the two ranges of pulse. For this rationale we divide the entire field into two parts and cultivate one range in one part and the other range in the other part. By comparing the yield of two parts, we can get results for comparative analysis. To apply the principle of replication to this trial, firstly we divide the entire field into several parts; cultivate one range in half of these parts and other range in remaining parts. By collecting the statistics yield of two ranges, we can draw conclusion by comparing the same. Therefore, results are more reliable when we are applying the principle of replication in comparison to the results attained without applying the principle. The more we repeat the experiment, the better the results that we get.

(ii) Principle of randomization

The principle of randomization provides protection against the effects of extraneous factors by randomization when we conduct an experiment. The principle of randomization indicates the need for a design or plans the experiment in such a way that variations caused by extraneous factors can be united under the universal course of chance.

(iii) Principle of local control

Under this principle, the extraneous factors, the identified basis of variability, is made to vary intentionally over as wide a range as necessary and this needs to be done in such a way that the variability it causes can be measured and hence eliminated from the experimental error. This means that we should plan the trial in a manner that we can perform a two-way analysis of variance in which the total variability of data is divided into three components attributed to treatments, the extraneous factors and experimental error.

Formal Experimental Designs

The following are the various kinds of formal experimental designs:

- **Completely Randomized Design (CR design):** It involves only two principles, viz., the principle of replication and randomization. The CR design is used when the experimental areas are homogenous for study.
- **Randomized Block Design (RB design):** It is an improvisation of the CR design. Along with the other two principles, local control can be applied in the RB design.

NOTES

- **Latin Square Design (LS design):** For agriculture-related researches, the LS design is used. The Latin square design is used where the researcher desires to control the variation in an experiment that is related to rows and columns in the field.
- **Factorial Design:** Factorial design is used for studies where more factors show more than one effect.

CHECK YOUR PROGRESS

1. List the two methods used to search facts.
2. What is the first step in the research process?
3. List three methods of data collection.
4. List the three types of research designs.
5. List the three principles of experimental designs enumerated by Prof. Fisher.
6. List any two formal experimental designs.

1.3 SURVEY AND RESEARCH

Survey is an important tool in research. No research can be performed without them. Survey can be defined in various ways. Some of the common definitions of survey are as follows:

- To view with a scrutinizing eye; to examine.
- To inspect, or take a view of; to view with attention, as from a high place; to overlook; as, to stand on a hill, and survey the surrounding country.
- To determine the form, extent, position, etc., of, as a tract of land, a coast, harbour, or the like, by means of linear and angular measurements, and the application of the principles of geometry and trigonometry; as, to survey land or a coast.
- A particular view; an examination, especially an official examination, of all the parts or particulars of a thing, with a design to ascertain the condition, quantity, or quality; as, a survey of the stores of a ship; a survey of roads and bridges; a survey of buildings.
- To examine with reference to condition, situation, value, etc.; to examine and ascertain the state of; as, to survey a building in order to determine its value and exposure to loss by fire.

1.3.1 Types of Survey

There are basically two types of surveys:

(i) Descriptive: These surveys generally collect information on what people think and do.

(ii) Analytic: These surveys are generally used to either test hypotheses or to answer particular research questions.

While the most common method of collecting survey data is the 'questionnaire', the means by which you gather the information that goes into the survey responses may vary. If the survey makes use of a questionnaire, the measuring instruments must have demonstrable reliability and validity, especially with regard to sampling, questioning and mode of questioning.

Some examples of collecting survey data include self-administered posted questionnaires, web-based forms, telephone question and answer interviews or face-to-face interviews. There are advantages and disadvantages of each approach, primarily to do with sample size and open versus closed questions. In order to make a judgement, the key areas for consideration include the cost, co-ordination, size of the sample, rate of return, nature and quality of the data obtained and the ability to clarify questions or responses. The success of using surveys depends strongly on the design of appropriate body of questions and the skill of the interviewer.

Other important methods of collecting survey data are Interview, Observation and Case Study. All these methods have been discussed in detail later.

1.3.2 Stages in Survey Method

Surveys go through the following seven stages:

(i) Planning and designing the survey

In this stage, you must define the goals and objectives of your survey. You should write down the outline of your research and also establish a budget for the project. You are also required to plan your schedule, define the population and estimate the required sample size. The method of data collection and the method for determination of the results should also be decided at this stage. Finally, you must write down the questions and design and pretest the questionnaire.

(ii) Collecting data

In this stage, first you have to decide on the survey method that will suit your research needs. There is no best method of collection of data and you must gather the required data keeping your resources in mind. You must also decide what steps to take in case sufficient data is not collected from the respondents.

(iii) Accessing data

The only purpose of this stage is to transfer the data into the analytical software for processing it further.

(iv) Preparing and managing data

The main aim of this stage is to get the data ready for analysis. In this step, you are required to formulate a 'codebook'. This codebook must include variable names,

NOTES

variable formats and descriptive variable labels. You should also set up multiple item indices and scales, i.e., multiple variables that have exactly the same answer set. In this step, you should transform your data, which will help to get the data in the form and structure required for analysis. Also, the missing data values should be replaced with estimates so that better summary statistics are obtained.

(v) Analysing data

In this stage, you take out all the useful information that you require from the data that you have collected. This helps you make informed decisions.

(vi) Reporting

After analysing the data, the results need to be reported. The main aim of reporting is to produce results from the data analysis which can be easily understood by others, who can use this information.

(vii) Deployment

You must tailor your results according to the needs of the target audience. This will ensure the effectiveness of the results.

1.3.3 Research Methods

This section discusses the various research methods such as interview, schedule, questionnaire, and case study.

1. Interview

Interviewing is a very effective method of data collection. It is a systematic and objective conversation between an investigator and respondent for collecting relevant data for a specific research study. Along with conversation, learning about the gestures, facial expressions and environmental conditions of a respondent are also very important. Generally, interview collects a wide range of data from factual demographic data to highly personal and intimate information relating to a person's opinions, attitudes, values and beliefs, past experience and future intentions. The interview method is very important in the collection of data from the respondent who is less educated or illiterate. Personal interview is more feasible when the area covered for survey is compact. Probing is a very important part of an interview.

Types of Interview

The following are the various types of interviews:

(i) Structured or directive interview

In this type of interview, the investigator goes to the respondent with a detailed schedule. Some questions in same sequence are asked from all respondents.

(ii) Unstructured or non-directive interview

In this type of interview, the respondent is encouraged to give his honest opinion on a given topic without or with minimum help from others.

NOTES

(iii) Focused interview

This is a semi-structured interview where the respondent shares the effect of the experience to the given conditions with the researcher or investigator. It is conducted with those respondents only who have prior experience of conditions given by the investigator. Analysis of the attitude, emotional feelings for the situations under study is main purpose behind conducting these interviews. A set of fix questions may not be required in this interview but a relevant topic is required which is known to the respondent.

(iv) Clinical interview

While a focussed interview is concerned with effects of specific experience, clinical interviews are concerned with broad underlying feelings or motivations or the course of the individual's life experiences with reference to the research study. It encourages the interviewee to share his experience freely.

(v) Depth interview

To analyse or study the respondent's emotions, opinions, etc., depth interviews are conducted. This kind of interview aims to collect intensive data about individuals, especially their opinions. It is a lengthy process to get unbiased data from the respondent. Interviewers should avoid advising or showing this agreement. Instead, the investigator has to motivate the respondent to answer the questions.

Features of Interviews

The following are some of the features of interviews:

- The interviewer and the respondent are the participants in any interview. They both are unknown to each other and so it is important for an interviewer to introduce himself first to the respondent.
- An interview has a beginning and a termination point in the relationship between the participants.
- Interview is not a mere casual conversational exchange. It has a specific purpose of collecting data which is relevant to the study.
- Interview is a mode of obtaining a verbal response to questions to put verbally. It is not always face to face.
- Success of interview depends on the interviewer and respondent and how they perceive each other.
- It is not a standardized process.

Essentials for an Effective Interview

The following are the requirements for a successful interview:

- **Data availability:** The respondent should have complete knowledge of the information required for specific study.

NOTES

- **Role perception:** The interviewer and the respondent should be aware of their roles in the interview process. The respondent should be clear about the topic or questions which have to be answered by him. Similarly, it is the responsibility of the interviewer to make the respondent comfortable by introducing himself first. The investigator should not affect the interview situation through subjective attitude and argumentation.
- **Respondent's motivation:** The respondent can hesitate to answer the questions. In this case, the approach and skills of the interviewer are very important as he has to motivate the respondent to answer or express himself.

Advantages and Disadvantages of Interviews

The following are the advantages of the interview method:

- In-depth and detailed information is collected.
- The interviewer tries to improve the responses and quality of data received. He can control the conditions in favour of the research study.
- Interviews help in gathering supplementary information which can be helpful to the study.
- Interviews use special scoring devices, visuals and materials to improve the quality of data or information collected.
- Interviews use observation and probing by the interviewer to see the accuracy and dependability of given data by the respondent.
- Interviews are flexible in nature.

The following are the disadvantages of interviews:

- Interviews consume more time and cost.
- The respondent's responses can be affected by the way the interviewer asks the questions.
- The respondent may refuse to answer some personal questions which are relevant to the study.
- Recording and coding of data during the interview process may sometimes be difficult for the interviewer.
- The interviewer may not have good communication or interactive skills.

Interview Process

The following are the stages in an interview process:

(i) Preparation

The interviewer needs to make certain preparations to make an interview successful. The interviewer should keep all the copies of the schedule or guide ready. They need to prepare the lists of respondents with their addresses, contact number and meeting time. They should prepare themselves with all the approaches and skills

required to conduct an interview. They should prepare themselves to face all adverse situations during the interview. If the interviewer is not doing such planning, they can fail to collect the right information from respondent.

NOTES

(ii) Introduction

The interviewer is not known to the respondent. Therefore, the interviewer must introduce himself first to every respondent. In the introduction, the interviewer should tell about himself, his organization details and the purpose of his visit. If the interviewer knows someone who the respondent is familiar with, then he can use that person's reference to make the respondent more comfortable. The following are some steps which help in motivating the respondent:

- The interviewer should introduce himself with a smiling face and always greet the respondent.
- He should identify and call the respondent by name.
- He must describe how the respondent is selected.
- He should explain the purpose and usefulness of the study.
- He should focus on the value of the respondent's cooperation.

(iii) Developing report

It is important for an interviewer to develop a rapport with the respondent before starting the interview. By doing this, a cordial relationship is established between them. It helps the interviewer understand the inherent nature of the respondent which helps in building a rapport and the discussion can be started with some general topic or with the help of a person who is commonly known to both of them.

(iv) Carrying the interview forward

After establishing a rapport, the skills of the interviewer are required to carry the interview forward. The following are some guidelines which should be followed:

- Start the interview in an informal and natural manner.
- Ask all the questions in the same sequence as in the schedule.
- Do not take an answer for granted. It is not necessary that an interviewee will know all answers or will give all answers. The interviewer has to create interest for answering questions.
- The objective of the question should be known to the interviewer to ensure that the correct information is collected for research study.
- Explain the question if it has not been understood properly by the respondent.
- Listen to the respondent carefully with patience.
- Never argue with the respondent.
- Show your concern and interest in the information given by the respondent.

NOTES

- Do not express your own opinion for answers of any question in the schedule.
- Continue to motivate the respondent.
- If the respondent is unable to frame the right answer, the interviewer should help him by providing alternate questions.
- Ensure that the conversation does not go off track.
- If the respondent is unable to answer a particular question due to some reasons, drop the question at that moment. This question can be asked indirectly later on.

(v) Recording the interview

Responses should be recorded in the same sequence as they are given by the respondent. The response should be recorded at the same time as it is generated. It may be very difficult to remember all the responses later for recording them. Recording can be done in writing but there may be some problems if the writing skills of an interviewer are not good. Hence, use of electronic devices like tape recorders can help in this purpose. The interviewer should also record all his probes and other comments on the schedule, but they should be in brackets to ensure that they are set off from response.

(vi) Closing the interview

After the interview is over, the interviewer must thank the respondent for his cooperation. He must collect all the papers before leaving the respondent. If the respondent wants to know the result of the survey, the interviewer must ensure that the results are mailed to him when they are ready.

(vii) Editing

At the end, the interviewer must edit the schedule to check that all the questions have been asked and recorded. Also, abbreviations in recording should be replaced by full words.

Problems Faced in an Interview

The following are some of the main problems faced in an interview:

(i) Inadequate response

Kahn and Cannel laid down five principal symptoms of inadequate response. They are given as follows:

- **Partial response** in which the respondent gives a relevant but incomplete answer.
- **Non-response** in which the respondent remains silent or refuses to answer the questions.
- **Irrelevant response** in which the respondent's answer is not relevant to the question asked.

- **Inaccurate response** in which the reply is biased.
- **Verbalized response problem** which arises because of the respondent's failure to understand the question.

NOTES

(ii) Interviewer's biasness, refusal, incapability to understand questions, etc.

An interviewer can affect the performance of an interview with his own responses and suggestions. Such biasing factors can never be overcome fully, but their effect can be reduced by training and development techniques.

(iii) Non response

Some respondents out of the total respondents fail to respond to the schedule. The reasons for this non response can be non availability, refusal, incapability to understand questions, etc.

(iv) Non availability

Some respondents are not available at their places at the time of call. This could be because of odd timings or working hours.

(v) Refusal

Some respondents refuse to answer the questions. There can be many reasons for this, such as language, odd hours, sickness, no interest in such studies, etc.

(vi) Inaccessibility

Some respondents can be inaccessible because of various reasons such as migration, touring job, etc.

Methods and aims of controlling non response

Kish suggests the following methods to reduce either the percentage of non-response or its effects:

1. Improved procedure for collecting data is the most obvious remedy for non-response. The improvements advocated are as follows:
   - Guarantee of anonymity
   - Motivation of the respondent to cooperate
   - Arousing the respondent's interest by clever opening remarks and questions
   - Advance notice to the respondent
2. Call backs are the most effective way of reducing not-at-home responses in personal interviews, as are repeated mailings in no-returns in mail survey.
3. Substitution for non-response is often suggested as a remedy. Usually this is a mistake because the substitutes resemble the responses rather than the non-responses. Nevertheless, beneficial substitution methods can sometimes be designed with references to important characteristics of population.

NOTES

Attempts to reduce the percentage or effect of non-response is aimed at reducing the bias caused by vast differences non respondents and respondents. The response bias should not be confused with the reduction of sample size due to non-response. The latter effect can be easily overcome either by anticipating the size of non-response in designing the sample size or by compensating for it with a supplement. These adjustments increase the size of the response and the sampling precision, but they do not reduce the non-response percentage or bias.

1. Telephonic Interview

Telephonic interview is a non-personal method of data collection. It may be used as a major method or supplementary method of data collection. It is useful in the following conditions:

- When the population is composed of those people who are listed in telephone directories.
- When less number of questions have to be answered by the respondents.
- When the time available for the survey is less.
- When the subject is of the interest to the respondent.
- When the respondents are widely scattered.

Advantages

The following are the advantages of telephonic interviews:

- Less time and low cost
- Good quality of response
- Less demanding on interviewer
- No field work is required
- Easy to contact those respondents who cannot be reached

Disadvantages

Telephonic interviews have the following limitations:

- Restricted to persons who are listed in telephone or other relevant directories
- Not feasible to conduct long interviews
- Limitation of information collected
- No answer to personal questions by respondents
- Respondent's emotions, facial expressions and other environmental factors cannot be recorded
- Difficult to develop rapport

Group Interview

Group interview is the method of collecting primary data from a number of individuals with common interests. In group interviews, the interviewer performs the role of a

discussion leader. Free discussion is encouraged on the same aspects of the subject under the study. Information is collected either through a self-administered questionnaire or through an interview. Samples for the group can be selected from schools, colleges, clubs and other associations.

NOTES

Advantages

The following are the advantages of this technique:

- Respondent gets freedom to express his views
- Flexible method
- Use of visual aids
- Less time consuming as group can be interviewed in the time required for one respondent's interview
- Respondents are more confident in groups
- Eliminates the limitation of individual interviews

Disadvantages

The following are the main disadvantages of group interviews:

- Difficulty in selecting the desired sample group
- Dominance of one individual in a group
- Respondents can be biased or they can try to please the interviewer or others

2. Observations

Observation can be defined as viewing or seeing. Observation means specific viewing with the purpose of gathering the data for a specific research study. Observation is a classical method of scientific study. It is very important in any research study as it is an effective method for data collection.

Characteristics of Observation Method

The following are the characteristics of observation method of data collection:

- **Physical and mental activity:** Eyes observe so many things in our surroundings but our focus or attention is only on data which is relevant to research study.
- **Observation is selective:** It is very difficult for a researcher to observe everything in his surroundings. He only observes the data which is purposive for his research study and meets with the scope of his study. The researcher ignores all the data which is not relevant to the study.
- **Observation is purposive and not casual:** Observation is purposive as it is relevant to a particular study. The purpose of observation is to collect data for the research study. It focusses on human behaviour which occurs in

NOTES

a social phenomenon. It analyses the relationship of different variables in a specific context.

- **Accuracy and standardization:** Observation of pertinent data should be accurate and standardized for its applications.

Types of Observation

Different concepts define the classification of observations.

With respect to an investigator's role, observation may be:

- Participant observation
- Non-participant observation

With respect to the method of observation, it can be classified into the following:

- Direct observation
- Indirect observation

With reference to the control on the system to be observed, observation can be classified into the following:

- Controlled observation
- Uncontrolled observation

(i) Participant observation

In participant type of observation, the observer is an active participant of the group or process. He participates as well as observes as a part of phenomenon; for example, to study the behaviour of management students towards studying and understanding marketing management, the observer or researcher has to participate in the discussion with students without telling them about the observation or purpose. When respondents are unaware of observations, then only their natural interest can be studied.

Advantages

The following are the main advantages of participant observation:

- In-depth understanding of the respondent group
- Context which is meaningful to observed behaviour can be recorded or documented by the researcher

Disadvantages

The following are the demerits of participant observation:

- If a participant is at lower level in hierarchy of group, his participation may be less.
- Emotions of the observer may result in loss of objectivity.

NOTES

(ii) Non-participant observation

In non-participant observation, the observer does not participate in the group process. He acknowledges the behaviour of the group without telling the respondents. It requires a lot of skills to record observations in an unnoticeable manner.

(iii) Direct observation

In direct observation, the observer and researcher personally observe all the happenings of a process or an event when the event is happening. In this method, the observer records all the relevant aspects of an event which are necessary for study. He is free to change the locations and focus of the observation. One major limitation of the method is that the observer may not be able to cover all relevant events when they are happening.

(iv) Indirect observation

Physical presence of an observer is not required and recording is done with the help of mechanical, photographic or electronic devices; for example, close circuit TV (CCTV) cameras are used in many showrooms to observe the behaviour of customers. It provides a permanent record for an analysis of different aspects of the event.

(v) Controlled observation

All observations are done under pre-specified conditions over extrinsic and intrinsic variables by adopting experimental design and systematically recording observations. Controlled observations are carried out either in the laboratory or the field.

(vi) Uncontrolled observation

There is no control over extrinsic and intrinsic variables. It is mainly used for descriptive research. Participant observation is a typical uncontrolled one.

Prerequisites of Observation

The following are the prerequisites of observation:

- The conditions of observation must provide accurate results. An observer should be in a position to observe the object clearly.
- The right number of respondents should be selected as the sample size for the observation to produce the desired results.
- Accurate and complete recording of an event.
- If it is possible, two separate observers and sets of instruments can be used in all or some observations. Then the result can be compared to measure accuracy and completeness.

NOTES

Advantages and Limitations of Observation

The following are the advantages of observations:

- It ensures the study of behaviour in accordance with the occurrence of events. The observer does not ask anything from the representatives, he just watches the doing and saying of the sample.
- The data collected by observation defines the observed phenomenon as they occur in their natural settings.
- When an object is not able to define the meaning of its behaviour, observation is best method for analysis; for example, animals, birds and children.
- Observation covers the entire happenings of an event.
- Observation is less biased as compared to questioning.
- It is easier to conduct disguised observation studies as opposed to disguised questioning.
- The use of mechanical devices can generate accurate results.

The following are the limitations of observation:

- Past studies and events are of no use to observation. For these events and study, one has to go through narrations, people and documents.
- It is difficult to understand attitudes with the help of observation.
- Observations cannot be performed by the choice of the observer. He has to wait for an event to occur.
- It is difficult to predict when and where the event will occur. Thus, it may not be possible for an observer to reach in every event.
- Observation requires more time and money.

Use of observation in business research

Observation is very useful in the following business research purposes:

- Buying behaviour of customer, lifestyles, customs, interpersonal relations, group dynamics, leadership styles, managerial style and actions.
- Physical characteristics of inanimate things like houses, factories, stores, etc.
- Movements in a production plant.
- Flow of traffic, crowd and parking on road.

3. Schedule and Questionnaire

Primary data can be collected with the help of mails and surveys. The respondents receive the questionnaires from the researcher and are asked to fill them completely and return them to the researcher. It can be performed only when the respondents are educated. The mail questionnaire should be simple and easy to understand, so

NOTES

that the respondents can answer all questions easily. In mail questionnaires, all the answers have to be given and recorded by the respondents and not by the researcher or investigator, as in the case of personal interview method. There is no face-to-face interaction between the investigator and respondent and so the respondent is free to give answers of his own choice.

Importance of questionnaires

A questionnaire is a very effective method as well as research tool in any research study. It ensures the collection of a diversified and wide range of scientific data to complete the research objectives. The questionnaire provides all the inputs in the form of relevant data to all statistical methods used in a research study.

Types of Questionnaire

The following are the various categories of questionnaires:

(i) Structured or standard questionnaire

Structured or standard questionnaires contain predefined questions in order to collect the required data for research study. These questions are the same for all the respondents. Questions are in the same language and in the same order for all the respondents.

(ii) Unstructured questionnaire

In unstructured questionnaires, the respondent has the freedom to answer all the questions in his own frame of reference and in his own terms.

Process of Data Collection

The researcher prepares the mailing list by collecting the addresses of all the respondents with the help of primary and secondary sources of data. A covering letter must accompany every questionnaire, indicating the purpose and importance of the research and importance of cooperation of the respondent for the success of the research study.

Alternate modes of sending questionnaire

The following are the alternate modes of distributing questionnaire to respondents:

(i) Personal delivery

The researcher or investigator himself delivers the questionnaire to the respondents and requests them to fill it within a specific duration, i.e., one day or two days, as per the convenience of the researcher. After the given duration, they collect the questionnaire from the respondents. This added the advantage of personal interview and mail survey. Alternatively, the questionnaire can be delivered personally to the respondents and the respondents return the questionnaire by mail to the researcher.

NOTES

(ii) Attaching questionnaire to a product

When a firm is launching a new product or wants to collect the feedback on old products, the firm attaches a questionnaire with its product and requests the customers to fill the questionnaire. The company can give some discount or gift to the respondent of every return questionnaire.

(iii) Advertising the questionnaire

The questionnaire is advertised in magazines and newspapers with instructions to complete it. After filling the questionnaire from the magazine or newspaper, the respondents mail it to the advertiser.

(iv) Newsstand inserts

In this method of sending questionnaires to the respondents, the questionnaire, along with covering letter and a self-addressed reply-paid envelope, is inserted into a random sample of newsstand copies of a newspaper or magazine.

Improving the Response in a Mail Survey

Generally, the response rate in mail surveys in countries like India is very low. The following techniques can be adopted to increase the rate of response:

- **Covering letter:** The covering letter should be in a language which generates the interest of the respondent. It should address the respondent by name.
- **Quality printing:** Sometimes the quality of the printed questionnaire is so bad that the respondent faces a lot of problems in reading it. This results in loss of interest and so, the quality of printing should be excellent and attractive.
- **Prior information:** Prior information can be given to the concerned respondent by telephone, e-mail, newsletters, etc. Such steps bring more success than follow-ups.
- **Incentives:** Monetary and non-monetary incentives can be given to respondents who are filling questionnaire. This generates a higher response.
- **Follow-ups:** The respondent can be approached with the help of an investigator to collect the questionnaire or to solve the problems faced by respondent in filling the questionnaire.
- **Larger sample size:** We should always select a sample size which is larger than what is actually required. This will help the researcher in getting answers from the effective sample size.

Advantages and Disadvantages of Questionnaires

The following are the advantages of questionnaires:

- Low cost
- Wide reach and extensive coverage
- Easy to contact the person who is busy

NOTES

- Respondent's convenience in completion of questionnaire
- More impersonal, provides more anonymity
- No interviewer's biasness
- Accuracy

The following are the disadvantages of questionnaires:

- Low response by respondent
- Low scope in many societies where literary level is low
- More time requirement

Preparation of an Effective Questionnaire

While preparing a questionnaire, the researcher must focus on some key parameters to prepare it. These key parameters are as follows:

- Proper use of open and close probe
- Proper sequence of questions
- Use of simple language
- Asking no personal question in which the respondent is hesitating to answer
- Should not be time consuming
- Use of control questions indicating reliability of the respondent

Collecting Data through Schedule

This method is very similar to the collection of data through questionnaires. The only difference is that in schedule, enumerators are appointed. These enumerators go to the respondents, ask the stated questions in the same sequence as the schedule and record the reply of respondents. Schedules may be given to the respondents and the enumerators should help them solve the problems faced while answering the question in the given schedule. Thus, enumerator selection is very important in data collection through schedules.

Distinction between schedule and questionnaire

Both questionnaire and schedule are popular methods of data collection. The following are the main differences between questionnaire and schedule:

- A questionnaire is generally sent to the respondents through mail, but in case of schedule, it is sent through enumerators.
- Questionnaires are relatively cheaper mediums of data collection as compared to schedules. In the case of questionnaires, the cost is incurred in preparing it and mailing it to respondent, while in schedule, more money is required for hiring enumerators, training them and incurring their field expenses.

NOTES

- The response rate in questionnaires is low as many people return it without filling. On the other hand, the response rate in schedules is high because they are filled by enumerators.
- In collecting data through questionnaires, the identity of the respondent may not be known, but this is not the case when it comes to schedules.
- Data collection through questionnaires requires a lot of time, which is comparatively very less in case of schedules.
- Generally, there is no personal contact in case of questionnaires, but in schedules, personal contact is always there.
- The literacy level of the respondent is very important while filling questionnaires, but in schedules, the literacy level of the respondent is not a major concern as the responses have to be recorded by enumerators.
- Wider distribution of questionnaires is possible but this is difficult with schedules.
- There is less accuracy and completeness of responses in questionnaires as compared to schedules.
- The success of questionnaires depends on the quality of questions but success of a schedule depends on the enumerators.
- The physical appearance of questionnaire matters a lot, which is less important in case of schedules.
- Observation method cannot be used along with questionnaires but it can be used along with schedule.

4. Case Study Method

We explore and analyse the life of a social chapter or entity, whether it be a family, a person, an institution or a community, with the help of a case study. The purpose of case study method is to identify the factors and reasons that account for particular behaviour patterns of a sample chapter and its association with other social or environmental factors. Generally social researchers use case study method to understand the complex social phenomenon and to identify the factors related to this phenomenon. Case study provides the clues and ideas to a researcher for further research study. By adopting case study method, a researcher gets to know about happenings in the past, which could be related to the research studies and analyse the problem with better perspectives.

Assumptions of case study method

The assumptions made in a case study method are as follows:

- Case study depends on the imagination of the investigator who is analysing the case study. The investigator makes up his procedure as he goes along.
- History related to the case is complete and as coherent as it could be.

NOTES

- It is advisable to supplement the case data by observational, statistical and historical data, since these provide standards for assessing the reliability and consistency of the case material.
- Efforts should be made to ascertain the reliability of life history data by examining the internal consistency of the material.
- A judicious combination of techniques of data collection is a prerequisite for securing data that is culturally meaningful and scientifically significant.

Advantages and Disadvantages of Case Study Method

Case study ensures several advantages to the researcher for his research work. Key advantages of the case study method are as follows:

- Provides the basis for understanding complex social phenomenon and all related factors affecting the social phenomenon.
- Provides clues and ideas for exploratory research. When the researcher is not able to get a fair idea about the research, past happenings mentioned in a case study help the researcher get clues and ideas.
- Case study helps in generating objectives for exploratory research.
- It suggests the new courses of inquiry.
- Case study helps in formulating research hypothesis.

Some important disadvantages of case study method are as follows:

- **Reliability:** Data collected through case study may not be reliable or it can be difficult to verify the reliability of data in the current scenario.
- **Adequacy:** Data collected through case studies may not be adequate for research work as data is not pertinent to the research conditions.
- **Representative:** Data presented by case studies represents the happenings with unknown circumstances to a researcher. Hence, it cannot be the true representation of events to a researcher.

Making Case Study Effective

The criteria for evaluating the adequacy of case history is of central importance for case study. John Dollard has proposed seven criteria for evaluating such adequacy. They are as follows:

(i) The subject must be viewed as a specimen in the cultural series, i.e., the case drawn out from its total context for the purpose of study must be considered as a member of the particular cultural group or community. The scrutiny of life histories of people must be done with a view to identify community values, standards and their shared way of life.

(ii) The organic motto of action must be socially relevant, i.e., the action of individual cases must be viewed as a series of reactions to social stimuli or situations. In other words, the social meaning of behaviour must be taken into consideration.

NOTES

(iii) The strategic role of the family group in transmitting the culture must be recognized, i.e., in case of the individual being the member of a family in shaping his behaviour must never be overlooked.

(iv) The specific method of elaboration of organic material onto social behaviour must be clearly shown, i.e., case history that portrays in detail how basically a biological organism, the man, gradually blossoms forth into a social person, are especially fruitful.

(v) The continuous related character of experience for childhood through adulthood must be stressed. In other words, the life history must be a configuration depicting the inter-relationships between the persons various experiences.

(vi) The social situation must be carefully and continuously specified as a factor. One of the important criteria for the life history is that a person's life must be shown as unfolding itself in the context of and partly owing to specific social situations.

(vii) The life history material itself must be organized according to some conceptual framework. This in turn would facilitate generalizations at a higher level.

Case study as a method of business research

A detailed case study helps the researcher identify the reasons behind business-related problems. As it can be possible that that particular incident has happened in past, so the current issues can be sorted out, by referring to the same case. In-depth analysis of selected cases is of particular value to business research when a complex set of variables may be at work in generating observed results and intensive study is needed to unravel the complexities. The exploratory investigator should have an active curiosity and willingness to deviate from the initial plan, when the finding suggests a new course of enquiry, which might prove more productive. With the help of case study method, the risk can be minimized in any decision-making process.

Check Your Progress

7. What are the two main types of surveys?
8. List the two categories of the types of questionnaires.
9. What are the two alternate modes of sending questionnaires?
10. What is the main purpose of the case study method?

1.4 HYPOTHESIS: MEANING, NEED AND NATURE

According to Theodorson, 'a hypothesis is a tentative statement asserting a relationship between certain facts.' Kerlinger describes it as 'a conjectural statement of the relationship between two or more variables'.

1.4.1 Need and Nature of Hypothesis

NOTES

The following points help in understanding the importance of hypothesis:

- A hypothesis is a proposal intended to explain a fact or an observation.
- A hypothesis specifies the sources of data which shall be studied and in what context they shall be studied.
- It determines the data needs.
- Hypothesis suggests the type of research which is likely to be the most appropriate.
- A hypothesis contributes to the development of the theory.

Nature of hypothesis

Hypothesis is more useful when stated in precise and clearly defined terms. A good hypothesis implies that hypothesis which fulfills its intended purposes and is up to the mark. The following are some important points to be kept in mind:

- A good hypothesis should be stated in the simplest possible terms. It is also called the principle of the economy or business. It should be clear and precise.
- A good hypothesis is in agreement with the observed facts. It should be based on original data derived directly.
- It should be so designed that bits test will provide an answer to the original problem which farms the primary purpose of the investigation.
- Hypothesis should state relationship between variables, if, it happens to be a rational hypothesis.

1.4.2 Criteria for Hypothesis Construction

Once the investigative questions are set up for each of the objectives, the researcher should identify the anticipated or possible answers to the investigative questions. A survey of related theories and earlier studies and discussions with co-scientists will facilitate this process. He, then, should write down those answers as appropriate types of hypotheses—descriptive, relational or causal—as the case may be. He should evaluate these tentative hypotheses in terms of the characteristics of a good hypothesis and refine and record them into logical and testable hypotheses, keeping in mind the rules given below:

Rules for constructing hypotheses

According to Smith, there are certain rules for constructing good hypotheses. These are as follows:

(i) Search for variable measurements with the most quantitative characteristics available: Precise quantitative measurements are more critical in testing theory than qualitative characteristics. Hage gives four techniques to search for

NOTES

and create variables from non-variable concepts. First, the researcher can search for implied dimensions underlying non-variable concepts. Campbell provides a good example of this method in the study of the non-variable concept 'social group'. He identifies four underlying dimensions of degrees of proximity, similarity, perceived common fate, and perceived spatial pattern. Second, one can create new variables by comparing conceptual synonyms or analogies. Price used the synonym technique in his study of organizational measurement concepts like 'participation in decision making', 'organization control', 'power' and 'influence' refer to the degree of organizational centralization. Third, one can search the literature for rarely occurring associations between phenomena. The cognitive dissonance theory in sociology got its start in this manner. Fourth, one can generate new variables through ordering many concepts from more or less abstract extension of the research variables and their applicability in the new order of society, say virtual or network organization.

(ii) Make the variable scale properties explicit by stating all of the variable's mutually exclusive and totally inclusive categories by degrees. For example, a variable like 'income' may be categorized into (1) up to ₹ 5000 per month, (2) ₹ 5001 to 10000, (3) ₹ 10001 to 20000, (4) ₹ 20001 and above.

(iii) Describe the means used to sort observations into your variables categories in sufficient detail so that your methods may be evaluated and replicated by others. 'Personality disintegration' is a good example of a poorly operationalised variable. It is an unreliable measure and cannot be replicated.

(iv) Always consider alternative operations that might be more appropriate for a given variable.

(v) Analyse variables through their relationships. Non-ratio uni- or multi- variable distribution is arbitrary, since it has no intrinsic lower boundary.

(vi) Link two or more formal propositions through a shared independent or dependent variable where possible. For example, from following concrete observations:

(a) 'Married persons are less likely than unmarried persons to commit suicide.'

(b) 'Married persons with children are less likely than married persons without children to commit suicide.'

The following abstract formal hypotheses may be inductively produced:

(i) 'Sucide rates vary directly with the degree of individualism.'

(ii) 'Suicide rates vary indirectly with the degree of group cohesion.'

Types of Statistical Hypothesis

In context of statistical analysis, we generally consider two types of hypothesis:

- Null hypothesis
- Alternative hypothesis

When comparing the superiority of both the methods A and B, if we assume that both the methods are equally good, then the assumption is known as 'null hypothesis'. On the other hand, if we consider method A to be better, it is alternative hypothesis. These may be, symbolically presented as:

Null Hypothesis = H_0

Alternative Hypothesis = H_a

NOTES

1.4.3 Difference between Proposition, Hypothesis and Theory

The difference between a proposition, hypothesis and theory are as follows

(i) A proposition is a logical statement of relationship between two or more variables which has, generally, been confirmed by empirical research. (A proposition should be distinguished from a hypothesis which is a logical statement of an assumed relationship between two or more variables which must be empirically tested, replicated and elaborated before being accepted as confirmed).

(ii) Proposition is a broad statement drawn from a theory, whereas a hypothesis takes this one step further and formulates a more specific statement that is empirically testable. Proposition states a relationship between two concepts and a hypothesis operationalizes this relationship and puts it in an empirically testable form.

(iii) The term hypothesis is used to refer to an explanation of things that occur. In some cases, it may refer to a simple guess. In other instances it may be a well-developed set of propositions that are crafted to explain the detailed workings of some occurrence or occurrences. One definition states specifically that it is the antecedent to a conditional proposition.

(iv) The hypothesis is formed and tested within the scientific process. One may develop the hypothesis while observation is occurring, but that may also be considered premature. The act of observation (outside of experimentation) may actually present opportunity to disprove a hypothesis. The hypothesis though is necessarily well defined and inclusive of details. This allows for accurate testing. It also, in many cases distinguishes it from a theory.

(v) The term theory is one of a rather scientific nature, but of a less limited nature. Some uses can refer to explanations of occurrences; some do include usage as referencing a simple guess. There is more though. Theory is used to refer to a branch of study that is focused on the general and conceptual, as compared to the practical and the applied of the same subject. It is significant that a theory is conjectural in nature.

NOTES

(vi) A hypothesis is a proposed explanation for something. We call it a theory when that hypothesis has been tested with considerable evidence. As a result, a theory is usually a much larger set of statements than a hypothesis because a theory can grow with every new piece of evidence it explains. In other words, a theory can explain far more than the phenomenon it originally was proposed to explain.

(vii) A hypothesis attempts to answer questions by putting forth a plausible explanation that has yet to be rigorously tested. A theory, on the other hand, has already undergone extensive testing by various scientists and is generally accepted as being an accurate explanation of an observation. This doesn't mean the theory is correct; only that current testing has not yet been able to disprove it, and the evidence as it is understood, appears to support it. A theory will often start out as a hypothesis — an educated guess to explain observable phenomenon. The more a hypothesis is tested and holds up, the better accepted it becomes as a theory.

Check Your Progress

11. State the definition of the term 'hypothesis' as given by Theodorson.
12. What is a good hypothesis?
13. List the two types of statistical hypothesis.

1.5 SUMMARY

- Research is a systematic approach to a purposeful investigation. Research is scientific by nature as it involves many scientific methods.
- On the basis of intent, research can be classified as pure research, applied research, exploratory research, descriptive research and diagnostic research.
- As per the methods of study, research can be classified as fundamental, applied, historical, formulative, experimental, ex post facto and case study.
- Quantitative approach and qualitative approach are the two main approaches for data collection.
- Defining the problem and objectives, developing data source, data collection, data analysis and presenting the findings are the steps in the research process.
- A research design is a systematic, objective and scientific plan developed for directing a research study. It constitutes the overview for data collection, measurement and analysis of data.
- The ideal research design is concerned with specifying the optimum research procedure that could be followed where there are no practical restrictions. To manage with the future changes, a researcher must have a flexible research design. This flexibility ensures the desired achievements in a research.

NOTES

- Formulative research is another term used for exploratory research. The main objective of such studies is problem formation with more precision for research and developing research hypothesis to get the results for operations. The key concern in such type of studies is to generate ideas and finding the insights.
- Studies describing the individuality of a particular person or group are called descriptive research, whereas research studies defining the occurrence of any happening or association of one happening with others are called diagnostic research.
- Survey is an important tool in research. No research can be performed without them. Descriptive and analytical are the two types of surveys.
- While the most common method of collecting survey data is the 'questionnaire', the means by which you gather the information that goes into the survey responses may vary.
- Planning and designing the survey, collection of data, accessing data, preparing and managing data, analysing data, reporting and deployment are the various stages of a survey method.
- Telephonic interview is a non-personal method of data collection. It may be used as a major method or supplementary method of data collection.
- Group interview is the method of collecting primary data from a number of individuals with common interests. In group interviews, the interviewer performs the role of a discussion leader. Free discussion is encouraged on the same aspects of the subject under the study. Information is collected either through a self-administered questionnaire or through an interview.
- Observation is a classical method of scientific study. It is very important in any research study as it is an effective method for data collection.
- A questionnaire is a very effective method as well as research tool in any research study. It ensures the collection of a diversified and wide range of scientific data to complete the research objectives. The questionnaire provides all the inputs in the form of relevant data to all statistical methods used in a research study.
- The purpose of case study method is to identify the factors and reasons that account for particular behaviour patterns of a sample chapter and its association with other social or environmental factors. Generally social researchers use case study method to understand the complex social phenomenon and to identify the factors related to this phenomenon. Case study provides the clues and ideas to a researcher for further research study. By adopting case study method, a researcher gets to know about happenings in the past, which could be related to the research studies and analyse the problem with better perspectives.

NOTES

- A hypothesis is a tentative statement asserting a relationship between certain facts.
- A good hypothesis implies that hypothesis which fulfills its intended purposes and is up to the mark.
- Null hypothesis and alternative hypothesis are the two types of hypothesis considered in the context of statistical analysis.
- A proposition is a logical statement of relationship between two or more variables which has, generally, been confirmed by empirical research.
- Proposition is a broad statement drawn from a theory, whereas a hypothesis takes this one step further and formulates a more specific statement that is empirically testable. Proposition states a relationship between two concepts and a hypothesis operationalizes this relationship and puts it in an empirically testable form.

1.6 KEY TERMS

- **Research:** It refers to a systematic and objective study to find facts which can be answers to questions and solutions to problems.
- **Observation:** It refers to specific viewing with the purpose of gathering the data for a specific research study.
- **Sampling:** It refers to selecting a part of the universe with a view to draw conclusions about the 'universe' or 'population' for a study.
- **Research design:** It is a systematic, objective and scientific plan developed for directing a research study.
- **Experience survey:** It is a survey of the respondents who are familiar with the research problem (to be studied).
- **Diagnostic research:** It is a research study defining the occurrence of any happening or association of one happening with others.
- **Descriptive surveys:** They are surveys that generally collect information on what people think and do.
- **Analytic surveys:** They are surveys that are generally used to either test hypotheses or to answer particular research questions.
- **Group interview:** It is a method of collecting primary data from a number of individuals with common interests.
- **Proposition:** It is a logical statement of relationship between two or more variables which has, generally, been confirmed by empirical research.
- **Hypothesis:** It is a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.

1.7 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Arbitrary and scientific method are the two methods used to search for facts.
2. Defining the problem and objectives is the first step in the research process.
3. Interviewing, observation and sampling procedure are the three methods of data collection.
4. Research design in case of exploratory research studies, research design in case of descriptive and diagnostic research studies and research design in case of hypothesis-testing research studies are the three types of research designs.
5. Principle of replication, principle of randomization and principle of local control are the three principles of experimental designs enumerated by Prof. Fisher.
6. LS Design and Factorial Design are two formal experimental designs.
7. Descriptive and analytical surveys are the two main types of surveys.
8. Questionnaires can either be structured or unstructured.
9. Personal delivery and advertising the questionnaire are two alternate modes of sending questionnaires.
10. The purpose of case study method is to identify the factors and reasons that account for particular behaviour patterns of a sample chapter and its association with other social or environmental factors.
11. According to Theodorson, 'a hypothesis is a tentative statement asserting a relationship between certain facts.'
12. A good hypothesis implies that hypothesis which fulfills its intended purposes and is up to the mark.
13. Null hypothesis and alternative hypothesis are the two types of statistical hypothesis.

1.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What do you understand by research?
2. Write a short note on research and the scientific method.
3. State the characteristics of research.
4. What are the different types of research approaches?
5. What is research design?

NOTES

6. Why is research design needed in research?
7. What are the characteristics of a good research design?
8. What are the various types of surveys?
9. List the various features of interviews.
10. Write a short note on the case study method of data collection.
11. Write a short note on the need and nature of hypothesis.
12. What are the various types of hypotheses in the context of statistical analysis?

Long-Answer Questions

1. What is the purpose of research? Explain the various types of research.
2. Explain the research process in detail.
3. Explain the different types of research designs.
4. What are the features of an exploratory research design?
5. How is a research design made in case of descriptive and diagnostic research studies?
6. Explain the various principles of experimental design.
7. Explain the various stages in the survey method.
8. Describe the process of data collection.
9. What are the various types of interviews? Explain the various advantages and limitations of interviews.
10. Discuss the interview process in detail.
11. What is the observation method of data collection? Explain its various types.
12. Explain the various criteria for hypothesis construction.
13. Differentiate between proposition, hypothesis and theory.

1.9 FURTHER READING

Booth, Wayne. 2008. *The Craft of Research*, Third edition. Illinois: University of Chicago Press.

Creswall, John W. 2008. *Research Designs: Quantitative, Qualitative and Mixed Methods Approaches*. London: Sage Publications.

Christenson, Larry B. *et al.* 2010. *Research Methods, Design and Analysis*, Eleventh edition. New Jersey: Allyn and Bacon.

Kothari, C. R. 2008. *Research Methodology: Methods and Techniques*. New Delhi: New Age International.

Kumar, Ranjit. 2010. *Research Methodology: A Step-by-Step Guide for Beginners*, Third edition. New Delhi: Sage Publications.

UNIT 2 SAMPLING AND DATA COLLECTION

NOTES

Structure

2.0 INTRODUCTION

Data collection is essentially an important part of the research process so that the inferences, hypotheses or generalizations tentatively held may be identified as valid, verified as correct, or rejected as untenable. In order to collect the requisite data for any research problem, the researcher has to sample the population concerned, since it is not possible to encompass the entire population, to devise appropriate tools and techniques for measuring the attributes concerned, and to administer these tools on the selected sample or samples for collecting the relevant data.

The main aim of research is to discover principles that have universal application. Generally, research includes all such assumptions that are based on a large number of samples/units/objects. It would be impractical if not impossible to test or observe each unit of population under controlled conditions in order to arrive at principles having universal validity. A 'population' is any group of individuals/units that have one or more characteristics in common which are of interest to the researcher, for a particular research. A 'sample' is a small percentage of the larger group who are selected for research. A sample can be statistically explained as being a subset of a population. The sample will be able to give an idea of the characteristics of the larger group from where it has been drawn. It is possible to make deductions about the larger population on the basis of the sample.

NOTES

For selecting a sample, it is necessary to have a sampling frame. After defining a population and listing all the units, a researcher selects a sample of units from the sampling frame. Sampling design refers to a definite plan for obtaining a sample from the sampling frame. It refers to the technique or procedure, which a researcher adopts in selecting some sampling units from where inferences about population are drawn. An error in statistics is the difference between the value of a statistic and that of the corresponding parameter. These errors arise due to chance differences between the members of population included in the sample and those not included. This unit discusses the concept of research design, sample, methods of sampling, sampling design and sampling errors.

In this unit, you will be taught about sampling and sampling design. You will learn about the various methods of sampling such as probability or random sampling, non-probability or non-random sampling, purposive or judgment sampling, quota sampling and snow-ball sampling. The unit will also discuss the various sources of data. This includes collection of primary and secondary data and qualitative and quantitative data. You will finally learn about Talcott Parson's opinions and contribution on social system and pattern variables.

2.1 UNIT OBJECTIVES

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

- Define sampling and sampling design
- Discuss the various criteria related to choice of sampling procedure
- Describe the various methods of sampling
- Analyse the various sources of data—primary, secondary, qualitative and quantitative
- Assess Talcott Parson's opinions and contribution on social system and pattern variables

2.2 SAMPLING AND SAMPLING DESIGN

A part of the population is called sample. Selecting a part of the 'universe' with a view to draw conclusions about the 'universe' or 'population' for a study is known as sampling. A researcher uses sampling for saving time and costs as a selected sample is a replica of the population.

Sampling design: Census and sample survey

All items in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry, when all the items are covered, no element of chance is left and highest accuracy is obtained. In practice, this may not be true.

NOTES

Even the slightest element of bias in such an enquiry will get larger and larger as the number of observations increase. Moreover, there is no way of checking the element of bias or its extent, except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Therefore, when the field of inquiry is large, this method becomes difficult to adopt because of the resources involved. At times, this method is practically beyond the reach of ordinary researchers. Perhaps, government is the only institution which can get the complete enumeration carried out. Even the government adopts this method in very rare cases such as population census conducted once in a decade. Further, many a times it is not possible to examine every item in the population and sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population. In such cases there is no utility of census surveys.

However, it needs to be emphasized that when the universe is a small one, it is no use resorting to a simple survey. When field studies are undertaken in practical life, consideration of time and cost invariably lead to a selection of respondents, i.e., selection of only a few items. The respondents selected should act as representatives of the total population in order to produce a miniature cross-section. The selected respondents constitute what is technically called a 'sample' and the selection process is called 'sampling technique'. The survey so conducted is known as 'sample survey'.

Algebraically, let the population size be N and if a part of size n (which is < N) of this population is selected according to some rule for studying some characteristics of the population, the group consisting of these n units is known as 'sample'. The researcher must prepare a sample design for his study, i.e., he must decide how a sample should be selected and of what size such a sample would be.

Implications of a sample design

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. Sample design may as well lay down the number of items to be included in the sample, i.e. the size of the sample. Sample design is determined before data are collected. There are many sample designs from which a researcher can choose. Some designs are relatively more precise and easier to apply than others. The researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

Merits of sampling

The following are the advantages of sampling:

- **Size of population:** It is very difficult to study a large population for a research study; hence, a sample from the population is selected for the study and represents all characteristics of population.
- **Funds requirement for the study:** When the funds availability is lesser than the anticipated cost of census survey, sampling is an effective method.

NOTES

- **Facilities:** When facilities like technology and staff members are limited, sampling is preferable.
- **Time:** The time required for the sampling procedure is less, so a researcher prefers this method.

Sampling procedure

Sampling is a complicated process. A researcher has to identify all the factors which can affect the sample. The various criteria related to choice of sampling procedure are as follows:

- **Purpose of survey:** Defining the purpose of survey helps the researcher in the selection of a particular method of sampling. A particular method of sampling choice depends on the geographical area of the survey and size and nature of the study.
- **Measurability:** The application of statistical inference theory requires computation of the sampling error from the sample itself. Probability samples only allow such computation. Hence, where the research objectives require statistical inference, the sample should be drawn by applying simple random sampling method or stratified random sampling method, depending whether the population is homogeneous or heterogeneous.
- **Degree of precision:** A desired level of precision of the result of the survey decides the method adopted for sampling.
- **Information about population:** Details of information available about the population to be studied help in deciding the method of sampling. If no data is available about population, it is difficult to apply probability random sampling. In this condition, the non-probability sampling method can be used for getting an idea of the population.
- **Nature of population:** Whether the population is homogeneous or heterogeneous decides the variables to be studied. Simple random sampling can be used for a homogeneous population. If the population is heterogeneous, stratified random sampling is a better option.
- **Geographical area of study and size of population:** Multi-stage, cluster sampling is used for the study of wide geographical area and large size of population.
- **Financial resources:** Availability of finance decides the need of sampling method.
- **Time limitation:** The time limit to complete a study decides the method of sampling.

Characteristics of a good sample

The following are the characteristics of a good sample:

- Representative
- Accurate

- Precise
- Right sized

Methods of Sampling

Sampling methods can be classified as:

NOTES

2.2.1 Probability or Random Sampling

Probability sampling is based on the theory of probability. It is also known as random sampling. It provides the known non-zero chance of selection to each population element. When generalization is the objective of study and high accuracy of estimation of population parameter is required, random sampling is used. The following are the types of random sampling:

(i) Simple random sampling

It provides each element an equal and independent chance of being selected. Equal chance means equal probability of selection. Independent chance means that draw of one element will not affect the chances of other elements being selected. It is used for small homogeneous population.

Advantages

The following are the advantages of simple random sampling:

- Easy to use
- Equal and independent chance of selection of every element
- No need of prior information of population

Disadvantages

The following are the disadvantages of simple random sampling:

- Impractical because of non-availability of population details
- Does not represent proportionate representation because observations are selected randomly so there is a chance of selecting same type of observation in one sample whereas population may consist of different type of observations
- May be expensive and time consuming

(ii) Stratified random sampling

When the population to be studied is heterogeneous in nature, it is divided into a homogeneous group or strata and from each homogeneous group, a random sample is drawn. Stratified random sampling can be classified into the following:

(a) Proportionate stratified sampling

Proportionate stratified sampling involves drawing a sample from each stratum in proportion to the latter's share in total population. In this, proper representation is

NOTES

given to each stratum. For example, let us assume that the management faculty of a university consists of the following specialization groups:

Specialization Stream	No. of Students	Proportion of Each Stream
Marketing	40	0.40
Hr	20	0.20
Finance	30	0.30
It	10	0.10
Total	100	1.00

If the researcher wants to draw an overall sample of 40, then the strata sample sizes would be:

Strata	Sample Size
Marketing	$40 \times 0.40 = 16$
HR	$40 \times 0.20 = 8$
Finance	$40 \times 0.30 = 12$
IT	$40 \times 0.10 = 4$
Total	40

Advantages

The following are the advantages of proportionate stratified sampling:

- Enhancement of representativeness to each sample as proportionate stratified sample consists of observations in proportion of various strata
- Higher statistical efficiency
- Easy to carry out
- Giving self-weighing sample

Disadvantages

The following are the limitations of proportionate stratified sampling:

- Prior knowledge of composition and of distribution of population
- Time consuming and expensive
- Classification error

(b) Disproportionate stratified random sampling

Proportionate representation is not given to strata. It necessarily involves giving over-representation to some strata and under representation to others. The desirability of disproportionate sampling is usually determined by the following three factors:

- The sizes of strata
- Internal variances among strata
- Sampling costs

NOTES

This method is suitable when the population has some small but important subgroups, when certain groups are homogeneous and it is expected that there will be significant differences in the response of the subgroups in the population.

Advantages

The following are the advantages of disproportionate stratified random sampling:

- Less time consuming
- Appropriate weightage to a particular group which is small but more important

Disadvantages

The following are the limitations of disproportionate stratified random sampling:

- No proportionate representation to each stratum
- Prior knowledge of composition of population required
- Doubtful practical feasibility
- Classification errors

(iii) Systematic random sampling

Systematic random sampling is an alternate to random selection. It consists of taking *n* item in population after a random start with an item from 1 to *n*. It is also known as fixed interval method; for example, 1st, 11th, 21st and so on. It possesses the characteristics of randomness and some non-probability traits.

Systematic selection can be applied to various populations such as students in class, houses in a street, yellow pages, telephone directory, etc.

Advantages

The following are the advantages of systematic random sampling:

- Simpler than random sampling
- Easy to use
- Easy to instruct
- Less time consuming
- Cost effective
- Statically more efficient

Disadvantages

The following are the disadvantages of systematic random sampling:

- Ignorance of all other elements between two *n* elements
- Each element does not get an equal chance
- Method gives biased sample

NOTES

(iv) Cluster sampling

Each sample chapter is a cluster of the population elements in this method. There is random selection of the sampling chapter, which consists of the population elements. Then from each selected sampling chapter, a sample of the population element is drawn.

Cluster sampling is used in socio-economic surveys, public opinions polls, ecological studies, farm management services, rural credit services, demographic studies and large scale surveys of political and social behaviour, attitude surveys, etc.

Advantages

The following are the advantages of cluster sampling:

- Easier and more convenient
- Cost effective
- Convenience of field work as it would be done in compact places
- Less time consumption
- Substitution of units for other units
- More flexible

Disadvantages

The following are the limitations of cluster sampling:

- Variation in cluster size
- Increased bias of resulting sample because of variation
- Sampling error

(v) Area sampling

Area sampling is also a form of cluster sampling. In a large field survey cluster consisting of specific geographical areas like districts, tallukas, blocks, villages, in a city are randomly drawn. When geographical areas are selected as sampling units, their sampling is known as area sampling.

(vi) Multi-stage sampling

Sampling is carried out in two or more stages. Firstly, a sample of the first stage sampling chapter is drawn, then from each of the selected first stage sampling units, a sample of the second stage sampling chapter is drawn. The procedure continues up to final sampling units or population elements. An appropriate random sampling method is adopted at each stage.

The population is scattered over a wider geographical area and no details are available for sampling.

NOTES

Advantages

The following are the advantages of multi-stage sampling:

- Time effective
- Cost effective

Disadvantages

The following are the limitations of multi-stage sampling:

- Procedure of estimating sampling error
- Cost advantage is complex

In multi-stage sampling, sampling at the second stage is called sub-sampling.

(vii) Random sampling with probability proportional to size

The procedure of selecting clusters with probability proportional to size (PPS) is used widely. If one primary cluster has twice as large a population as another, it is given twice the chance of being selected. If the same number of people are then selected from each of the selected clusters, the overall probability of any person will be the same. Thus, PPS is a better method of securing a representative sample of population elements in multi-stage cluster sampling.

Advantages

The advantage of a PPS cluster is that it is more accurate than a simple random sample of clusters.

Disadvantages

The following is the limitation of random sampling with probability proportional to size:

- PPS cannot be used if the sizes of the primary sampling clusters are not known.

(viii) Double sampling and multi-phase sampling

Double sampling refers to the subsection of the final sample form a pre-selected larger sample that provided information for improving the final selection. When the procedure is extended to more than two phases of selection, it is then called multi-phase sampling. This is also known as sequential sampling, as sub-sampling is done form a main sample in phases. Double sampling or multi-phase sampling is a compromise solution for a dilemma posed by undesirable extremes. The statistics based on the sample of 'n' can be improved by using ancillary information from a wide base, but this is too costly to obtain from the entire population of N elements. Instead, information is obtained from a larger preliminary sample which includes the final sample n.

NOTES

(ix) Replicated or interpenetrating sampling

Replicated sampling involves selection of a certain number of sub-samples rather than one full sample from a population. All the sub-samples should be drawn using the same sampling technique and each is a self-contained and adequate sample of the population. Replicated sampling technique can be used with any basic sampling technique: simple or stratified, single or multi-stage or single or multi-phase sampling. It provides a simple means of calculating the sampling error. It is practical. The replicated samples can throw light on variables and non-sampling errors. The only disadvantage is that it limits the amount of stratification that can be employed.

2.2.2 Non-Probability or Non-Random Sampling

Non-probability sampling or non-random sampling is not based on the theory of probability. This sampling does not provide a chance of selection to each population element.

Advantages

The only merit of this type of sampling is simplicity, convenience and low cost.

Disadvantages

The demerit of this type of sampling is that it does not ensure a selection chance to each population chapter. The selection probability sample may not be a representative one. The selection probability is unknown. It suffers from sampling bias which will distort results.

This sampling method is used when there is no other feasible alternative due to non-availability of a list of population, when the study does not aim at generalizing the findings to the population, when the costs required for probability sampling may be too large, when probability sampling require more time, but the time constraints and the time limit for completing the study do not permit it. It may be classified into the following:

Convenience or accidental sampling

Convenience sampling means selecting sample units in a just 'hit and miss' fashion; for example, interviewing people who we happen to meet. This sampling also means selecting whatever sampling units are conveniently available; for example, a teacher may select students in his class. This method is also known as accidental sampling because the respondents who the researcher meets accidently are included in the sample.

This type of sampling may be used for simple purposes such as testing ideas or rough impressions about a subject of interest.

Advantage

The main advantage of this method is that it is cheap and simple. Also, it does not require a list of population or statistical expertise.

NOTES

Disadvantage

The only disadvantage of this method is that it is highly biased because of the researcher's subjectivity. This method is used the least and its findings cannot be generalized.

2.2.3 Purposive (or Judgment) Sampling

Purposive sampling method means deliberate selection of the sample units that confirm to some predetermined criteria. This is also known as judgment sampling. It involves selection of cases which we judge as the most appropriate ones for the given study. It is based on the judgment of the researcher or some expert. It does not aim at securing a cross section of a population. The chance that a particular case is selected for the sample depends on the subjective judgment of the researcher.

This method is used when what is important is the typicality and specific relevance of the sampling units to the study and not their overall representativeness of the population.

Advantages

The following are the advantages of this sampling method:

- It is less costly and more convenient.
- It guarantees inclusion of the relevant elements in the sample.

Disadvantages

The following are the disadvantages of this sampling method:

- It is less efficient for generalizing.
- It does not ensure representativeness.
- It requires more prior extensive information and does not lend itself for using inferential statistics.

2.2.4 Quota Sampling

Quota sampling is a form of convenient sampling which involves selection of quota groups of accessible sampling units by traits such as sex, age, social class, etc. It is a method of stratified sampling in which the selection within strata is known as random. It is this known random element that constitutes its greatest weakness.

This sampling method is used in studies like marketing service, opinion polls and readership service which do not aim at making decisions, but try to get some crude results quickly.

Advantages

The following are the advantages of quota sampling:

- It is less costly.
- It takes less time.

NOTES

- There is no need for a list of the population.
- The fieldwork can easily be organized.

Disadvantages

The following are the disadvantages of quota sampling:

- It is impossible to estimate sampling error.
- Strict control if field work is difficult.
- It is subject to a higher degree of classification.

2.2.5 Snow-ball Sampling

Snow-ball sampling is the colourful name for a technique of building up a list or a sample of a special population by using an initial set of its members as informants. This sampling technique may also be used in socio-metric studies. It is used in studies involving respondents who are rare to find. To start with, the researcher compiles a short list of sample units from various sources. Each of these respondents is contacted to provide names of other probable respondents.

Snow-ball sampling is very suitable in studying social groups, informal groups in a formal organization and diffusion of information among professionals of various kinds.

Advantage

The main advantage of snow-ball sampling is that it is useful for smaller populations for which no frames are readily available.

Disadvantages

The following are the disadvantages of this method:

- It does not allow the use of probability statistical methods.
- It is difficult to apply when the population is large.
- It does not ensure the inclusions of all the elements in the list.

CHECK YOUR PROGRESS

1. List two characteristics of a good sample.
2. List two types of random sampling.
3. State the main advantage of accidental sampling.

2.3 SOURCES OF DATA

Finding the answers of questions for research study is called data collection. Data is facts, and other relevant materials, past and present, for study and analysis. For

the study of industrial marketing research, the data requirement can be classified as:

(a) Data relating to human beings

(b) Data relating to the organization

(c) Data relating to territorial or geographical areas

NOTES

Personal data related to human beings consists of the following:

- **Demographic and socio-economic characteristics of individuals:** Education, location, etc.
- **Behavioural variables:** Intentions, opinions, knowledge, awareness, attitudes, practice, etc.
- **Organizational data:** It consists of data relating to ownership, organizational beginning, objectives, functions, resources, performance and growth.
- **Territorial data:** This data is related to geo-physical characteristics, resource endowment, population pattern infrastructure degree of development, etc., of special divisions like villages, tallukas, districts, cities, state and the nation.

The data acts as the basic raw material for analysis. No study can be completed without a proper analysis of the available data. No results can be drawn on research questions. Correct answers are not presented through inferences based on imagination. The quality of findings in the research study depends on adequacy, relevance and reliability of data used in the research study.

2.3.1 Primary and Secondary Data

The classification of sources of data can be done as follows:

(a) Primary sources

(b) Secondary sources

(a) Primary sources of data: The data which is collected by the researcher himself/ herself directly from the original sources is called primary data and the original source is called primary source for data collection. Primary data can be gathered with the help of various methods such as mailing, interviewing, observation, etc.

Merits of primary data

The following are the merits of primary data:

- As primary data is the main source of data, the collected data can be used for several purposes in the research study.
- Primary data helps in understanding the changes that take place over a period of time. As time progresses, many facts get changed. In this context, secondary data may not be very useful and the researcher will require primary data, which is relevant as well as accurate.

- It is flexible, which is an advantage to the researcher. The same data can be utilized by the researcher for different purposes.
- Primary data is based on extensive and in-depth research study.

NOTES

Demerits of primary data

The following are the demerits of primary data:

- The process of collection of primary data is very expensive. A very high cost is involved in collecting data from primary sources as it uses manpower, survey tools, etc.
- It is time consuming. A lot of time is required in collecting data from various primary sources.
- The process of primary data collection requires skilled personnel, who have to conduct in-depth research. Availability of skilled people for research work is also a big challenge in research study at any place.
- It is difficult to administer.

Methods of collecting primary data

A researcher uses original sources to collect primary data. In this regard, the researcher collects the data according to the needs of the research. He collects the data according to the suitability of the research, timeliness of research and need of the research. Primary data collection is not only a costly affair, but is also time consuming. For many types of social science research, secondary data is not available. Therefore, it becomes difficult to conduct the research further. Now, the researcher needs the data which he collects by using primary sources and this is the original data collected by him.

Primary data has to be collected in a situation where the available data is not pertinent to the research study or the available data is inappropriate or obsolete. Various methods can be used to collect the data for research study. A 'method' is different from a 'tool'. The mode or way of gathering the data is known as method and a tool is an instrument used for the method. For instance, a schedule is used for interviewing. The important methods are as follows:

- Interviewing
- Observation
- Experimentation
- Mail survey
- Projective technique
- Simulation

Secondary sources of data: The data which has been collected and compiled for some other purposes by someone else and is used by a researcher for his own research study, is known as secondary data and the source of information is called

secondary source. The secondary sources contain previously analysed and compiled information and reports whose data can be utilized by researchers for their research work. Published records and reports are not the only secondary sources of data collection, even unpublished records like accounting and financial records, personnel records, register of members, minutes of meeting, inventory records, etc., can be used as a secondary source.

Features of secondary sources

The contents of the secondary sources of data are detailed and diversified, yet some of the certain common characteristics are as follows:

- Secondary data is easily available to all researchers and does not consist of the hassles of developing tools and using them.
- The data contained in a secondary source is free from the researcher's control in its collection and classification. The content as well as the form of the secondary sources is not developed by the researcher, in fact they are shaped by others. Thus, the value of research conducted with the help of secondary data can be useless if it has not been tested correctly.
- The researcher has to estimate the time, place and conditions of secondary sources and secondary data, which is not so easy. Practically, a researcher is not required to be present at the time of collecting data from the original source.

Use of secondary data

Secondary data can be used by the researchers in the following three ways:

(i) Secondary sources provide the needed information to the researcher for reference purposes. For instance, the general statistical information in the number of co-operative credit societies in the country, their coverage of villages, their capital structure, volume of business, etc., may be taken from published reports and quoted as background information in a study on the evaluation of the performance of cooperative credit societies in a selected district/state.

(ii) Secondary data is used as a benchmark to test the findings of research for making comparative analysis. For instance, the findings of a local or regional survey may be compared with the national averages; the performance indicators of a particular bank may be tested against the corresponding indicators of the banking industry as a whole and so on.

(iii) Secondary data is also used as the only source of information for research work such as market trend analysis, financial analysis of companies, etc. Year books, statistical report of government departments, report of public organizations of bureau of public enterprises, census reports, etc., serve as major data sources for such research studies.

NOTES

Merits of secondary data

The following are the advantages of secondary data:

- Secondary data is easily accessible and does not cost a lot to the researcher, depending on its availability. Once the source of documents and reports are located, collection of data is just a matter of desk work. Even the tediousness of copying the data from the source can now be avoided with the help of photocopying facilities.
- Wider and scattered geographical area and long historical period can be analysed by a researcher with less cost which is an added advantage of using secondary data. Therefore, secondary data's usage provides new horizons to the researcher's limits.
- The secondary data's use ensures the availability of data for making scientific generalizations from the studies.
- Environmental and cultural settings are required for the study.
- The secondary data helps a researcher test the findings obtained through primary data. It provides additional data for analysis of primary data. Thus, a researcher need not wait for the time to collect primary data for analysis.

Demerits of secondary data

The following are the disadvantages of secondary data:

- The secondary data availability cannot be as per the research study needs and it may be possible that the available secondary data is not significant. Significance may vary as time progresses ahead. It may be possible that the available data is not relevant to the current conditions.
- The available data is not needed. To assess the accuracy of secondary data we need to know the procedure as to how the data was collected.
- Even up-to-date secondary data becomes obsolete when it is presented in the printed form because of time lag in producing it. For instance, population census data is published two or three years after compilation and no new figures are added to it for another ten years.
- Information about the sources of secondary data may not be available to all researchers. Even if the location of the source is known, the accessibility depends primarily on proximity. For instance, most of the unpublished official records and compilations are located in the capital city and they are not within easy reach of researchers based in far off places.

Evaluation of secondary data

When a researcher wants to use secondary data for his research, he should evaluate it before deciding to use it. The following are the steps in the process of data evaluation:

NOTES

(i) Data pertinence

The first consideration in evaluation is to examine the pertinence of the available secondary data to the research problem under study. The following questions should be considered to check the relevance of the data:

- What are the definition and classifications? Are they consistent?
- What are the measurements of variables used? What is the degree to which they conform to the requirements of our research?
- What is the coverage of the secondary data in terms of topic and time? Does this coverage fit the needs of our research?

The pertinence of the secondary data to the research on hand should be determined on the basis of these considerations, as a researcher who is imaginative and flexible may be able to redefine his research problem so as to make use of otherwise unusual available data.

(ii) Data quality

If the researcher is convinced about the available secondary data for his needs, the next step is to examine the quality of the data. The quality of data refers to its accuracy, reliability and completeness. The assurance and reliability of the available secondary data depends on the organization which collected it and the purpose for which it was collected. What is the authority and reputation of the organization? Is it well recognized? Is it noted for reliability? Is it capable of collecting reliable data? Does it use trained and well qualified investigators? The answers to these questions determine the degree of confidence we can have in the data and its accuracy. It is important to go to the original source of the secondary data rather than to use an immediate source which has been quoted from the original. Then only, can the researcher evaluate the cautionary and other comments that were made in the original source.

(iii) Data completeness

The completeness refers to the actual coverage of the published data. This depends on the methodology and sampling design adopted by the original organization. Is the methodology sound? Is the sample size small or large? Is the sample method appropriate? The answers to these questions may indicate the appropriateness and adequacy of the data for the problem under study. The question of possible bias should also be examined. Whether the purpose for which the original organization collected the data had a particular orientation? Has the study been made to promote the organization's own interest? How was the study conducted? These are important clues. The researcher must be careful when the source does not give the details of the methodology and sampling design. Then it is not possible to determine the adequacy of the secondary data for the researcher's study.

NOTES

2.3.2 Qualitative and Quantitative Data

The data can be classified as **qualitative** and **quantitative**. Generally, data which is in the numerical form is known as 'quantitative', whereas any type of data which is not in the numerical form is known as 'qualitative' data. Qualitative data includes everything from words and text to sound recordings, videos, photographs, etc.

Quantitative = Numerical variables

Qualitative = Categorical variables

Quantitative data is anything that is **numerical** in nature. This includes how many, how frequently, or how much of something. For instance, how many times the rat pressed the lever, how frequently a child hit another child or how much implicit prejudice the participant showed.

Qualitative data is data that cannot be gathered numerically, but is instead **categorical** in nature. This includes the types of behaviour, rather than the number of behaviours. For instance, what type of career someone has, what parenting style someone has and so on.

For instance, if you want to look at how education is related to career success. You can look at this in a few different ways:

(i) **Quantitatively:** You can compare the number of years of education (a numerical variable) with income at age 50 (also a numerical variable). You might study this relationship using a correlation (requires 2 quantitative variables).

(ii) **Qualitatively:** You can compare the type of education (like no college degree, Associate's degree, Bachelor's degree, more than Bachelor's degree—a categorical variable) with the type of career (such as managerial, clerical, service, industrial—also a categorical variable). You would study this relationship using Chi-Square (requires two qualitative variables).

(iii) **Both:** You can use one quantitative variable and one qualitative variable. For instance, you might look at the type of education (like no college, Associate's degree, Bachelor's degree, more than Bachelor's degree—a categorical variable) and income at age 50 (a numerical variable). You would study this relationship using ANOVA (which requires one qualitative and one quantitative variable).

Which type of analysis you use (correlation vs. Chi-Square vs. ANOVA) depends on your exact research question.

2.3.3 Talcott Parsons (on Social System and Pattern Variables)

Talcott Parsons (1902–82) was for many years the best-known sociologist in the United States and indeed one of the best-known in the world. He produced a general theoretical system for the analysis of society that came to be called

structural functionalism. Parsons' analysis was largely developed within his major published works:

- *The Structure of Social Action* (1937)
- *The Social System* (1951)
- *Structure and Process in Modern Societies* (1960)
- *Sociological Theory and Modern Society* (1968)
- *Politics and Social Structure* (1969)

NOTES

Parsons was an advocate of the 'grand theory', an attempt to integrate all the social sciences into an overarching theoretical framework. His early work, *The Structure of Social Action* reviewed the output of his great predecessors, especially Max Weber, Vilfredo Pareto, and Émile Durkheim and attempted to derive from them a single 'action theory' based on the assumptions that human action is voluntary, intentional and symbolic. Later, he became intrigued with and involved in an astonishing range of fields: from medical sociology (where he developed the concept of the sick role to psychoanalysis-personally undergoing full training as a lay analyst) to anthropology, to small group dynamics to race relations and then economics and education.

Parsons is also well-known for his idea that every group or society tends to fulfill four **functional imperatives**. These are as follows:

- Adaptation to the physical and social environment
- Goal attainment, which is the need to define primary goals and enlist individuals to strive to attain these goals
- Integration, the coordination of the society or group as a cohesive whole
- Latency, maintaining the motivation of individuals to perform their roles according to social expectations

Parsons contributed to the field of social evolutionism and neo-evolutionism. He divided evolution into four sub-processes:

(i) Division, which creates functional subsystems from the main system

(ii) Adaptation, where those systems evolve into more efficient versions

(iii) Inclusion of elements previously excluded from the given systems

(iv) Generalization of values, increasing the legitimization of the ever-more complex system

Furthermore, Parsons explored these sub-processes within three stages of evolution:

(i) Primitive (ii) Archaic (iii) Modern

While archaic societies have the knowledge of writing, modern societies have the knowledge of law. Parsons viewed the Western civilization as the pinnacle of modern societies, and out of all western cultures, he declared the United States

NOTES

as the most dynamically developed. For this, he was attacked as an ethnocentrist. Parsons' late work focussed on a new theoretical synthesis around four functions common (he claimed) to all systems of action—from the behavioural to the cultural and a set of symbolic media that enable communication across them. His attempt to structure the world of action according to a mere four concepts was too much for many American sociologists, who were at that time retreating from the grand pretensions of the 1960s to a more empirical, grounded approach.

Pattern variables

Parsons asserted that there were two dimensions to societies: *instrumental* and *expressive*. By this he meant that there are qualitative differences between kinds of social interaction. Essentially, he observed that people can have personalized and formally detached relationships based on the roles that they play. He called the characteristics that were associated with each kind of interaction, pattern variables. Some examples of expressive societies would include families, churches, clubs, crowds and smaller social settings. Examples of instrumental societies would include bureaucracies, aggregates and markets.

1. **Affectivity vs affective neutrality:** When actor is oriented towards maximum satisfaction from a given choice.
2. **Particularism vs universalism:** Situations are judged according to uniform criteria (universalism) and not according to actor or individual relation with the given subject (particularism).
3. **Quality vs performance:** Defining people on the basis of biological difference and performance is judging people according to their performance and capacity.
4. **Self-orientation vs collective orientation:** When the actor acts out of personal interest, it is self-orientation.

Check Your Progress

4. How can data requirement be classified for the study of industrial marketing research?
5. What is qualitative data?
6. List two important methods of collecting primary data.

2.4 SUMMARY

- A researcher uses sampling for saving time and costs as a selected sample is a replica of the population.
- A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would

NOTES

adopt in selecting items for the sample. Sample design may as well lay down the number of items to be included in the sample, i.e., the size of the sample. Sample design is determined before data are collected.

- Probability sampling is based on the theory of probability. It is also known as random sampling. It provides the known non-zero chance of selection to each population element. When generalization is the objective of study and high accuracy of estimation of population parameter is required, random sampling is used.
- Non-probability sampling or non-random sampling is not based on the theory of probability. This sampling does not provide a chance of selection to each population element.
- Quota sampling is a form of convenient sampling which involves selection of quota groups of accessible sampling units by traits such as sex, age, social class, etc. It is a method of stratified sampling in which the selection within strata is known as random.
- Snow-ball sampling is a technique of building up a list or a sample of a special population by using an initial set of its members as informants.
- Finding the answers of questions for research study is called data collection. Data is facts, and other relevant materials, past and present, for study and analysis.
- For the study of industrial marketing research, the data requirement can be classified as:
  - o Data relating to human beings
  - o Data relating to the organization
  - o Data relating to territorial or geographical areas
- The classification of sources of data can be done as follows:
  - o Primary sources
  - o Secondary sources
- The data which is collected by the researcher himself/ herself directly from the original sources is called primary data and the original source is called primary source for data collection.
- A researcher uses original sources to collect primary data. In this regard, the researcher collects the data according to the needs of the research. He collects the data according to the suitability of the research, timeliness of research and need of the research.
- The data which has been collected and compiled for some other purposes by someone else and is used by a researcher for his own research study, is known as secondary data and the source of information is called secondary source.

NOTES

- The data can be classified as qualitative and quantitative. Generally, data which is in the numerical form is known as 'quantitative', whereas any type of data which is not in the numerical form is known as 'qualitative' data.
- Talcott Parsons (1902–82) was for many years the best-known sociologist in the United States and indeed one of the best-known in the world. He produced a general theoretical system for the analysis of society that came to be called structural functionalism.
- Parsons asserted that there were two dimensions to societies: instrumental and expressive. By this he meant that there are qualitative differences between kinds of social interaction.

2.5 KEY TERMS

- **Sampling:** It refers to selecting a part of the 'universe' with a view to draw conclusions about the 'universe' or 'population' for a study.
- **Census inquiry:** It is a complete enumeration of all the items in the 'population'.
- **Sample design:** It is a definite plan for obtaining a sample from a given population.
- **Primary data:** The data which is collected by the researcher himself/ herself directly from the original sources is called primary data.
- **Secondary data:** The data which has been collected and compiled for some other purposes by someone else and is used by a researcher for his own research study, is known as secondary data.

2.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Representativeness and accuracy are the characteristics of a good sample.
2. Simple random sampling and stratified random sampling are two types of random sampling.
3. The main advantage of accidental sampling is that it is cheap and simple. Also, it does not require a list of population or statistical expertise.
4. For the study of industrial marketing research, the data requirement can be classified as:
   (a) Data relating to human beings
   (b) Data relating to the organization
   (c) Data relating to territorial or geographical areas
5. Qualitative data is data that cannot be gathered numerically, but is instead categorical in nature.

6. Interviewing and observation are the two important methods of collecting primary data.

2.7 QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. Write a short note on the implications of a sample design.
2. List the advantages of sampling.
3. Write a short note on snow-ball sampling.
4. What are the various sources of data?
5. Who was Talcott Parsons? What is his contribution to sample designing?

Long-Answer Questions

1. Explain the sampling procedure in detail.
2. Discuss the various methods of sampling.
3. What are the advantages and disadvantages of primary and secondary data?
4. Explain Parsons' concept of structural functionalism.

2.8 FURTHER READING

Booth, Wayne. 2008. *The Craft of Research*, Third edition. Illinois: University of Chicago Press.

Creswall, John W. 2008. *Research Designs: Quantitative, Qualitative and Mixed Methods Approaches*. London: Sage Publications.

Christenson, Larry B. *et al.* 2010. *Research Methods, Design and Analysis*, Eleventh edition. New Jersey: Allyn and Bacon.

Kothari, C. R. 2008. *Research Methodology: Methods and Techniques*. New Delhi: New Age International.

Kumar, Ranjit. 2010. *Research Methodology: A Step-by-Step Guide for Beginners*, Third edition. New Delhi: Sage Publications.

UNIT 3 DATA PROCESSING, REPRESENTATION AND ANALYSIS

Structure

3.0 INTRODUCTION

In this unit, you will be introduced to the various concepts of data processing, data representation and data analysis. You will learn how to check for analysis along with editing, coding, classification and transcription of data. The unit will also discuss the representation of data. This will include tabulation, construction of a frequency table, components of the table, principles of table construction, frequency distribution, class intervals and graphs, charts and diagrams. You will also be taught about qualitative and quantitative data analysis. Under this, you will learn about the various measures of central tendency, the measures of dispersion, correlation analysis and the coefficient of determination.

3.1 UNIT OBJECTIVES

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

- Describe the processing of data which includes editing, coding and classification
- Assess data representation through tabulation, frequency tables, graphs, charts, diagrams and line graphs
- Analyse data analysis which includes measures of central tendency, measures of dispersion and correlation

3.2 DATA PROCESSING

Data is available in various types and forms. Interpreting this data is not easy. Most researches, especially social science researches, come to conclusions by using both primary and secondary data. To arrive at a meaningful interpretation on the research hypothesis, the researcher has to prepare his data for this purpose. This preparation involves the identification of data structure, the coding of data and the grouping of data for preliminary research interpretation. This data preparation for research analysis is termed as processing of data. Further selections of tools for analysis would, to a large extent, depend on the results of this data processing.

Data processing is an intermediary stage of work between data collection and data interpretation. The data gathered in the form of questionnaires/interviews/schedules/field notes/data sheets is mostly in the form of large volume of research variables. These research variables are a result of the execution of the preliminary research plan, which also sets out the data processing methods beforehand. Processing of data requires advanced planning and this planning may cover aspects, such as the identification of variables, hypothetical relationship among the variables and the tentative research hypothesis.

The various steps in processing of data are as follows:

- Identification of data structure
- Editing the data
- Coding and classifying the data
- Transcription of data

3.2.1 Checking for Analysis

In the data preparation step, the data is prepared in a data format, which allows the analyst to use modern analysis software such as SAS or SPSS.

The major criterion in this is to define the data structure. A data structure is a dynamic collection of related variables and can be conveniently represented as a graph where nodes are labelled by variables.

The data structure also defines the stages of the preliminary relationship between variables/groups that have been pre-planned by the researcher. Most data structures can be graphically presented to give clarity to the framed researched hypothesis. A sample structure could be a linear structure, in which one variable leads to the other and finally, to the resultant variable.

The identification of the nodal points and the relationship among the nodes could sometimes be a complex task. When the task is complex, which involves several types of instruments being collected for the same research question, the procedure for drawing the data structure would involve a series of steps. In several intermediate steps, the heterogeneous data structure of the individual data sets can be harmonized to a common standard and the separate data sets are then integrated into single data sets. However, the clear definition of such data structures would help in the further processing of data.

3.2.2 Editing

The next step in the processing of the data is editing of the data instruments. Editing is the process of checking to detect and correct errors and omissions. Data editing happens at two stages, one at the time of recording of the data and second at the time of analysis of data.

Data editing at the time of recording of data

Editing and testing of the data at the time of data recording is done keeping the following questions in mind:

- Do the filters agree?
- Is the data inconsistent?
- Have the 'missing values' been filled up?
- Have variable descriptions been specified?
- Have labels for variable names and value labels been defined and written?

All editing and cleaning steps are documented so that the redefining of variables or later analytical modification could be easily incorporated into the data sets.

Data editing at the time of analysis of data

Data editing is also a prerequisite of the analysis of data. This ensures that the data is complete in all respects for subjecting them to further analysis.

NOTES

Some of the usual checklist questions used by a researcher for editing data sets before analysis would be:

- Is the coding frame complete?
- Is the documentary material sufficient for the methodology description of the study?
- Is the storage medium readable and reliable?
- Has the correct data set been framed?
- Is the number of cases correct?
- Are there differences between questionnaires, coding frame and data?
- Are there undefined and so-called 'wild codes'?
- Has the original document been compared with the first counting of the data?

The editing step checks for completeness, accuracy and the uniformity of the data, as created by the researcher.

Completeness: The first step of editing is to check whether there is an answer to all the questions/variables set out in the data set. If there were an omission, the researcher sometimes would be able to deduce the correct answer from other related data on the same instrument. If this is possible, the data set has to be rewritten on the basis of the new information. For example, the approximate family income can be inferred from other answers to the probes such as occupation of family members, sources of income, approximate spending and saving and borrowing habits of family members, etc. If the information is vital and has been found to be incomplete, then the researcher can take the step of contacting the respondent personally again and solicit the requisite data again. If none of these steps are useful, the data should be marked 'missing'.

Accuracy: Apart from checking for omissions, the accuracy of each recorded answer should be checked. A random checking process can be applied to trace the errors at this stage. Consistency in responses can also be checked at this step. The cross verification to a few related responses would help in checking for consistency in responses. The reliability of the data set would heavily depend on this step of error correction. Clear inconsistencies should be rectified in the data sets.

Uniformity: In editing data sets, the researcher should be on the lookout for lack of uniformity, interpretation of questions and instruction by the data recorders. For instance, the responses towards a specific feeling could have been queried from a positive as well as the negative angle. While interpreting the answers, care should be taken to record the answer as a 'positive question' response or as a 'negative question' response in all uniformity checks for

consistency in coding throughout the questionnaire/interview/schedule response/data set.

The final step in the editing of data is to maintain a log of all corrections that have been carried out at this stage. The documentation of these corrections helps the researcher retain the original data set.

3.2.3 Coding

The edited data is then subject to *codification* and *classification*. The coding process assigns numerals or other symbols to the several numerals or other symbols to the responses of the data set. If there is a prerequisite to prepare a coding scheme for the data set, the recording of the data is done on the basis of this coding scheme.

The responses collected in the data sheets vary, sometimes the responses could be the choice among the multiple responses, sometimes the response could be values and sometimes the response could be alphanumeric. At the recording stage itself, if some codification were done to the responses collected, it could be useful in the data analysis. When codification is done, it is imperative to keep a log of the codes allotted to the observations. This code sheet will help in the identification of variables/observations and the basis for such codification.

The first coding done to primary data sets are the individual observations themselves. This response sheet coding gives a benefit to the research, in that, the verification and editing of the recordings and further contact with the respondent can be achieved without any difficulty. The codification can be done at the time of distribution of the primary data sheets itself. The codes can be alphanumerical to keep track of where and to whom it had been sent. For instance, if the data consists of several people in different localities, the sheets that are distributed in a specific locality may carry a unique part code which is alphabetic. To this alphabetic code, a numeric code can be attached to distinguish the person to whom the primary instrument was distributed. This also helps the researcher to keep track of who the respondent is and who the probable respondents are from whom primary data sheets are yet to be collected. Even at a latter stage, any specific queries on a specific response sheet can be clarified.

The variables or observations in the primary instrument would also need codification, especially when they are categorized. The categorization could be on a scale, i.e., most preferable to not preferable, or it could be very specific, such as gender classified as male and female. Certain classifications can lead to open-ended classification such as education classification, with classifications like illiterate, graduate, professional, others (please specify). In such instances, the codification needs to be carefully done to include all

NOTES

possible responses under 'others please specify'. If the preparation of the exhaustive list is not feasible, then it will be better to create a separate variable for the 'others please specify' category and record all the responses.

Numeric coding: Coding need to not necessarily be numeric. It can also be alphabetic. Coding has to be done only numerically when the variable is subject to further parametric analysis.

Alphabetic coding: Mere tabulation or frequency count or graphical representation of the variable may be given in an alphabetic coding.

Zero coding: A coding of zero has to be assigned carefully to a variable. In many instances, when manual analysis is done, a code of zero would imply a 'no response' from the respondents. Hence, if a value of zero is to be given to specific responses in the data sheet, it should not lead to the same interpretation of non- response. For instance, if there is a tendency to give a code of 0 to a 'no.' then a coding other than zero should be given in the data sheet. An illustration of the coding process of some of the demographic variable is given in the following table.

Question Number	Variable Observation	Response categories	Code
1.1	Organization	Private	Pt
		Public	Pb
		Government	Go
3.4	Owner of vehicle	Yes	2
		No	1
4.2	Vehicle performs	Excellent	5
		Good	4
		Adequate	3
		Bad	2
		Worst	1
5.1	Age	Upto20 years	1
		21-40 years	2
		40-60 years	3
5.2	Occupation	Salaried	S
		Professional	P
		Technical	T
		Business	B
		Retired	R
		Housewife	H
		Others	=

= could be treated as a separate variable/observation and the actual response could be recorded. The new variable could not be termed as 'other occupation'.

The coding sheet needs to be prepared carefully if the data recording is not done by the researcher but is outsourced to a data entry firm or individual. In order to enter the data in the same perspective as the researcher would like to view it, the data coding sheet is to be prepared first and a copy of the data coding sheet should be given to the outsourcer to help in the data

entry procedure. Sometimes, the researcher might not be able to code the data from the primary instrument itself. He may need to classify the responses and then code them. For this purpose, classification of data is also necessary at the data entry stage.

3.2.4 Classification

When open-ended responses have been received, classification is necessary to code the responses. For instance, the income of the respondent could be an open-ended question. From all responses, a suitable classification can be arrived at. A classification method should meet certain requirements or should be guided by certain rules.

First, classification should be linked to the theory and the aim of the particular study. The objectives of the study will determine the dimensions chosen for coding. The categorization should meet the information required to test the hypothesis or investigate the questions.

Second, the scheme of classification should be exhaustive, i.e., there must be a category for every response. For example, the classification of marital status into three categories viz., married, single and divorced is not exhaustive because responses like 'widower' or 'separated' cannot be fitted into the scheme. Here, an open-ended question will be the best mode of getting the responses. From the responses collected, the researcher can fit a meaningful and theoretically supportive classification. The inclusion of classification 'others' tends to fill the cluttered, but few responses from the data sheet. But the 'others' categorization has to be carefully used by the researcher. However, this categorization tends to defeat the very purpose of classification which is designed to distinguish between observations in terms of the properties under study. The classification 'others' will be very useful when a minority of respondents in the data set give varying answers. For instance, the reading habits of newspaper may be surveyed. 95 respondents out of 100 could be easily classified into 5 large reading groups while 5 respondents could have given a unique answer. These given answers, rather than being separately considered, could be clubbed under the 'others' heading for meaningful interpretation of respondents and reading habits.

Third, the categories must also be mutually exhaustive so that each case is classified only once. This requirement is violated when some of the categories overlap or different dimensions are mixed up.

The number of categories for a specific question/observation at the coding stage should be maximum permissible number, since reducing the categorization at the analysis level would be easier than splitting an already classified group of responses. However, the number of categories is limited by the number of cases and the anticipated statistical analysis that is to be used on the observation.

NOTES

3.2.5 Transcription of Data

When the observations collected by the researcher are not very large, the simple inferences, which can be drawn from the observations, can be transferred to a data sheet, which is a summary of all the responses on all observations from a research instrument. The main aim of transition is to minimize the shuffling proceeds between several responses and several observations. Suppose a research instrument contains 120 responses and the observations have been collected from 200 respondents, a simple summary of one response from all 200 observations would require shuffling of 200 pages. The process is quite tedious if several summary tables are to be prepared from the instrument. The transcription process helps in the presentation of all responses and observations on data sheets which can help the researcher arrive at preliminary conclusions of the nature of the sample collected, etc. Transcription is hence an intermediary process between data coding and data tabulation.

Methods of transcription

The researcher may adopt a manual or computerized transcription. Long worksheets, sorting cards or sorting strips could be used by the researcher to manually transcript the responses. The computerized transcription could be done using a database package such as spreadsheets, text files or other databases.

The main requisite for a transcription process is preparation of the data sheets where observations are the row of the database and the responses/ variables are the columns of the data sheet. Each variable should be given a label so that long questions can be covered under the label names. The label names are thus the links to specific questions in the research instruments. For instance, opinion on consumer satisfaction could be identified through a number of statements (say 10 thus); the data sheet does not contain the details of the statement, but gives a link to the question in the research instrument though variables labels. In this instance, the variable names could be given as CS1, CS2, CS3, CS4, CS5, CS6, CS7, CS8, CS9 and CS10. The label CS indicates Consumer Satisfaction and the numbers 1 to 10 indicate the statement measuring consumer satisfaction. Once the labelling process has been done for all the responses in the research instrument, the transcription of the response is done.

Manual transcription

When the sample size is manageable, the researcher need not use any computerization process to analyse the data. The researcher could use a manual transcription and analysis of responses. The choice of manual transcription would be taken when the number of responses in a research instrument is very less, say 10 responses, and the number of observations

collected are within 100. A transcription sheet with 100×50 (assuming each response has 5 options) row /column can be easily managed by a researcher manually. On the other hand, if the variables have 5 options, it results in a worksheet of 100×200 sizes which might not be easily managed by the researcher manually. In the second instance, if the number of responses is less than 30, then the worksheet could be attempted manually. In all other instances, it is advisable to use a computerized transcription process.

Long worksheets

Long worksheets require quality paper, preferably chart sheets thick enough to last several usages. These worksheets normally are ruled both horizontally and vertically, allowing responses to be written in the boxes. If one sheet is not sufficient, the researcher may use multiple ruled sheets to accommodate all the observations. Heading of the responses which are variable and their coding (optional) are filled in the first two rows. The first column contains the code of observations. For each variable, the responses from the research instrument are then transferred to the worksheet by ticking the specific option that the observer has chosen. If the variable cannot be coded into categories, sufficient length for recording the actual response of the observer should be provided for in the worksheet.

The worksheet can then be used for preparing the summary tables or can be subjected to further analysis of data. The original instrument can now be kept aside as safe documents. Copies of the data sheets can also be kept for further reference. As has been discussed under the editing section, the transcription data has to be subjected to a testing to ensure error-free transcription of data.

Transcription can be made as and when the edited instrument is ready for processing. Once all the schedules/questionnaires have been transcripted, the frequency tables can be constructed straight from the worksheet. Other methods of manual transcription involve adoption of sorting strips or cards.

In olden days, data entry and processing were made through mechanical and semi-metric devices such as key punch using punch cards. The arrival of computers has changed the data processing methodology altogether.

Check Your Progress

1. List the various steps in the processing of data.
2. Why is the data edited?
3. List the various ways in which data can be coded.
4. List the two methods of transcription of data.

NOTES

3.3 DATA REPRESENTATION

Data can be represented in various forms such as tabulation, frequency table, graphs, charts, diagrams and line graphs. These are described in this section.

3.3.1 Tabulation

The transcription of data can be used to summarize and arrange the data in a compact form for further analysis. This process is called **tabulation**. Thus, tabulation is the process of summarizing raw data and displaying them in compact statistical tables for further analysis. It involves counting the number of cases falling into each of the categories identified by the researcher. Tabulation can be done manually or through the computer. The choice depends on the size and type of study, cost considerations, time pressures and the availability of software packages. Manual tabulation is suitable for small and simple studies.

Manual tabulation

When data is transcribed in a classified form as per the planned scheme of classification, category-wise totals can be extracted from the respective columns of the worksheets. A simple frequency table counting the number of 'yes' or 'no' responses can be made easily by counting the 'Y' and 'N' response columns in the manual worksheet table prepared earlier. This is a one-way frequency table and can be readily inferred from the totals of each column in the worksheet.

Sometimes the researcher has to cross-tabulate two variables, for instance, the age group of vehicle owners. This requires a two-way classification and cannot be inferred straight from any technical knowledge or skill. If one wants to prepare a table showing the distribution of respondents by age, a tally sheet showing the age groups horizontally is prepared. Tall remarks are then made for the respective groups, i.e., vehicle owners, from each line of response in the worksheet. After every fourth tally, the fifth tally is cut across the previous four tallies. This represents a group of five items. This arrangement facilitates easy counting of each of the class groups. Illustration of this tally sheet is shown in the following:

Age group of Responses	Tally marks	No.
Below 20	II	2
20–39	IIII IIII IIII IIII III	23
40–59	IIII IIII IIII	15
Above 59	IIII IIII	10
Total		50

Although manual tabulation is simple and easy to construct, it can be tedious, slow and error-prone as the responses increase.

Computerized tabulation can be done easily with the help of software packages. The input requirement will be the column and row variables. The software package then computes the number of records in each cell of three row column categories. The most popular package is the Statistical Package for Social Science (SPSS). It is an integrated set of programme suitable for analysis of social science data. This package contains programmes for a wide range of operations and analysis such as handling, missing data, recording variable information, simple descriptive analysis, cross tabulation, multivariate analysis and non- parametric analysis.

3.3.2 Construction of Frequency Table

Frequency tables provide a 'shorthand' summary of the data. The importance of presenting statistical data in tabular form needs no emphasis. Tables facilitate comprehending masses of data at a glance; they conserve space and reduce explanations and descriptions to a minimum. They give a visual picture of relationships between variables and categories. They facilitate summation of items and the deduction of errors and omissions and provide the basis for computation.

It is important to make a distinction between general purpose tables and specific tables. General purpose tables are primary or reference tables designed to include large amounts of source data in a convenient and assessable form. Specific purpose tables are analytical or derivate ones that demonstrate significant relationships in the data or the result of statistical analysis. Tables in government reports on population, vital statistics, agriculture, industries, etc., are of general purpose type. They represent extensive repositories and statically information. Special purpose tables are found in monographs, research reports and articles and are reused as instruments of analysis. In research, we are primarily concerned with a specific goal.

3.3.3 Components of a Table

The following are the major components of a table:

(i) Heading:

- Table number
- Title of the table
- Designation of the units

(ii) Body:

- Sub-head, headings of all rows or blocks of sub-items
- Body-head: Headings of all columns or main captions and their sub-captions

NOTES

- Field/body: The cell in rows and columns

(iii) Notations:

- Footnotes, where applicable
- Source, where applicable

3.3.4 Principles of Table Construction

There are certain generally accepted principles or rules related to the construction of tables. These are given as follows:

- Every table should have a title. The title should represent a concise description of the contents of the table. It should be clear and concise. It should be placed above the body of the table.
- A number facilitating easy reference should identify every table. The number can be centred above the title. The table numbers should run in constitutive serial order. For instance, tables in chapter 1 should be numbered as 1.1, 1.2 and so on, in chapter 2 as 2.1, 2.2, 2.3 and so on.
- The captions (or column headings) should be clear and brief.
- The units of measurement under each heading must always be indicated.
- If there is any additional information related to the table, it should be written as a footnote and placed immediately below the table. This will help avoid any confusion with the footnotes of the main text.
- The source(s) from where the data has been taken should be mentioned immediately below the table.
- The data is represented in a columnar form. Lines are drawn at the top and bottom of the table and below the captions.
- The columns may be numbered to facilitate reference.
- All column figures should be properly aligned. Decimal point and 'plus' or 'minus' signs should be in perfect alignment.
- Columns and rows that are to be compared with one another should be brought close together.
- Totals of rows should be placed in the extreme right column and totals of columns at the bottom.
- In order to emphasize the relative significance of certain categories, different kinds of typeface, spacing and identifications can be used.
- The arrangement of categories in a table may be chronological, geographical, alphabetical or according to magnitude. Numerical categories are usually arranged in descending order of magnitude.
- Miscellaneous and exceptional items are generally placed in the last row of the table.

- Items in a table are usually listed vertically.
- Abbreviations should be avoided whenever possible and ditto marks should not be used in a table.
- The table should be made as logical, clear, accurate and simple as possible.

Text references should identify tables by number, rather than by expressions such as, 'The table above' or 'the following table'. Tables that are too wide can be formatted in landscape. While writing a report or a thesis, some writers place both, special purpose and general purpose tables in an appendix and refer to them in the text by numbers. This practice has the disadvantage of inconveniencing the reader who wants to study the tabulated data as the text is read. A more appropriate procedure is to place special purpose tables in the text and primary tables, if needed at all, in appendices.

3.3.5 Frequency Distribution and Class Intervals

Variables that are classified according to magnitude or size are often arranged in the form of a frequency table. While constructing this table, it is necessary to determine the number of class intervals to be used and the size of the class intervals.

A distinction is usually made between continuous and discrete variables. A continuous variable has an unlimited number of possible values between the lowest and the highest, with no gaps or breaks. Examples of continuous variables are age, weight, temperature, etc. A discrete variable can have a series of specified values with no possibility of values between these points. Each value of a discrete variable is distinct and separate. Examples of discrete variables are gender of persons (male/female), occupation (salaried, business, profession) and car size (800cc, 1000cc, 1200cc).

In practice, all variables are treated as discrete units, the continuous variables being stated in some discrete chapter size according to the needs of particular situations. For example, length is described in discrete units of millimeters or a tenth of an inch.

Class intervals: Ordinarily, the number of class intervals may not be less than five and not more than fifteen, depending on the nature of the data and the number of cases being studied. After noting the highest and lowest values and the features of the data, the number of intervals can be easily determined.

For many types of data, it is desirable to have class intervals of uniform size. The intervals should neither be too small nor too large. Whenever possible, the intervals should represent convenient numerical divisions such as five or ten, rather than odd division such as three or seven. Class intervals must be clearly designated in a frequency table in such a way any possibility

of misinterpretation or confusion is obviated. For example, to represent the age group of the population, use of intervals 1–20, 20–50, and 50 and above would be confusing. This may be presented as 1–20, 21–50, and above 50.Every class interval has a mid point. For example, the mid point of an interval 1–20 is 10.5. Once the class intervals are determined, it is routine work to count the number of cases that fall in each interval.

One-way tables

One-way frequency tables present the distribution of cases on only a single dimension or variable. For example, the distribution of respondents by gender, religion and social or economic status are shown in one-way tables. One-way tables are used because the resultant frequency distributions can be described in simple sentences. For instance, the gender distribution of a sample study may be described as 'The sample data represents 58 per cent are males and 42 per cent of the sample are females.'

Two-way tables

Distributions in terms of two or more variables and the relationship between the two variables are shown in two-way table. The categories of one variable are presented one below another, on the left margin of the table, and those of another variable at the upper part of the table, one along the other. The cells represent a particular combination of both variables. To compare the distributions of cases, raw numbers are converted into percentages based on the number of cases in each category.

The following illustrates two-way tables:

	Extent of participation						
Category Members	**Low no. of Respondents**	**%**	**Medium No. of respondents**	**%**	**High no. of respondents**	**%**	**Total**
Ordinary	65	41.9	83	56.8	2	1.3	115
Committee	4	10.3	33	84.6	2	5.1	39

Another method of constructing a two-way table is to state the percent of representation as a within brackets term rather than as a separate column. Here, special care has been taken as to how the percentage is calculated, either on a horizontal representation of data or as vertical representation of data. Sometimes, the table heading itself provides a meaning to the method of representation in the two-way table.

	Democratic Participation			
Economic Status	**Low**	**Medium**	**High**	**Total**
Low	6(35.3)	11(64.7)	0(0.0)	17
Medium	13(38.2)	18(53.0)	3(8.8)	34
High	6(62.5)	10(62.5)	0(0.0)	16
Very High	2(33.3)	3(50.0)	1(16.7)	6
Total	27	42	4	73

NOTES

3.3.6 Graphs, Charts and Diagrams

In presenting the data of frequency distributors and statistical computations, it is often desirable to use appropriate forms of graphical presentations. In addition to tabular forms, graphics presentation involves the use of **graphics**, **charts** and other pictorial devices such as **diagrams**. These forms and devices reduce large masses of statistical data to a form that can be quickly understood. The meaning of figures in tabular form may be difficult for the mind to grasp or retain. Properly constructed graphs and charts relieve the mind of burdensome details by portraying facts concisely, logically and simply. Then emphasizing new and significant relationships are also useful in discovering new facts and in developing hypothesis.

The device of graphic presentation is particularly useful when the prospective readers are non-technical people or the general public. It is useful to even technical people for dramatizing certain points about data, for important points can be more effectively captured in pictures than in tables. However, graphics forms are not substitutes for tables, but are additional tools for the researcher to emphasize the research findings.

Graphics presentation must be planned with utmost care and diligence. Forms of **graphics** used should be simple, clear, accurate and appropriate to the data. In planning this work, the following questions must be considered:

- What is the purpose of the diagram?
- What facts are to be emphasized?
- What is the educational level of the audience?
- How much time is available for the preparation of the diagram?
- What kind of chart will portray the data most clearly and accurately?

Types of graphs and general rules

The most commonly used graphical forms may be grouped into the following categories:

- Line graphs or charts
- Bar charts
- Segmental presentations
- Scatter plots
- Bubble charts
- Stock plots
- Pictographs
- Chesnokov faces

NOTES

The general rules to be followed in graphics representations are as follows:

- The chart should have a title placed directly above the chart.
- The title should be clear, concise and simple and should describe the nature of the data presented.
- Numerical data upon which the chart is based should be presented in an accompanying table.
- The horizontal line measures the time or independent variable and the vertical line measures the variable.
- Measurements proceed from left to right on the horizontal line and from bottom to top on the vertical.
- Each curve or bar on the chart should be labelled.
- If there is more than one curve or bar, they should be clearly differentiated from one another by distinct patterns or colours.
- The zero should always be represented and the scale intervals should be equal.
- Graphical forms should be used sparingly. Too many forms distract rather than illuminate a presentation.
- Graphic forms should follow and not precede the related textual discussion.

3.3.7 Line Graphs

The line graph is useful for showing changes in data relationship over a period of time. In this graph, figures are plotted in relation to two intersecting lines or axes. The horizontal line is called the abscissa or X-axis and the vertical, the ordinal or Y-axis. The point at which the two axes intersect is zero for both X and Y-axis. The point at which the two axes intersect is zero for both X and Y axes. 0 is the origin of coordinates. The two lines divide the region of the plane into four sections known as quadrants, which are numbered anti-clockwise. Measurements to the right and above 0 are positive (plus) and measurement to the left and below 0 is negative (minus). Any point on plane of the two axes is plotted in terms of the two axes reading from the origin 0. Scale intervals in both the axes should be equal. If a part of the scale is omitted, a set of parallel jagged lines should be used to indicate the break in the scale. The time dimension or independent variable is represented by the X-axes and the other variable by Y-axes.

CHECK YOUR PROGRESS

5. What are the limitations of manual tabulation?
6. List the major components of a table.

NOTES

3.4 DATA ANALYSIS: QUANTITATIVE AND QUALITATIVE

This section discusses data analysis which includes measures of central tendency, measures of dispersion, correlation analysis and coefficient of determination.

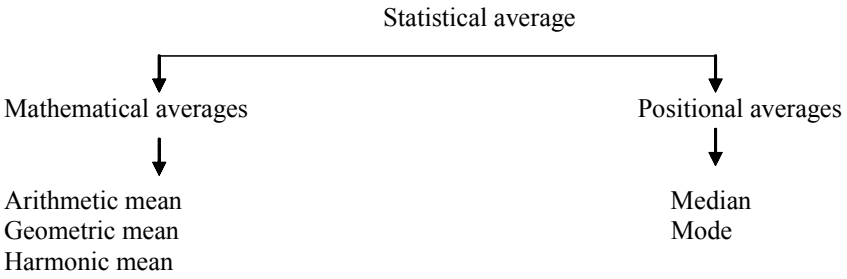
3.4.1 Measures of Central Tendency

Analysis of data involves understanding the characteristics of the data. The following are the important characteristics of statistical data:

- Central tendency
- Dispersion
- Skewness
- Kurtosis

In a data distribution, the individual items may have a tendency to come to a central position or an average value. For instance, in a distribution of marks, individual students may score marks between zero and hundred. In such distributions, the tendency of the data to concentrate towards the central position of the distribution is called central tendency. Central tendency of the data is measured by statistical averages. Averages are classified into the following two groups:

- Mathematical averages
- Positional averages



Arithmetic mean, geometric mean and harmonic mean are mathematical averages. Median and mode are positional averages. These statistical measures try to understand how individual values in a distribution concentrate to a central value like average. If the value of distribution approximately comes near to the average value, we conclude that the distribution has tendency.

Arithmetic mean

Arithmetic mean is the most commonly used statistical average. It is the value obtained by dividing the sum of the items by the number of items in a series. Symbolically we say:

NOTES

Arithmetic mean = $\Sigma X/n$

Where ΣX = the sum of the items

N = the number of items in the series.

If $X_1\ X_2\ X_3 \ldots X_n$ are the values of a series, then arithmetic mean of the series obtained by

$(X_1 + X_2 + X3 \ldots + X_n)/n$. If put $(X_1 + X_2 + X_3 + \ldots .. X_n) = \Sigma X$,

Then arithmetic mean = $\Sigma X/n$

When frequency is given along with the values, to calculate arithmetic mean, the values are first multiplied with the corresponding frequency. Then their sum is divided by the frequency. Thus, in a discrete series, arithmetic mean is calculated by the following formula:

Arithmetic mean = $\Sigma fx/\Sigma f$

Where, Σfx = Sum the values multiplied by the corresponding frequency

Σf = Sum of the frequency

If $X_1, X_2, X_3, \ldots .. X_n$ are the values of a series, and $f_1, f_2, f_3, \ldots .. f_n$ are their corresponding frequencies.

Arithmetic mean is calculated by $(f_1 X_1 + f_2 X_2 + f_3 X_3 \ldots + f_n X_n)/(f_1 + f_2 + f_3 \ldots + f_n)$ or

Arithmetic mean = $\Sigma fx/\Sigma f$

Individual series

1. Find arithmetic mean of the following data.

   58 67 60 84 93 98 100

   Arithmetic mean = $\Sigma X/n$

   Where,

   ΣX = the sum of the item

   N = the number of items in the series

   $\Sigma X = 58 + 67 + 60 + 84 + 93 + 98 + 100 = 560$

   N = 7

   $\Sigma X = 560/7 = 80$

2. Find arithmetic mean of the following distribution.

   2.0 1.8 2.0 2.0 1.9 2.0 1.8 2.3 2.5 2.3

   1.9 2.2 2.0 2.3

   Arithmetic mean = $\Sigma X/n$

   Where,

   ΣX = the sum of the items

n = the number of items in the series

ΣX = 2.0 + 1.8 + 2.0 + 2.0 + 1.9 + 2.0 + 1.8 + 2.3 + 2.5 + 2.3 + 1.9 + 2.2 + 2.0 + 2.3 = 29

N = 14

ΣX = 29/14=2.07

Discrete series

3. Calculate arithmetic mean of the following 50 workers according to their daily wages.

Daily Wages : 15 18 20 25 30 35 40 42

Number of workers: 2 3 5 10 12 10 5 2

Arithmetic mean using direct formula

Wages (X)	Frequency(F)	Fx
15	2	30
18	3	54
20	5	100
25	10	250
30	12	360
35	10	350
40	5	200
42	2	84
45	1	45
	Σf = 50	Σfx = 473

Arithmetic mean = Σfx/Σf

Where,

Σfx = 473

Σf = 50

Arithmetic mean = 1473/50 = 29.46

Continuous series

4. Find arithmetic series for the following distribution.

Marks : 10–20 20–30 30–40 40–50 50–60 60–70 70–80 80–90

No. of students : 6 12 18 20 20 14 8 2

Marks	Frequency(f)	Mid value(X)	fx
10–20	6	15	90
20–30	12	25	300
30–40	18	35	630
40–50	20	45	900
50–60	20	55	1100
60–70	14	65	910
70–80	8	75	600
80–90	2	85	170
	Σf = 100		Σfx = 4700

Arithmetic mean = Σfx/Σf

Where,

Σfx = 4700

Σf = 100

Arithmetic mean = 4700/100

= 47

Geometric mean

Geometric mean is defined as the nth root of the product of N items of a series. If there are two items in the data, we take the square root; if there are three items we take the cube root, and so on.

Symbolically,

GM = $\sqrt{x1.x2 \ldots xn}$

Where,

X1, X2, X3, …..Xn are the items of the given series.

To simplify calculations, logarithms are used.

Accordingly,

GM = Anti log of (Σlog X/n)

In discrete series

GM = Anti log of Σf . log x/Σf

Illustration

1. Find geometric mean for the following data.

   25 279 112 3675 84 9 18 54 73 648

Value(X)	Log X
25	1.3979
279	2.4456
112	2.0492
3675	3.5652
84	1.9242
9	0.9542
18	1.2552
54	1.7323
73	1.8633
648	2.8116
	19.9986

GM = Anti log of (“log X/n)

= Anti log of (19.9986/10)

= Anti log of 1.9986

= 99.967

NOTES

Geometric mean for discrete series:

Calculate the geometric mean of the following data.

Class	No. of families	Income
Landlords	1	100
Cultivators	50	80
Landless Laborers	25	40
Money lenders	2	750
Scholl Teachers	3	100
Shop Keepers	4	150
Carpenters	3	120
Weavers	5	60

Income	Frequency	LogX	f.Log X
1000	1	3.0000	3.0000
80	50	1.9031	95.1550
40	25	1.6021	40.0525
750	2	2.8751	5.7502
100	3	2.0000	6.0000
150	4	2.1716	8.7044
120	3	2.0792	6.2376
60	5	1.7782	8.8910
	93		173.7907

GM = Anti log of Σf.Log X/Σf

= Anti log of 173.7907/93

= Anti log 1.86871

= 73.91

Harmonic mean

In individual series

HM = N/”f (1/m)

N = Total Frequency

M = Mi values of the class

Illustration

For individual series

1. Find harmonic mean of the following data:

5 10 3 7 125 58 47 80 45 26

Values X	Factorial 1/X
5	.2
10	.1
3	.33
7	.14
125	.008
58	.017
47	.021
80	.014
45	.022
26	.038
	{(1/X) =.89

HM = $N/\Sigma(1/x)$

HM = 10/.89

= 11.235

Harmonic mean for discrete series

Compute harmonic mean for the following data

Marks : 10 20 25 30 40 50

Frequency : 20 10 15 25 10 20

Marks	Frequency	1/x	f.1/x
10	20	.1	2.0
20	10	.05	0.5
25	15	.04	.6
30	25	.033	.83
40	10	.025	.25
50	20	.02	.4
	{f=100		{f(1/x)=4.58

HM = $N/\Sigma f(1/x)$

HM = 100/4.58

= 21.834

Harmonic mean for continuous series

1. Calculate harmonic mean for the given data.

Class : 10–20 20–30 30–40 40–50 50–60 60–70

Frequency : 5 7 3 15 12 8

Class	Frequency	Midx	1/x	F.1/x
10-20	5	15	.0661	.33
20-30	7	25	.04	.28
30-40	3	35	.0285	.085
40-50	15	45	.0222	.333
50-60	12	55	.0181	.218
60-70	8	65	.0153	.123
	{f=50			{F(1/x)=1.369

HM = $N/\Sigma(1/x)$
HM = 50/1.369
= 37.8689

NOTES

Median

Median is the middlemost item of a given series. In individual series, we arrange the given data according to ascending or descending order and take the middlemost item as the median.

When two values occur in the middle, we take the average of these two values as median. Since median is the central value of an ordered distribution, there are equal number of values to the left and right of the median.

Individual series

Median = (N+ 1/2)th item

Illustration:

1. Find the median of the following scores.

   97 50 95 51 90 60 85 64 81 65 80 70 75

   First we arrange the series according to ascending order.

   50 51 60 64 65 70 75 80 81 85 90 95 97

   Median = (N+1)2nd item
   = (13+1)2nd item
   = (14/2)th item
   = (7)th item
   = 75

   Median for distribution with even number of items

2. Find the median of the following data.

   95 51 91 60 9064 85 69 80 70 78 75

   First we arrange the series according to ascending order.

   51 60 64 69 70 75 78 80 8590 91 95

   Median = (N +1)/2th item
   = (12+1)/2th item
   = (13/2)th item
   = (6.5)th item
   = (6th item + 7th item)/2
   = (75 + 78)/2
   = 153/2
   = 76.5

Median of discrete series

To find the median of a grouped series, we first of all, cumulate the frequencies. Locate median at the size of (N+1)/2th cumulative frequency. N is the cumulative frequency taken.

Steps:

- Arrange the values of the data in ascending order of magnitude.
- Find out cumulative frequencies
- Apply the formula (N=1)/2th item
- Look at the cumulative frequency column and find the value of the variable corresponding to the above.

Find median for the following data.

Income	:	100	150	80	200	250	180
Number of persons	:	24	26	16	20	6	30

First of all arrange the data according to ascending order.

Income	Frequency	Cumulative Frequency
80	16	16
100	24	40
150	26	(N=1)/2 66
180	30	96
200	20	116
250	6	122

Median = (N+1)2th item

= (122+1)/2th item

= (123)/2th item

= (61.5)th item

= value at the 61.5 cumulative frequency is taken as median

Therefore,

Median = 150

Median for continuous series

To find the median of a grouped series, with class interval, we should first of all cumulate the frequencies. Then we should locate the median at the size of (N)/2th cumulative frequency. Now apply the interpolation formula to obtain the median.

Median = L1 + (N/2–m)/fXC

L1 = Lower limit of the median class

N/2 = Cumulative frequency/2

m = Cumulative frequency of the class preceding the median class

F = Frequency of the median class

C = Class interval

Find the median of the following data.

Class : 12–14 15–17 18–20 21–23 24–26

Frequency : 1 3 8 2 6

Class	Frequency	CF
12–14	1	1
15–17	3	4
18–20	8	12(N/2=10)
21–23	2	14
24–26	6	20

Median = L1 + (n/2 = m)/fXC

L1 = 18

N/2 = 10

M = 4

F = 8

C = 2

= 18 + (10 – 4)/8X2

= 18 + 6/8X2

= 18 + (12/8)

= 18 + 1.5

= 19.5

Advantages of median

The following are the advantages of median:

- The value of the median can be determined graphically.
- Averages for data which has open-ended classes, can be calculated easily.
- It is the most convenient and the easiest measure of central tendency.
- The extreme values of the data do not affect the median.
- With the help of the median, the values of the data are divided equally to either of its sides.

NOTES

- It is not only simple to calculate, but also very simple to understand.

Disadvantages of median

The following are the disadvantages of median:

- Simple and very little fluctuation also affects the value of the median.
- Further algebraic treatment of the median is not possible.
- It is not a true representation of all the values since its value is determined by observation.
- The calculation of median requires the data to be arranged in an ascending order. This is very tedious when there are too many items in a series.

Mode

In business, mode is a popular measure of central tendency. It is that value of a distribution which repeats itself the maximum number of times. When one item repeats itself more than any other item, or in cases where two items repeat themselves an equal number of times, the mode is ill-defined. In such a case, mode is calculated using the following formula:

Mode = 3 median – 2 median

Advantages of mode

The following are the advantages of mode:

- Mode is the most typical frequency value of the distribution.
- It is not affected by extreme values.
- Mode can be determined even for series with open-ended classes.
- Mode can be graphically determined.

Disadvantages of mode

The following are the disadvantages of mode:

- It is difficult to calculate mode when one item repeats itself more than others.
- Mode is not capable of further algebraic treatment.
- Mode is not based on all the items of the series.
- Mode is not rigidly defined. There are several formulae for calculating mode.

Mode of individual series

1. Calculation of mode for the following data:

   7 10 8 5 8 6 8 9

   Since item 8 repeats itself more than the other items, mode = 8

NOTES

2. Calculation of mode for the following data:

   15 25 14 18 21 16 19 20

   Since no item repeats itself more than once, mode is ill defined.

   Mode = (3 median – 2 median)

   Mean = 18.5

   Median = (18+19)/2

   = 18.5

   Mode = (3X18.5) – (2X18.5)

   = 55.5 – 36.5 = 19

Mode of discrete data series

In discrete series the item with highest frequency is taken as mode.

3. Find mode for the following data.

Size of shirt	No. of persons
28	10
29	20
30	40
31	65
32	50
33	15
34	5

Since 65 is the highest frequency, its size is taken as mode.

Mode = 31

Calculation of mode using grouping table and analysis table

To make grouping table:

- Group the frequency in two.
- Frequencies are grouped in two leaving the first frequency.
- Group the frequency in three.
- Frequencies are grouped in three leaving the first frequency.
- Frequencies are grouped in three leaving the first and second frequency.

To make analysis table:

- Analysis table is made on the basis of the grouping table.
- Circle the highest value of each column.
- Assign marks to classes, which constitute the highest value of the column.
- Count the number of marks.

NOTES

- The class with the highest marks is selected as the model class.
- Apply the interpretation formula and find the mode.

Mode = L1+(F1-F0/2F1-F0-F2)XC

L1 = Lower limit of the model class

F1 = Frequency of the model class

F0 = Frequency of the class preceding the model class

F2 = Frequency of the class succeeding the model class

C = Class interval

Illustration:

Find mode for the following data using grouping table and analysis table.

Expenditure	0-20	20-40	40-60	60-80	80-100	100-120	120-140
No. of families	14	15	27	13	12	17	2

Grouping table

Class	Frequency	I	II	III	IV	V
0-20	14					
20-40	15	29				
40-60	27		42	56		
60-80	13	40			55	
80-100	12		25			52
100-120	17	29		42		
120-140	2		29		31	

Steps:

- In column I, the frequencies are grouped in two.
- In column II, frequencies are grouped in two, leaving the first frequency.
- In column III, frequencies are grouped in three.
- In column IV frequencies are grouped in three, leaving the first frequency.
- In column V frequencies are grouped in three, leaving the first and second frequency.

Analysis table:

Class	Frequency	I	II	III	IV	V	Total
0-20	14			1			1
20-40	15		1	1	1		3
40-60	27	1	1	1	1	1	5
60-80	13	1			1	1	4
80-100	12					1	1
100-120	17						0
120-140	2						0

The highest mark is 5 and is obtained by the class 4 – 60.

Therefore, model class = 40 – 60

Mode is calculated by the following formula:

Mode = L1+(F1 – F0)/(2F1 – F0 – F2)XC

L1 = Lower limit of the model class = 40

F1 = Frequency of the model class = 27

F0 = Frequency of the class preceding the model class = 15

F2 = Frequency of the class succeeding the model class = 13

C = Class interval = 20

Mode = 40 +(27 – 15)/(2X27 – 15 – 13)X20

= 40 + (12/54 – 28)20

= 40 + (12/26)20

= 40 + (.4615)20

= 40 + 9.23

= 49.23

3.4.2 Measures of Dispersion

Dispersion is the tendency of the individual values in a distribution to spread away from the average. Many economic variables like income, wages, etc., are widely varied from the mean. Dispersion is a statistical measure, which understands the degree of variation of items from the average.

Objectives of measuring dispersion

Study of dispersion is needed to:

- Test the reliability of the average
- Control variability of the data
- Enable comparison with two or more distribution with regard to their variability
- Facilitate the use of other statistical measures

Measures of dispersion point out as to how far the average value is representative of the individual items. If the dispersion value is small, the average tends to closely represent the individual values and is reliable. When dispersion is large, the average is not a typical representative value. Measures of dispersion are useful in controlling the cause of variation. In industrial production, efficient operation requires control of quality variation. Measures of variation enable comparison of two or more series with regard to their variability. A high degree of variation would mean little consistency and low degree of variation would mean high consistency.

NOTES

Properties of a good measure of dispersion

The following are the properties of a good measure of dispersion:

- Good measures of dispersion should be simple to understand.
- They should be easy to calculate.
- They should be rigidly defined.
- They should be based on all the values of a distribution.
- They should be amenable to further statistical and algebraic treatment.
- They should have sampling stability.
- They should not be unduly affected by extreme values.

Measures of dispersion

The following are the measures of dispersion:

- Range
- Quartile deviation
- Mean deviation
- Standard deviation
- Lorenz curve

Range, quartile deviation, mean deviation and standard deviation are mathematical measures of dispersion. Lorenz curve is a graphical measure of dispersion.

Measures of dispersion can be absolute or relative. An absolute measure of dispersion is expressed in the same chapter of the original data. When two sets of data are expressed in different units, relative measures of dispersion are used for comparison. A relative measure of dispersion is the ratio of absolute measure to an appropriate average.

The following are the important relative measures of dispersion:

- Coefficient of range
- Coefficient of quartile deviation
- Coefficient of mean deviation
- Coefficient of standard deviation

Range

Range is the difference between the lowest and the highest value.

Symbolically,

Range = Highest value – Lowest value

Range = H-L

H = Highest value

L = Lowest value

Relative measure of dispersion is co-efficient of range. It is obtained by the following formula:

Coefficient of range = (H – L)/(H + L)

1. Calculate the range of the following distribution, showing income of 10 workers. Also calculate the co-efficient of range.

25 37 40 23 58 75 89 20 81 95

Range = H-L

H = Highest value = 95

L = Lowest value = 20

Range = 95 – 20 = 75

Coefficient of range = (H-L)/(H+L)

= (95-20)/(95+20)

= 75/115

= .6521

Range is simple to understand and easy to calculate, but it is not based on all items of the distribution. It is subject to fluctuations from sample to sample. Range cannot be calculated for open-ended series.

Quartile deviation

Quartile deviation is defined as inter-quartile range. It is based on the first and the third quartile of a distribution. When a distribution is divided into four equal parts, we obtain four quartile, Q1, Q2, Q3, and Q4.

The first quartile Q1 is point of the distribution where 25 per cent of the items of the distribution lie below Q1 and 75 per cent of the items of the distribution lie above the Q1. Q2 is the median of the distribution, where 50 per cent of the items of the distribution lie below Q2 and 50 per cent of the items of the distribution lie above Q2. The third quartile Q3 is point of the distribution where 75 per cent of the items of the distribution lie below Q3 and 25 per cent of the items of the distribution lie above the Q3.

Quartile deviation is based on the difference between the third and first quartile. So quartile deviation is defined as the inter-quartile range. Symbolically,

Inter-quartile range = Q3 – Q1

Quartile deviation = (Q3 – Q1)/2

Co-efficient of quartile deviation = (Q3 – Q1)/(Q3 + Q1)

NOTES

Advantages of quartile deviation

The following are the advantages of quartile deviation:

- Quartile deviation is superior to range as a rough measure of dispersion.
- It has a special merit in measuring dispersion in open-ended series.
- Quartile deviation is not affected by exchange values.

Limitations of quartile deviation

The following are the limitations of quartile deviation:

- Quartile deviation ignores the first 25 per cent of the distribution below Q1 and 25 per cent of the distribution above the Q3.
- Quartile deviation is not open to further mathematical treatment.
- Quartile deviation is affected a lot by sampling fluctuations.

Problems

Individual series

1. Find the quartile deviation and its co-efficient.

   20 58 40 12 30 15 50

   First of all arrange the data in ascending order.

   12 15 20 28 30 40 50

   Q1 = Size of (N+1)/4th item
   = Size of (7+1)/4th item
   = Size of (8/4)th item
   = 2nd item
   = 15

   Q3 = Size of 3 X (N+1)4th item
   = Size of 3 X (7+1)/4th item
   = Size of 3 X8/4th item
   = (3X2)nd item
   = 6th item
   = 40

   Co-efficient Deviation = (Q3 – Q1)/(Q3 + 1)
   = (40 – 15)/(40 + 15)
   = 25/55
   = .4545

Discrete series

2. Find quartile deviation and its co-efficient for the following data.

Income : 110 120 130 140 150 160 170 180 190 200

Frequency: 50 45 40 35 30 25 20 15 10 5

Income	Frequency	CF
110	50	50
120	45	95(N+1)4th item=69=120
130	40	135
140	35	170
150	30	200
160	25	225 3(3+1)/4th item =207=160
170	20	245
180	15	260
190	10	270
200	5	275

Q1 = Size of (N+1)/4th item

= size of (275+1)/4th item

= size of (278/4)th item

= Size of 69th cumulative frequency

= 120

Q3 = Size of 3(N+1)/4th item

= size of 3X(275+1)/4th item

= size of 3X69th item

= size of 207th cumulative frequency

= 160

Quartile Deviation = (160-120)/2

= 40/2

= 20

Co-efficient of Quartile Deviation = (Q3-Q1)/(Q3+Q1)

= (160-120)/(160+120)

= 20/280

= .0714

Continuous series

Find quartile deviation for the following series:

Marks : 0–20 20–40 40–60 60–80 80–100

Frequency : 10 30 36 30 14

Income	Frequency	CF
0-20	10	10
20-40	30	40 (N)/4th class = 20-40
40-60	36	76
60-80	30	106 3(N)/4th class = 60-80
80-100	14	120

Q1 = Lies in (N)/4th class

= lies in (120)/4th class

= lies in (30)th cumulative frequency class

= lies in 20-40

Q1 can be obtained by applying the interpolation formula

= L1+(N/4)-m/fXC

= 20+(30-10)/30X20

= 20+400/30

= 20+13.33

= 33.33

Q3 = lies in 3(30)th cumulative frequency class

= lies in 60-80 class

Q3 can be obtained by applying the interpolation formula

= L1+3(N/4)-m/fXC

= 60+(90-76)/30X20

= 60+(14/30)X20

= 60+280/30

= 60+9.33

= 69.33

Quartile deviation = (Q3-Q1)/2

= (69.33-33.33)2

= 36/2

= 18

Co-efficient of Quartile deviation = (Q3-Q1)/(Q3+Q1)

= (69.33-33.33)/(69.33+33.33)

= 36/102.66

= .3505

Mean deviation

Range and quartile deviation do not show any scattering from the average. However, mean deviation and standard deviation helps us achieve the dispersion.

Mean deviation is the average of the deviations of the items in a distribution from an appropriate average. Thus, we calculate mean deviation from mean, median or mode. Theoretically, mean deviation from median has an advantage because the sum of the deviations items from median is the minimum when signs are ignored. However, in practice, mean deviation from mean is frequently used. That is why it is commonly called mean deviation.

Formula for calculating mean deviation = $\Sigma D/N$

Where,

ΣD = Sum of the deviation of the items from mean, median or mode

N = Number of items

Steps

The following are the steps in the calculation of mean deviation:

- Calculate mean, median and mode of the series.
- Find the deviation of items from the mean, median or mode.
- Add the deviations and obtain D.
- Take the average of the deviations (D/N) which is the mean deviation. The coefficient of mean deviation is the relative measure of mean deviation. It is obtained by dividing the mean deviation by a particular measure of average used for measuring mean deviation.

  If mean deviation is obtained from median, the coefficient of mean deviation is obtained by dividing mean deviation by the median.

  The co-efficient of mean deviation = mean deviation/median

  If mean deviation is obtained from mean, the co-efficient of mean deviation is obtained by dividing mean deviation by mean.

  The co-efficient of mean deviation = mean deviation/mean

  If mean deviation is obtained from mode, the co-efficient of mean deviation is obtained by dividing mean deviation by mode.

  The co-efficient of mean deviation = mean deviation/ mode

Problems

Calculate mean deviation for the following data from the mean:

Daily wages:	15	18	20	25	30	35	40	42	45
Frequency :	2	3	5	10	12	10	5	2	1

Daily wages	Frequency	f.x	D=x-20	Fd
15	2	30	5	10
18	3	54	2	6
20	5	100	0	0
25	10	250	5	50
30	12	360	10	120
35	10	250	15	150
40	5	200	20	100
42	2	84	22	44
45	1	45	25	25
	50	1473		505

Mean = 1473/50

= 20

Mean deviation = ΣfD/N

= 505/50

= 10.1

The co-efficient of mean deviation

= mean deviation/mean

= 10.1/20

= .505

Continuous series

The procedure remains the same. The only difference is that we have to obtain the midpoints of the various classes and take deviations of these midpoints. The deviations are multiplied by their corresponding frequencies. The value so obtained is added and its average is the mean deviation.

Calculate mean deviation for the following data.

Class : 5–10 10–15 15–20 20–25 25–30 30–35 35–40 40–45

Frequency: 6 5 15 10 5 4 3 2

Class	Frequency	Mid x	d	fd	D=x-28.8	fD
5-10	6	7.5	-15	-90	21.5	129
10-15	5	12.5	-10	-50	16.3	81.5
15-20	15	17.5	-5	-75	11.3	169.5
20-25	10	22.5	0	0	6.3	63
25-30	5	27.5	5	25	1.3	6.5
30-35	4	32.5	10	40	3.7	14.8
35-40	3	37.5	15	45	8.7	26.1
40-45	2	42.5	20	40	13.7	27.4
	50			-65		516.6

Arithmetic mean = A+Σfx/Σf

= 22.5 + 65/50

= 22.5 + 1.3

= 28.8

NOTES

Mean deviation from mean = ΣfD/N
= 516.6/50
= 10.332

The co-efficient of mean deviation = mean deviation/mean
= 10.332/28.8
= .3762

Mean deviation from median

To find median

Class	Frequency	CF	Mid x	D=X-17	
5-10	6	6	7.5	9.5	57
10-15	5	11	12.5	4.5	22.5
15-20	15	26 (N/2)=25	17.5	.5	7.5
20-25	10	36	22.5	5.5	55
25-30	5	41	27.5	10.5	52.5
30-35	4	45	32.5	15.5	62
35-40	3	48	37.5	20.5	61.5
40-45	2	50	42.5	25.5	51
	50				369

Median = L1+(n/2-m/f)C
= 15+25-11/15X5
= 15+6/15X5
= 15+30/15
= 15+2
= 17

Mean deviation from median = ΣfD/N
= 369/50
= 7.38/17
= 0.434

Mean deviation from mode = model class 15-20
= L1+(F1-F0/2F1-F0-F2)C
= 15+(15-5/2X15-5-10)X5
= 15+(10/15)X5
= 15+3.33
= 18.33

Class	Frequency	Mid X	D=X-18.33	fD
5-10	6	7.5	10.83	64.98
10-15	5	12.5	5.83	29.15
15-20	15	17.5	.83	12.45
20-25	10	22.5	4.17	41.7
25-30	5	27.5	9.17	45.85
30-35	4	32.5	14.17	56.68
35-40	3	37.5	19.17	57.57
40-45	2	42.5	24.17	48.34
	50			356.72

Mean deviation from mode = ΣfD/N

= 356.72/50

= 7.13

The co-efficient of mean deviation = mean deviation/mode

= 7.16/18.3

= 0.3912

Advantages of mean deviation

The following are the advantages of mean deviation:

- Mean deviation is simple to understand and easy to calculate.
- It is based on each and every item of the distribution.
- It is less affected by the values of extreme items compared to standard deviation.
- Since deviations are taken from a central value, comparison about formation of different distribution can be easily made.

Limitations of mean deviation

The following are the limitations of mean deviation:

- Algebraic signs are ignored while taking the deviations of the items.
- Mean deviation gives the best result when it is calculated from the median but the median is not a satisfactory measure when variability is very high.
- Various methods give different results.
- It is not capable of further mathematical treatment.
- It is rarely used for sociological studies.

Standard deviation

Standard deviation is the most important measure of dispersion. It satisfies most of the properties of a good measure of dispersion. It was introduced by Karl Pearson in 1893. Standard deviation is defined as the mean of the squared deviations from the arithmetic mean.

Mean deviation and standard deviation are calculated from deviation of each and every item. Standard deviation is different from mean deviation in two respects. First of all, algebraic signs are ignored in calculating mean deviation. Secondly, signs are taken into account in calculating standard deviation while mean deviation can be found from mean, median or mode. Standard deviation is found only from mean.

Standard deviation can be computed in the following two methods:

(i) By taking deviations from actual mean

(ii) By taking deviation from assumed mean

Formula for finding standard deviation is:

The most common estimator for ó used is an adjusted version, the **sample standard deviation**, denoted by *s* and defined as follows:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_1\bar{x})^2},$$

where are the observed values of the sample items and is the mean value of these observations

Steps

The following steps are used to compute the mean:

- Calculate the actual mean of the series Σx/N
- Take deviation of the items from the mean (x – x)
- Find the square of the deviation from actual mean (x – x)2/N
- Sum the squares of the deviation $\Sigma(x - x)^2$
- Find the average of the squares of the deviations Σ(x – x)2/N
- Take the square root of the average of the sum of the deviation

Problems

1. Calculate the standard deviation of the following data:

   49 50 65 58 42 60 51 48 68 59

   Standard deviation from actual mean

   Arithmetic mean = Σx/N

   = 550/10

   = 55

Values	(x – 55)	(x – 55)2
49	–6	36
50	–5	25
65	10	100
58	3	9
42	–13	169
60	5	25
51	–4	16
48	–7	49
68	13	169
59	4	16
550		$\Sigma(x - x)^2 = 614$

$$\text{S.D.} = \sqrt{\Sigma(x - x)^2/N}$$
$$= \sqrt{614/10}$$
$$= \sqrt{61.4}$$
$$= 7.836$$

Standard deviation from assumed mean

Assumed mean = 50

Values	(x – 50)	$(x - 50)^2$
49	–1	1
50	0	0
65	15	225
58	8	64
42	–8	64
60	10	100
51	1	1
48	–2	4
68	18	324
59	9	81
550	$\Sigma(x - x) = 50$	$\Sigma (x - x)^2 = 864$

$$\text{S.D.} = \sqrt{\Sigma(x - x)^2/N - \Sigma\{(x - x)/N\}^2}$$
$$= \sqrt{614/10 - (50/10)^2}$$
$$= \sqrt{86.4 - 25}$$
$$= \sqrt{61.4}$$
$$= 7.836$$

Discrete series

Standard deviation can be obtained by the following three methods:

(i) Direct method

(ii) Short cut method

(iii) Step deviation

(i) Direct method

Under this method, the formula is:

$$S.D. = \sqrt{\Sigma(fx)^2/N - \Sigma\{(fx)/N\}^2}$$

Calculate standard deviation for the following frequencies distribution.

Marks	:	20	30	40	50	60	70
Frequency	:	8	12	20	10	6	4

Marks	Frequency	x^2	fx	fx2
20	8	400	160	3200
30	12	900	360	10800
40	20	1600	800	32000
50	10	2500	500	25000
60	6	3600	360	21600
70	4	4900	280	19600
	60		2460	112200

S.D. =

$= \Sigma 112200/60 - \Sigma\{2460/60\}^2$

$= \Sigma 1870 - 1681$

$= \Sigma 189$

$= 13.747$

3.4.3 Correlation Analysis

Business and economic variables are associated with each other. For example, supply and demand of a product is always connected to its price. As the price of the commodity falls, its demand immediately goes up. Thus, it can be said that the price and demand of a commodity are related inversely to each other or are negatively correlated. Also, the supply of a commodity goes up with the increase in price and the supply falls with the decrease in price. Thus, it can be said that price and supply are related to each other directly or are positively co-related. Hence it can be said that correlation shows the connection between two such variables in which changes in the value of one variable is accompanied with a change in the value of other variable.

According to L.R. Connor, 'if two or more quantities vary in sympathy so that movements in the one tend to be accomplished by corresponding movements in the other(s) they are said to be correlated.' According to W.I. King, 'Correlation means that between two series or group of data, there exits some causal connection.'

Thus, correlation is a statistical device which studies the relationship between two or more variables. Also, if any two variables are correlated, the change in the value of any one of them will result in a corresponding change in the value of the other.

Correlation and causation

Although the term correlation is used in the sense of mutual dependence of two or more variable, it is not always necessary that they have cause and effect relation. Even a high degree of correlation between two variables does not necessarily indicate a cause and effect relationship between them.

- Cause and effect relationships are being affected by a third variable. For instance, price of rice and price of sugar are affected by rainfall. Here there may not be any cause and effect relation between price of rice and price of sugar.
- Both the correlated variables are being affected by a third variable. For instance, price of rice and price of sugar are affected by rainfall. Here there may not be any cause and effect relation between price of rice and price of sugar.
- Related variable may be mutually affecting each other so that none of them is either a cause or an effect. Demand may be the result of price. There are cases when price rise due to increased demand.
- The correlation may be due to chance. For instance, a small sample may show correlation between wages and productivity, i.e., higher wage leading to lower productivity. In real life it need not be true. Such correlation is due to chance.
- There might be a situation of spurious correlation between two variables. For instance, relationship between number of divorces and television exports may be correlated. There cannot be any relationship between divorce and exports of television.

These points make it clear that correlation is only a statistical relationship and it does not necessarily signify a cause and effect relationship between the variables.

Types of correlation analysis

Correlation can be:

- Positive or negative
- Linear or non-linear
- Simple, multiple or partial

Positive and negative correlation

When values of two variables move in the same direction, correlation is said to be positive. When prices rise, supply increases and when prices fall supply decreases.

If, on the other hand, values of two variables move in the opposite direction, correlation is said to be negative. When price rises, demand

decreases and when prices fall, demand increases. In this case, an increase in the value of one variable on an average results in a decrease in the value of the other variable.

Linear and non-linear correlation

When a change in one variable leads to a constant ratio of change in the other variable, correlation is said to be **linear**. In case of linear correlation, points of correlation plotted on a graph will give a straight line. Correlation is said to be non-linear when change in one variable is not accompanied by a constant ratio of change in another variable. In case of non-linear correlation, points of correlation plotted on a graph do not give a straight line. It is called **curvilinear correlation** because the graph of such a correlation results in a curve.

Simple, partial and multiple correlations

Simple correlation studies the relationship between two variables only. For instance, correlation between price and demand is simple as only two variables are studied in this case. Multiple correlation studies the relationship of one variable with many variables. For instance, correlation of agricultural production with rainfall, fertilizer use and seed quality is a multiple correlation. Partial correlation studies the relationship of a variable with one of the many variables with which it is related. For instance, seed quality, temperature and rainfall are three variables which determine yield of a crop. In this case, yield and rainfall is a partial correlation.

Utility of correlation

Study of correlation is of immense practical use in business and economics.

- Correlation analysis enables us to measure the magnitude of one variable on the basis of other variables and studies.
- Once we establish correlation, we can estimate the values of one variable on the basis of the other. This is done with the help of regression equations.
- The correlation study is useful for formulation of economic policies. In economics, we are interested in finding the important dependent variables on the basis of independent variable.
- Correlation study helps us make relatively more dependable forecasts.

Methods of studying correlation

The following methods are used in the study of correlation:

- Scatter diagram
- Karl Pearson's method of correlation

NOTES

- Spearman's rank correlation method
- Concurrent deviation method

Scatter diagram

Scatter diagram is a graphical method of studying correlation between two variables. In a scatter diagram, one variable is measured on the x-axis and the other is measured on the y-axis of the graph. Each pair of values is plotted on the graph by means of dot marks. If plotted points do not show any trend, the two variables are not correlated. If the trend shows upward rising movement, correlation is positive. If the trend is downward sloping, correlation is negative.

Karl Pearson's co-efficient of correlation

Karl Pearson's co-efficient of correlation is a mathematical method for measuring correlation. Karl Pearson developed the correlation from the covariance between two sets of variables. Karl Pearson's co-efficient of correlation is denoted by the symbol 'r'. The formula for obtaining Karl Pearson's co-efficient of correlation is:

Direct method:

r = covariance between x and y

$\sqrt{SD_X}\sqrt{SD_Y}$

Covariance between x and y = $\Sigma xy/N - (\Sigma x/N \times \Sigma y/N)$

SD_x = standard deviation of x series = $\sqrt{(\Sigma x^2/N) - (\Sigma x/N)^2}$

SD_y = standard deviation of y series = $\sqrt{(\Sigma dy^2\ (\Sigma y^2/N) - (\Sigma y/N)^2}$

Short method using assumed mean:

If short cut method is used using assumed mean, the formula for obtaining Karl Pearson's co-Efficient of correlation is:

Covariance between x and y = $\Sigma dxdy/N - (\Sigma dxN\ X\ \Sigma dy/N)$

$SD_x = \sqrt{(\Sigma dx^2/N) - (\Sigma dx/N)^2}$

$SD_y = \sqrt{(\Sigma dy^2/N) - (\Sigma dy/N)^2}$

$r = \Sigma dxdy/N - (\Sigma dxN\ X\ \Sigma dy/N)$

$\sqrt{(\Sigma dx^2/N) - (\Sigma dx/N)^2}\ X\ \Sigma(\Sigma dy^2/N) - (\Sigma dy/N)^2$

The steps in calculating Karl Pearson's correlation co-efficient using short cut method is as follows:

- Assume mean of x and y series.
- Take deviations of x and y series from assumed mean and get Σdx and Σdy.

- Square of dx and dy and find the sum of squares and get Σdx^2 and Σdy^2
- Multiply the corresponding deviations of x and y series and total the products to get $\Sigma dxdy$

If the deviations are taken from the arithmetic mean $\Sigma dx = 0$ and $\Sigma dy = 0$, then formula becomes:

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2}\sqrt{\Sigma dy^2}}$$

Interpreting co-efficient of correlation

If the short cut method is used using actual mean, the formula for obtaining Karl Pearson's co-efficient of correlation is:

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2}\sqrt{\Sigma dy^2}}$$

Interpreting co-efficient of correlation

The co-efficient of correlation measures the correlation between two variables. The value of co-efficient of correlation always lies between +1 and –1. It can be interpreted in the following ways:

- If the value of co-efficient of correlation r is 1, it is interpreted as perfect positive correlation.
- If the value of co-efficient of correlation r is –1, it is interpreted as perfect negative correlation.
- If the value of co-efficient of correlation r is $0 < r < 0.5$, it is interpreted as perfect poor positive correlation.
- If the value of co-efficient of correlation r is $0.5 < r < 1$, it is interpreted as perfect good positive correlation.
- If the value of co-efficient of correlation r is $0 > r > –0.5$, it is interpreted as perfect poor negative correlation.
- If the value of co-efficient of correlation r is $–0.5 > r > –1$, it is interpreted as good negative correlation.
- If the value of co-efficient of correlation r is 0 it is interpreted as zero correlation.

Probable error of correlation is estimated to find out the extent to which the value of r is dependable. If probable error is added to or subtracted from the correlation coefficient, it would give such limits within which we can reasonably expect the value of correlation to vary. If the coefficient of correlation is less than the probable error it will not be significant. If the coefficient of correlation r is more than six times the probable error, correlation is definitely significant. If probable error is 0.5 or more, it is generally

NOTES

considered as significant. Probable error is estimated by the following formula:

$$PE = 0.6745(1 - r^2/\sqrt{N})$$

3.4.4 Coefficient of Determination

Besides probable error, another important method of interpreting co-efficient of correlation is the coefficient of determination. Coefficient of determination is the square of correlation or r^2. For instance, suppose the coefficient of correlation between price and supply is 0.8, we calculate the coefficient of determination as r^2, which is $.8^2$ or .64. This means that 64 per cent of the variation in supply is on account of changes in price.

Spearman's rank correlation method

Charles Edward Spearman, a British psychologist, devised a method for measuring correlation between two variables based on ranks given to the observations. This method is adopted when the variables are not capable of quantitative measurements like intelligence, beauty, etc. It is in such cases that rank correlation is useful.

Spearman's rank correlation coefficient is given by,

$$r_k = 1 - 6"D^2/n(n2 - 1)$$

Where,

D is the difference between ranks

n is number of pairs correlated

Concurrent Deviation Method

In the concurrent deviation method, correlation is calculated between direction of deviations and not their magnitudes. As such only the direction of deviations is taken into account in the calculation of this coefficient and their magnitude is ignored.

The formula for the calculation of coefficient of concurrent deviations is as follows:

$$r_c = +- \sqrt{|2C - n/n|}$$

Steps in the calculation of concurrent deviation

The following are the steps in the calculation of concurrent deviation:

- Find out the direction of change of x-variable. When a successive figure in the series increases in direction it is marked '+' and when a successive figure in the series decreases in direction it is marked '"' and is denoted by dx.
- Find out the change in direction of y-variable. It is denoted by dy.

- Multiply dx and dy and determine the value of C. C is the number of positive products of dxdy.

  (–X – or + X +)

- Use the formula '$r_c = +– \sqrt{|2C – n/n|}$' to obtain the value of coefficient of r_c .

Problems

1. Calculate Karl Pearson's co-efficient of correlation for the following data:

X : 43 44 46 40 44 42 45 42 38 40 42 57

Y : 29 31 19 18 19 27 27 29 41 30 26 10

X	Y	Dx	Dy	Dx2	Dy2	Dxdy
43	29	3	–1	9	1	3
44	31	4	1	16	1	4
46	19	6	–11	36	121	–66
A(40)	18	0	–12	0	144	0
44	19	4	–11	16	121	–44
42	27	2	–3	4	9	–6
45	27	5	–3	25	9	–15
42	29	2	–1	4	1	
38	41	-2	11	4	121	–22
40	A(30)	0	0	0	0	0
42	26	2	–4	4	16	–8
57	10	17	–20	289	400	–340
		43	54	407	944	494

Direct Method

$$r = \frac{\text{covariance between x and y}}{\sqrt{SD_X}\sqrt{SD_Y}}$$

Covariance between x and y $= \Sigma dxdy/N – (\Sigma dxN \text{ X } \Sigma dy/N)$

$SD_x = \sqrt{(\Sigma dx^2/N) – (\Sigma dx/N)^2}$

$SD_y = \sqrt{(\Sigma dy^2/N) – (\Sigma dy/N)^2}$

Short cut method using assumed mean

$$r = \frac{\text{covariance between x and y}}{\Sigma D_X \times \Sigma D_Y}$$

NOTES

Covariance between x and y = $\Sigma dxdy/N - (\Sigma dxN \text{ X } \Sigma dy/N)$

$SD_x = \sqrt{(\Sigma dx^2/N) - (\Sigma dx/N)^2}$

$SD_y = \sqrt{(\Sigma dy^2/N) - (\Sigma dy/N)^2}$

$r = \Sigma dxdy/N - (\Sigma dxN \text{ X } \Sigma dy/N)$

$\sqrt{(\Sigma dx^2/N) - (\Sigma dx/N)^2} \text{ X } \Sigma(\Sigma dy^2/N) - (\Sigma dy/N)^2$

$\Sigma dxdy = 494$

$N = 12$

$\Sigma dx = 43$

$\Sigma D_Y = 54$

$\Sigma dx^2 = 407$

$\Sigma dy^2 = 944$

494/12 (43/12 X 54/12)

$\Sigma 407/12 - (43/12)^2 \text{ X } \Sigma 944/12 - (54/12)^2$

41.17 – (3.58 X 4.5)

Σ33.96 – 12.91 X Σ78.66 – 20.25

41.16 – 16.11

Σ21.09 Σ58.41

25.05

o X 7.64

25.05

35.08

= 0.714

Interpretation: There is a good positive correlation between x and y variable.

CHECK YOUR PROGRESS

7. List the important characteristics of statistical data.
8. List the various measures of dispersion.

3.5 SUMMARY

- Data is available in various types and forms. Interpreting this data is not easy. Most researches, especially social science researches, come to conclusions by using both primary and secondary data.

- Data processing is an intermediary stage of work between data collection and data interpretation. The data gathered in the form of questionnaires/interviews/schedules/field notes/data sheets is mostly in the form of large volume of research variables.
- In the data preparation step, the data is prepared in a data format, which allows the analyst to use modern analysis software such as SAS or SPSS. The major criterion in this is to define the data structure.
- The next step in the processing of the data is editing of the data instruments. Editing is the process of checking to detect and correct errors and omissions. Data editing happens at two stages, one at the time of recording of the data and second at the time of analysis of data.
- The first step of editing is to check whether there is an answer to all the questions/variables set out in the data set. If there were an omission, the researcher sometimes would be able to deduce the correct answer from other related data on the same instrument.
- The edited data is then subject to codification and classification. The coding process assigns numerals or other symbols to the several numerals or other symbols to the responses of the data set.
- When open-ended responses have been received, classification is necessary to code the responses.
- The transcription of data can be used to summarize and arrange the data in a compact form for further analysis. This process is called tabulation. Thus, tabulation is the process of summarizing raw data and displaying them in compact statistical tables for further analysis.
- Variables that are classified according to magnitude or size are often arranged in the form of a frequency table. While constructing this table, it is necessary to determine the number of class intervals to be used and the size of the class intervals.
- One-way frequency tables present the distribution of cases on only a single dimension or variable. For example, the distribution of respondents by gender, religion and social or economic status are shown in one-way tables.
- In presenting the data of frequency distributors and statistical computations, it is often desirable to use appropriate forms of graphical presentations. In addition to tabular forms, graphics presentation involves the use of graphics, charts and other pictorial devices such as diagrams.
- The line graph is useful for showing changes in data relationship over a period of time. In this graph, figures are plotted in relation to two intersecting lines or axes. The horizontal line is called the abscissa or X-axis and the vertical, the ordinal or Y-axis.

- Data can be either qualitative or quantitative in nature. To analyse data, we can use the various measures of central tendency, measures of dispersion, correlation analysis and the coefficient of determination.
- Analysis of data involves understanding the characteristics of the data. The following are the important characteristics of statistical data:
  - o Central tendency
  - o Dispersion
  - o Skewness
  - o Kurtosis
- Median is the middlemost item of a given series. In individual series, we arrange the given data according to ascending or descending order and take the middlemost item as the median.
- In business, mode is a popular measure of central tendency. It is that value of a distribution which repeats itself the maximum number of times. When one item repeats itself more than any other item, or in cases where two items repeat themselves an equal number of times, the mode is ill-defined.
- Dispersion is the tendency of the individual values in a distribution to spread away from the average. Many economic variables like income, wages, etc., are widely varied from the mean.
- Quartile deviation is defined as inter-quartile range. It is based on the first and the third quartile of a distribution.
- Mean deviation is the average of the deviations of the items in a distribution from an appropriate average. Thus, we calculate mean deviation from mean, median or mode.
- Scatter diagram is a graphical method of studying correlation between two variables. In a scatter diagram, one variable is measured on the x-axis and the other is measured on the y-axis of the graph.

3.6 KEY TERMS

- **Editing:** It is the process of checking to detect and correct errors and omissions.
- **Tabulation:** It is the process of summarizing raw data and displaying them in compact statistical tables for further analysis.
- **Arithmetic mean:** It is the value obtained by dividing the sum of the item by the number of items in a series.
- **Dispersion:** It is the tendency of the individual values in a distribution to spread away from the average.

- **Scatter diagram:** It is a graphical method of studying correlation between two variables.

3.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Identification of the data structure, editing the data, coding and classifying the data, transcription of the data and its classification are the various steps in the processing of data.
2. Data is edited to check its completeness, accuracy and uniformity.
3. Numeric coding, alphabetic coding and zero coding are the various ways in which data can be coded.
4. Manual transcription and long worksheets are the two methods of transcription of data.
5. Manual tabulation can be tedious, slow and error-prone.
6. Heading, body and notations are the major components of a table.
7. Central tendency, dispersion, skewness and kurtosis are the important characteristics of statistical data.
8. Range, quartile deviation, mean deviation, standard deviation and Lorenz curve are the various measures of dispersion.

3.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on editing of data.
2. How is data tabulated manually?
3. What are one-way and two-way tables?
4. Write a short note on line graphs.
5. What are the main objectives of measuring dispersion?
6. List the advantages of mean deviation.

Long-Answer Questions

1. Explain the coding and classification of data.
2. What do you understand by transcription of data? Explain the various methods of data transcription.
3. Explain the construction of a frequency table in detail.
4. Discuss the various measures of central tendency with the help of examples.

5. Explain the various measures of dispersion in detail.
6. What are the various types of correlation analysis?

NOTES

3.9 FURTHER READING

Booth, Wayne. 2008. *The Craft of Research*, Third edition. Illinois: University of Chicago Press.

Creswall, John W. 2008. *Research Designs: Quantitative, Qualitative and Mixed Methods Approaches*. London: Sage Publications.

Christenson, Larry B. *et al.* 2010. *Research Methods, Design and Analysis*, Eleventh edition. New Jersey: Allyn and Bacon.

Kothari, C. R. 2008. *Research Methodology: Methods and Techniques*. New Delhi: New Age International.

Kumar, Ranjit. 2010. *Research Methodology: A Step-by-Step Guide for Beginners*, Third edition. New Delhi: Sage Publications.

UNIT 4 MEASUREMENT AND SCALING TECHNIQUES, REGRESSION AND FORECASTING

Structure

4.0 INTRODUCTION

The term 'measurement' means assigning numbers or some other symbols to the characteristics of certain objects. When numbers are used, the researcher must have a rule for assigning a number to an observation in a way that provides an accurate description. We do not measure the object but some characteristics of it. Therefore, in research, people/consumers are not measured; what is measured only are their perceptions, attitude or any other relevant characteristics. There are two reasons for which numbers are usually assigned. First of all, numbers permit statistical analysis of the resulting data and secondly, they facilitate the communication of measurement results.

As mentioned earlier, the numbering is done based on certain rules. Therefore, the assignment of numbers to the characteristics must be isomorphic, i.e., there must be a one-to-one correspondence between the numbers and the characteristics being measured.

For example, same rupee figures should be assigned to a household with identical annual income. Only then numbers can be associated with specific characteristics of the measured object and vice versa. Further, they must not change over the objects or time. This means that the rules for a given assignment must be invariant over time or the object being measured.

Scaling is an extension of measurement. Scaling involves creating a continuum on which measurements on objects are located. Suppose you want to measure the satisfaction level towards Kingfisher Airlines and a scale of 1 to 11 is used for the said purpose. This scale indicates the degree of dissatisfaction, with 1 = extremely dissatisfied and 11 = extremely satisfied. Measurement is the actual assignment of a number from 1 to 11 to each respondent whereas the scaling is the process of placing the respondent on a continuum with respect to their satisfaction towards Kingfisher Airlines. In this unit, you will learn the concept of scales and scaling techniques, regression and prediction, hypotheses testing, the basic concepts of computer, and the importance of report writing in research.

4.1 UNIT OBJECTIVES

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

- Assess measurement in research and the various measurement scales
- Describe scaling and the important scaling techniques
- Discuss the meaning of regression and prediction and the linear aggression analysis
- Explain hypothesis testing and the steps involved in it
- Discuss the meaning, characteristics, history and types of computers
- Analyse the importance of computers in research
- Assess the significance of report writing in research

4.2 MEASUREMENT IN RESEARCH

In daily life we are often required to measure physical objects in order to determine their weight, height or other physical features. We also measure when we judge how much we like a song, a painting or the personalities of our friends. We, thus, measure physical objects as well as abstract contacts. Measurement is a relatively complex and demanding task, especially when it concerns qualitative or abstract phenomena. By measurement we mean the process of assigning numbers to objects

or observation, the level of measurement being a function of the rules under which the numbers are assigned.

It is easy to assign numbers with respect to properties of some objects, but it is relatively difficult with respect to others. For instance, measuring things such as social conformity, intelligence, or marital adjustment is much less obvious and requires much closer attention than measuring physical weight, biological age or a person's financial assets. In other words, properties like weight, height, etc., can be measured directly with some standard methods of measurement, that it is not that easy to measure properties like motivation to succeed, ability to stand stress and the like. We can expect high accuracy in measuring the length of pipe with a yardstick, but if the concept is abstract and the measurement tools are not standardized, we are less confident about the accuracy of the results of measurement.

Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence. In measuring, we device some form of scale in the range (in terms of set theory, range may refer to some set) and then transform or map the properties of objects from the domain (in terms of set theory, domain may refer to some other set) onto this scale. For example, in case we are to calculate the male to female attendance ratio, while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex. In terms of set theory, this process is one of mapping the observed physical properties of those coming to the show (the domain) on to a sex classification (the range).The rule of correspondence is: if the object in the domain appears to be male assign to '0' and if female assign to '1'. Similarly, we can record a person's marital status as 1, 2, 3 or 4, depending on whether the person is single/married/widowed or divorced. We can record 'yes or no' answers as well to a question as '0' and '1' (or as 1 and 2 or perhaps as 59 and 60). In this artificial or nominal way, categorical data, (qualitative or descriptive) can be made into numerical data and if we thus code the various categories. We refer to the numbers we record as nominal data. Nominal data are numerical in name only, because they do not share any of the properties of the numbers we deal in ordinary arithmetic. For instance, if we record marital status as 1, 2, 3 or 4 as stated above, we cannot write $4 > 2$ or $3 < 4$ or $3 - 1 = 4 - 2$, $1 + 3 = 4$ or $4/2 = 2$.

In situations where we cannot do anything except set up inequalities, we refer to the data as ordinal data. For instance, if one mineral can scratch another, it receives a higher hardness number on the Mohs scale of mineral hardness. In this scale, the numbers 1 to 10 are assigned respectively to talc, gypsum, calcite, fluorite, apatite, feldspar, quartz, topaz, sapphire and diamond. Thus, we can say that apatite is harder than gypsum ($5 > 2$) or that feldspar is softer than sapphire ($6 < 9$). However, we cannot assume, for example, that $10 - 9 = 5 - 4$, because the difference in hardness between diamond (10) and sapphire (9) is actually much greater than that between apatite (5) and fluorite (4). It would also be

meaningless to say that topaz is twice as hard as fluorite because their respective hardness numbers on Mohs' scale are 8 and 4. The greater than symbol (>) in connection with ordinal data may be used to designate 'happier than' 'preferred to' and so on.

When we can set up inequalities and also form differences, we refer to the data as interval data. Assume we are given the following temperature readings in degree Fahrenheit: 58°, 63°, 70°, 95°, 110°, 126° and 135°. In this case, we can write 100°, > 70° or 95° < 135°, which simply means that 110°, is warmer than 70° and that 95° is cooler that 135°. We can also write, for example, 95° – 70° = 135° – 110° since equal temperature differences are equal in the sense that the same amount of heat is required to raise the temperature of an object from 70° to 95° or from 110° to 135°. On the other hand, it would not mean much if we said that 126°, is twice as hot as 63°, even though 126°/63° = 2. If we convert these temperatures to Centigrade, then 126°F becomes 52°C and 63°F becomes 17°C. 52°C is more than thrice of 17°C!

When in addition to setting up inequalities and forming differences, we can also form quotients (i.e., when we can perform all the customary operations of mathematics), we refer to such data as ratio data. In this sense, ratio data includes all the usual measurement (or determinations) of length, height, money, amounts, weight, volume, area, pressures etc.

This distinction between nominal, ordinal, interval and ratio data is important for the nature of a set of data as it may suggest the use of particular statistical techniques. A researcher has to be very careful about this aspect while measuring properties of objects or of abstract concepts.

4.2.1 Measurement Scales

Scales of measurement can be considered in terms of their mathematical properties. The most widely used classification of measurement scales are as follows:

- Nominal scale
- Ordinal scale
- Interval scale
- Ration scale

1. Nominal scale

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence. The numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of

football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one with the numbers assigned to another. The counting of numbers in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales. Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. For example, if, according to Mendel's laws, you expected 10 of 20 offspring from a cross to be male and the actual observed number was 8 males, then you might want to know about the 'goodness to fit' between the observed and expected. Were the deviations (differences between observed and expected) the result of chance, or were they due to other factors. How much deviation can occur before you, the investigator, must conclude that something other than chance is at work, causing the observed to differ from the expected. The chi-square test is always testing what scientists call the **null hypothesis,** which states that there is no significant difference between the expected and observed result.

Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes the differences between things by assigning them to categories. Nominal data is thus, counted data. The scale wastes any information that we may have about varying degrees of attitude, skills, understanding, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other ex-post-facto research when data is being classified by major sub-groups of population.

2. Ordinal scale

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statements about scores based on ordinal scales. For instance, if Ram's position in his class is 10th and Mohan's position is 40th, it cannot be said that Ram's position is 4 times as good as that of Mohan. The statement would make no sense at all. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made.

Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the

NOTES

median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various ranks order methods. Measures of statistical significance are restricted to the non-parametric methods.

3. Interval scales

In case of interval scale, the intervals are adjusted in terms of some rule that has been established as bases for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero but it is not possible to determine for them, what may be called an absolute zero or the unique origin. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic. The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to 40° involves the same increase in temperature as in increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water. The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point.

Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of intervals as much more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used test for statistical significance are the 't' test and 'F' test.

4. Ratio scale

Ratio scale has absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive zero of length and similarly we can conceive of an absolute zero of time. For example, the zero point on a centimetre scale indicates the complete absence of length or height, but an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in scientist's mind. The number of minor traffic-rules violation and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeroes and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like 'Jyoti's typing speed was twice as good as that of Reetu.' The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale.

Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc., are examples. Generally all statistical techniques are usable with ratio scales and all manipulations that one can

carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variations may also be calculated.

Thus, proceedings from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variables permits, the researcher should use the scale that provides the most precise description. A researcher in physical sciences has the advantage to describe variables in interval scale form, a less precise type of measurement.

Sources of error in measurement

Measurement should be precise and unambiguous in an ideal research study. This objective, however, is often not met with an entirety. As such the researcher must be aware about the sources of error in measurement. The following are the possible sources of error in measurement.

- **Respondent:** At times the researcher may be reluctant to express strong negative feelings or it is just possible that he may have very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of 'guesses'. Transient factors like fatigue, boredom, anxiety, etc., may limit the ability of the respondent to respond accurately and fully.
- **Situation:** Situational factors may also come in the way of correct measurement. Any condition which places a strain on interview can have serious effects on the interviewer-respondent rapport. For instance, if the respondent feels that anonymity is not assured, he may reluctant to express certain feelings.
- **Measurer:** The interviewer can distort responses by rewording or reordering questions. His behaviour, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the finding. Errors may creep in incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data-analysis stage.
- **Instrument:** Error may arise because of the defective measuring instrument. The use of complex words, beyond the comprehension of the respondent, ambiguous meanings, poor printing, inadequate space for replies, response choice omissions, etc. are a few things that make the measuring instruments defective and may result in measurement errors. Another type of instrument deficiency is the poor sampling of the universe of items of concern.

  Researcher must know that correct measurement depends on successfully meeting all of the problems listed above. He must, to the extent possible, try to eliminate, neutralize or otherwise deal with all the possible sources of error so that the final results may not be contaminated.

4.2.2 Test of Sound Measurement

Sound measurement must meet the tests of validity, reliability and practicability. In fact, these are the three major considerations one should use in evaluating a measurement tool. 'Validity refers to the extent to which a test measures what we actually wish to measure. Reliability has to do with the accuracy and precision of a measurement procedure …. Practicality is concerned with a wide range of factors of economy, convenience, and interpretability.' We briefly take up the relevant details concerning these tests of sound measurement.

1. Test of validity

Validity is the most critical criterion and indicates the degree to which an instrument measures what it is supposed to measure. Validity can also be thought as utility. In other words, validity is the extent to which differences found with a measuring instrument reflect true differences among those being tested. But the question arises: how can one determine validity without direct confirming knowledge? The answer may be that we seek other relevant evidences that confirm the answers we have found with our measuring tool. What is relevant, evidence often depends upon the nature of the research problem and the judgment of the researcher. But one can certainly consider three types of validity in this connection:

(i) **Content validity:** is the extent to which a measuring instrument provides adequate coverage of topic under study. If the instruments contains a representative sample of the universe, the content validity is good its determination is primarily judgmental and intuitive. It can also be determined by using a panel of persons who shall judge how well the measuring instrument meets the standards, but there is no numerical way to express it.

(ii) **Criterion-related validity:** Relates to our ability to predict some outcome or estimate the existence of some correct condition. This form of validity reflects the success of measures used for some empirical estimating purpose. The concerned criterion must possess the following qualities:

- Relevance: (a criterion is relevant if it is defined in terms we judge to be the proper measure)
- Freedom from bias: (freedom from bias is attained when the criterion gives each subject an equal opportunity to score well.)
- Reliability: (a reliable criterion is stable or reproducible)
- Availability: (the information specified by the criterion must be available.)

In fact, a criterion-related validity is a broad term that actually refers to: (i) predictive validity and, (ii) concurrent validity. The former refers to the usefulness of a test in predicting the future performance whereas the later refers to the usefulness of the test in closely relating to other measures of known validity. Criterion–related validity is expressed as coefficient of correlation between test scores and some measure of future performance or between test scores on another measure of known validity.

(iii) **Construct validity:** Is the most complex and abstract. A measure is said to possess construct validity to the degree that it confirms to predicted correlations with other theoretical propositions. Construct validity is the degree to which scores on a test can be accounted for by the explanatory construct of the sound theory. For determining construct validity, we associate a set of other propositions with the results received from using our measurement instrument. If measurements on our devised scale correlate in a predicted way with these other propositions we can conclude that there is some construct validity.

If the above stated criteria and tests are met with, we may state that our measuring instrument is valid and will result in correct measurement. Otherwise we shall have to look for more information and/ or resort to exercise of judgment.

2. Test of reliability

The test of reliability is another important test of sound measurement. A measuring instrument is reliable if it provides consistent results. Reliable measuring instrument does contribute to validity, but a reliable imbruement need not be a valid instrument. For instance, a scale that consistently over- weighs objects by 5 kg., is a reliable scale, but it does not give a valid measure of weight. But the other way is not true, i.e., a valid instrument is always reliable. Accordingly reliability is not as valuable as validity, but is easier to assess reliability in comparison to validity. If the quality of reliability is satisfied by an instrument, then while using it we can be confident that the transient and situational factors are not interfering.

Two aspects of reliability, viz., stability and equivalence deserve special mention. The stability aspect is concerned with securing consistent results with repeated measurements of the same person and with the same instrument. We usually determine the degree of stability by comparing the results of repeated measurements. The equivalence aspect considers how much error may get introduced by different investigators or different samples of the items being studied. A good way to test for the equivalence of measurements by two investigators is to compare their observations of the same events. Reliability can be improved in the following two ways:

- By standardizing the conditions under which the measurements takes place, i.e., we must ensure that external sources of variations such as boredom, fatigue, etc., are minimized to the extent possible. That will improve stability aspect.
- By carefully designed directions for measurement with no variation from group to group, by using trained and motivated persons to conduct the research and also by broadening the sample of items used. This will improve equivalence aspect.

3. Test of practicality

The practicality characteristic of a measuring instrument can be judged in terms of economy, convenience and interpretability. From the operational point of view, the measuring instrument ought to be practical, i.e., it should be economical, convenient and interpretable. Economy considerations suggests that some trade–off is needed between the ideal research project and that which the budget cab afford. The length of measuring instrument is an important area where economic pressures are quickly felt. Although more items give greater reliability as stated earlier, but in the interest of limiting the interview or observation time, we have to take only few items for our study purpose. Similarly, data–collection methods to be used are also dependent at aims upon economic factors. Convenience test suggests that the measuring instrument should be easy to administer. For this purpose one should give due attention to the proper layout of the measuring instrument, for instance, a questionnaire, with clear instructions (illustrated by examples), is certainly more effective and easier to complete than one which lacks these features. Interpretability considerations are especially important when persons other than the designers of the test are to interpret the results. The measuring instrument, in order to be interpretable, must be supplemented by: (a) detailed instructions for administering the test; (b) scoring keys; (c) evidence about the reliability and (d) guides for using the test and for interpreting results.

Technique of developing measurement tools

The technique of developing measurement tools involves a four-stage process, consisting of the following:

(a) Concept development

(b) Specification of concept dimensions

(c) Selection of indicators

(d) Formation of index

The first and foremost step is that of concept development which means that the researcher should arrive at an understanding of the major concepts pertaining to his study. This step of concept development is more apparent in theoretical studies than in the more pragmatic research, where the fundamental concepts are often already established.

The second steps requires the researcher to specify the dimensions of the concepts that he developed in the first stage. This task may either be accomplished by deduction, i.e., by adopting a more or less intuitive approach or by the empirical correlation of the individual dimensions with the total concept and/or the other concepts. For instance, one may think of several dimensions such as product reputation, customer treatment, corporate leadership, concern for individuals, and sense of social responsibility and so forth when one is thinking about the image of a certain company.

Once the dimensions of a concept have been specified, the researcher must develop indicators for measuring each concept element. Indicators are specific questions, scales, or other devices by which respondent's knowledge, opinion, expectation, etc., are measured. As there is seldom a perfect measure of a concept, the researcher should consider several alternatives for the purpose. The use of more than one indicator gives stability to the scores and it also improves their validity.

The last step is that of combining the various indicators into an index, i.e., formation of an index. When we have several dimensions of a concept or different measurements of a dimension, we may need to combine them into a single index. One simple way for getting an overall index is to provide scale values to the responses and then sum up the corresponding scores. Such an overall index would provide a better measurement tool than a single indicator because of the fact that an 'individual indicator has only a probability relation to what we really want to know'. This way we must obtain an overall index for the various concepts concerning the research study.

4.2.3 Scaling

In research we quite often face measurement problems (since we want a valid measurement but we may not obtain it), especially when the concepts to be measured are complex and abstract and we do not possess the standardized measurement tools. Alternatively, we can say that while measuring attitudes and opinions, we face the problem of their valid measurement. Similar problems may be faced by a researcher, of course in a lesser degree, while measuring physical or institutional concepts. As such we should study some procedures which may enable us to measure abstract concepts more accurately. This brings us to the study scaling techniques.

Meaning of scaling

Scaling describes the procedures of assigning numbers to various degrees, of opinion, attitude and other concepts. This can be done in two ways viz.:

- Making a judgment about some characteristics of an individual and then placing him directly on a scale that has been defined in terms of that characteristics
- Constructing questionnaires in such a way that the score individual's responses assigns him a place on a scale

It may be stated here that a scale is a continuum, consisting of the highest point (in terms of some characteristics e.g. preference, fareableness etc.) and the lowest point along with several intermediate points between these two extreme points. These scale–point positions are so related to each other that when the first point happens to be the highest point, the second point indicates the higher degree in term as o f o a given characteristics as compared to the third point and the third

point indicates a higher degree as compared to the fourth and so on. Numbers for measuring the distinctions of degree in the attitudes/opinions are, thus, assign to individuals corresponding to their scale/positions. Or this is better understood when we talk about technique(s).

Scale classification basis

The number assigning procedures or the scaling procedures may be broadly classified on one or more of the following bases:

- **Subject orientation:** Under it a scale may be designed to measure characteristics of the respondent who completes it or to judge the stimulus object which is presented to the respondent. In respect of the former, we presume that the stimuli presented are sufficiently, genius so that the between-stimuli variation is small as compare to the variation among respondents. In the latter approach we ask the respondent to judge some specific object in terms of one or more dimensions and we presume that the between–respondent variation will be small as compare to the variation among the different stimuli presented to respondents for judging
- **Response form:** Under this we may classify the scales as categorical and comparative. Categorical scales are also known as rating scales. These scales are used when a respondent scares some object without direct reference to other objects. Under comparative scales which are also known as Ranking scales the respondent is asked to compare two or more objects. In this sense the respondent state that one object is superior to the other or that three models of pen rank in order 1, 2, and 3. The essence of ranking is infect, a relative comparison of a certain property of two or more objects.
- **Degree of subjectivity:** With these bases the scale data may be based on whether we measure subjective personal preferences or simply make non-preference judgments. In the former case, the respondent is asked to choose which person he favours or which solution he would like to see employed where as in the latter case he is simply asked to judge which person is more affective in some aspect or which solution will take fewer resources without reflecting any personal preference.
- **Scale properties:** Considering scale properties one may classify the scales as nominal, ordinal, interval and ratio scales. Nominal scales merely classify without indicating order, distance or unique origin. Ordinal scales indicate magnitude relationship of 'more than' or 'less than', but indicate no distance or unique origin. Interval scales have both order and distance values, but no unique origin. Ratio scales poses all these features.
- **Number of dimensions:** In respect of these bases, scale can be classified as 'unidimensional' and 'multidimensional' scales. Under the former we measure only one attribute of the respondent or the object whereas multidimensional scaling recognizes that an object might be described better

by using the concept of an attribute space of 'n' dimensions, rather than a single-dimension continuum.

- **Scale construction techniques:** Following are the five main techniques by which scales can be developed
  (a) *Arbitrary approach:* It is an approach where scale is developed on ad-hoc bases. This is the most widely used approach. It is presumed that such scales measure the concepts for which they have been designed, although there is little evidence to support such an assumption.
  (b) *Consensus approach:* here a panel of judges evaluate the items chosen for inclusion in the instrument in terms of whether they are relevant to the topic area and unambiguous in implication
  (c) *Item analysis approach:* Under it a number of individual items are developed into a test which is given to a group of respondents. After administering the test, the total scores are calculated for everyone. Individual items are than analysed to determine which items discriminate between persons or objects with high total scores and those with low scores.
  (d) *Cumulative scales* are chosen on the base of their confirming to some ranking of items with ascending and descending discriminating power. For instance, in such a scale the endorsement of an item representing an extreme position should also result in the endorsement of all items indicating less extreme positions.
  (e) *Factor scales* may be constructed on the bases of inter correlations of items which indicate that a common factor accounts for the relationship between items. This relationship is typically measured through factor analysis method

4.2.4 Important Scaling Techniques

Let us now discuss some of the important scaling techniques often used in the context of research especially in context of social or business research.

Rating scales: the rating scale involves qualitative description of a limited number of aspects of a thing or of trades of a person. When we use rating scales (or categorical scales), we judge an object in absolute terms against some specified criteria i.e., we judge properties of objects without reference to other similar objects. These ratings may be 'in such forms as 'Like- dislike', 'above average, average, below average', or other classification with more categories such as 'like very much, like somewhat-neutral-dislike somewhat, dislike very much'; 'Excellent, good, average, below average, poor', 'always, often, occasionally, rarely, never'; and so on. There is no specific rule whether to use a two-points scale, three-points scale or scale with still more points. In practice, three to seven points scales are generally use for the simple reason that more points on a scale provide an opportunity for greater sensitivity of measurement.

Rating scales may be either a graphic rating scale or an itemized rating scale:

(i) The graphic rating scale is quite simple and is commonly used in practice. Under it the various points are usually put along the line to form a continuum and the rater indicates his rating by simply making a mark (such as) at the appropriate point on a line that runs from one extreme to the other scale-points with brief descriptions may be indicated along the line, their function being to assist the rater in performing his job. The following is an example of five-points graphic rating scale when we wish to ascertain people's liking or disliking any product:

This type of scale has several limitations. The respondents may check at almost any position along the line which fact may increase the difficulty of analysis. The meanings of the terms like 'very much' and 'somewhat' may depend upon respondent's frame of reference so much so that the statement might be challenged in terms of its equivalency. Several other rating scale variants (e.g., boxes replacing line) may also be used.

(ii) The itemized rating scale (also known as numerical scale) presents a series of statements from which a respondent selects one as best reflecting his evaluation. These statements are ordered progressively in terms of more or less of some property. An example of itemized scale can be given to illustrate it.

Suppose we wish to inquire as to how well does a worker get along with his fellow workers? In such a situation we may ask the respondent to select one, to express his opinion, from the following:

- He is almost always involved in some friction with a fellow worker.
- He is often at odds with one or more of his fellow workers.
- He sometimes gets involved in friction.
- He infrequently becomes involved in friction with others.
- He almost never gets involved in friction with fellow workers.

The chief merit of this type of scale is that it provides more information and meaning to the rater, and thereby increases reliability. This form is relatively difficult to develop and the statements may not say exactly what the respondents would like to express.

Rating scales have certain good points. The results obtained from their use compare favourably with alternative methods. They require less time, are interesting to use and have a wide range of applications. Besides, they may also be used with a large number of properties or variables. But their value for measurement purposes depends upon the assumption that the respondents can and do make good judgments. If the respondents are not very careful while rating, errors may occur. Three types of errors are common viz., *the error of leniency*, *the error of central tendency* and the *error of halo effect*. The error of leniency occurs when certain respondents are

either easy raters or hard raters. When raters are reluctant to give extreme judgments, the result is the error of central tendency. The error of halo effect or the systematic bias occurs when the rater carries over a generalized impression of the subject from one rating to another. This sort of error takes place when we conclude for example, that a particular report is good because we like its form or that someone is intelligent because he agrees with us or has a pleasing personality. In other words, halo effect is likely to appear when the rater is asked to rate many factors, on a number of which he has no evidence for judgment.

Ranking scales: Under ranking scales (or comparative scales) we make relative judgments against other similar objects. The respondents under this method directly compare two or more objects and make choices among them. There are two generally used approaches of ranking scales viz.

(a) **Method of paired comparisons:** under it the respondent can express his attitude by making a choice between two objects, say between a new flavour of soft drink and an established brand of drink. But when there are more than two stimuli to judge, the number of judgments required in a paired comparison is given by the formula:

$$N = n(n-1) / 2$$

Where N = number of judgments.

n = number of stimuli or objects to be judged.

For instance if there are 10 suggestions for bargaining proposals available to a workers union, there are 45 paired comparisons that can be made with them. When N happens to be big figure there is a risk of respondents giving I'll considered answers or they may even refused to answer. We can reduce the number of comparisons per respondents either by presenting to each one of them only a sample of stimuli or by choosing a few objects which cover the range of attractiveness at about equal intervals and then comparing all other stimuli to these few standard objects. Thus, paired–comparison data may be treated in several ways. If there is substantial consistency we will find that if X is preferred to Y, and Y to Z, then X will consistently be preferred to Z. If this is true, we may take the total number of preferences among the comparisons as the score for that stimulus.

It should be remembered that paired comparison provides ordinal data, but the same may be converted into an interval scale by the method of the Law of Comparative Judgment developed by L. L. Thurstone. This technique involves the conversion of frequencies of preferences into a table of proportions which are then transformed into Z matrix by referring to the table of area under the normal curve. J. P. Guilford in his book *Psychometric Methods* has given a procedure which is relatively easier. The method is known as the Composite Standard Method and can be illustrated as under:

Suppose there are 4 proposals which some union bargaining committee is considering. The committee wants to know how the union membership ranks these proposals. For this purpose a sample of 100 members might express the views as shown in the following table:

Table 4.1 *Response Patterns of 100 Members' Paired Comparisons of 4 Suggestions for Union Bargaining Proposal Priorities Suggestions*

	A	B	C	D
A	—	65	32	20
B	40	—	38	42
C	45	50	—	70
D	80	20	98	—
Total	165	135	168	132

Rank order	2	3	1	4
Mp	o.5375	0.4625	0.5450	0.4550
Zj	0.09	(-).09	0.11	(-).11
Rj	0.20	0.02	0.22	0.00

Comparing the total number of preferences for each of the four proposals, we find that C is the most popular, followed by A, B and D respectively in popularity. The rank order shown in the above table explains all this.

By following the composite standard method, we can develop an interval scale from the paired-comparison ordinal data given in the above table for which purpose we have to adopt the following steps in order:

(i) Using the data in the above table, we work out the column mean with the help of the formula given below:

$$M p = C + 0.5 (N)/ Nn = 165 + 0.5 (100)/4(100) = 0.5375$$

Where

Mp = the mean proportion of the columns

C = the total number of choices for a given suggestions

n = number of stimuli (proposals in the given problem)

N = number of items in the samples.

The column means have been shown in the Mp row in the above table.

(ii) The Z values for the Mp are secured from the table giving the area under the normal curve when the Mp value is less than 0.5, the Z value is negative and for all Mp values higher than 0.5, the Z values are positive. These Z values are shown in Zj row in the above table.

(iii) As the Zj values represent an interval scale, zero is an arbitrary value. Hence we can eliminate negative scale values by giving the value of zero to the lowest scale value (this being (–).11 in our example which we shall take equal to zero) and then adding the absolute value of this lowest scale value

to all other scale items. This scale has been shown in Rj row in the above table.

Method of rank order: under this method of comparative scaling, the respondents are asked to rank their choices this method is easier and faster than the method of paired comparisons stated above. For example, with 10 items it takes 45 pair comparisons to complete the task, whereas the method of rank order simply requires ranking of 10 items only. The problem of transitivity (such as A prefers to B, B to C, but C prefers to A) is also not there in case we adopt method of rank order. Moreover, a complete ranking at times in not needed in which case the respondents may be asked to rank only their first, say, four choices while the number of overall items involved may be more than four, say, it may be 15 or 20 or more. To secure a simple ranking of all items involved we simply total rank values received by each item. There are methods through which we can as well develop an interval scale of these data. But then there are limitations of this method. The first one is that data obtained through this method are ordinal data and hence rank ordering is an ordinal scale with all its limitations then there may be the problem of respondents becoming careless in assigning ranks particularly when there are many (usually more than 10) items.

Scale constructing techniques

In social science studies, while measuring attitudes of the people we generally follow the technique of preparing the opinionnaire (or attitude scale) in such a way that the score of the individual responses assigns him a place on a scale. Under this approach, the respondent expresses his agreement or disagreement with a number of statements relevant to the issue. While developing such statements, the researcher must note the following two points:

- That the statements must elicit responses which are psychologically related to the attitude being measured
- That the statements need be such that they discriminate not merely between extremes of attitude but also among individuals who differ slightly

Researcher must as well be aware that inferring attitude from what has been recorded in opinionnaires has several limitations. People may conceal their attitudes and express socially acceptable opinions. They may not really know how they feel about a social issue. People may be unaware of their attitude about an abstract situation; until confronted with a real situation, they may be unable to predict their reaction. Even behaviour itself is at times not a true indication of attitude. For instance, when politicians kiss babies, their behaviour may not be a true expression of affection towards infants. Thus, there is no sure method of measuring attitudes; we only try to measure the expressed opinion and then draw inferences from it about people's real feelings or attitudes.

With all these limitations in mind, psychologists and sociologists have developed several scale construction techniques for the purpose. The researcher

NOTES

should know these techniques so as to develop an appropriate scale for his own study. Some of the important approaches, along with the corresponding scales develop under each approach to measure attitude are as follows:

Arbitrary scales

Arbitrary scales are developed on ad hoc bases and are designed largely through the researcher own subjective selection of items. The researcher first collects few statements or items which he believes are unambiguous and appropriate to a given topic. Some of these are selected for inclusion in the measuring instrument and then people are asked to check in a list the statements with which they agree.

The chief merit of such scale is that they can be developed very easily, quickly and with relatively less expense. They can also be designed to be highly specific and adequate. Because of these benefits, such scales are widely used in practice.

At the same time there are some limitations of these scales. The most important one is that we don't have objective evidence that such scales measure the concepts for which they have been develop. We have simply to rely on researchers in sight and competence.

Differential scales (or Thurstone-types scales)

The name of L. L. Thurstone is associated with differential scales which have been developed using consensus scale approach. Under such an approach the selection of items is made by a panel of judges who evaluate the items in terms of whether they are relevant to the topic area and unambiguous in implication. The detailed procedure is as under:

- The researcher gathers a large number of statements, usually 20 or more, that express various points of view towards a group, institution, idea, or practice (i.e., statements belonging to the topic area).
- These statements are then submitted to a panel of judges, each of whom arranges them in 11 groups or piles ranging from one extreme to another in position. Each of the judges is requested to place generally in the first pile statements which he thinks are most unfavourable to the issue, in the second pile to place those statements which he thinks are next most unfavourable and he goes on doing so in this manner till in the 11th pile he puts the statements which he considers to be the most favourable.
- This sorting by each judge yields a composite position for each the items. In case of marked disagreement between the judges in assigning a position to an item, that item is discarded.
- For items that are retained, each is given its median scale value between 1 and 11 as established by the panel. In other words, the scale value of any one statement is computed as the 'median' position to which it is assigned by the group of judges.

- A final selection of statements is then made. For this purpose a sample of statements, whose median scores are spread evenly from one extreme to other is taken. The statements so selected, constitute the final scale to be administered to respondents. The position of each statement on the scale is the same as determined by the judges.

After developing the scale as stated above, the respondents are asked during the administration of the scale to check the statements with which they agree. The median value of the statements that they check is worked out and this establishes their score or qualifies their opinion. It may be noted that in the actual instrument the statements are arranged in random order of scale value. If the values are valid if the opinionnaire deals with only one attitude dimensions, the typical respondent will choose one or several contiguous items (in terms of scale values) to reflect his views. However, at times divergence may occur when a statement appears to tap a different attitude dimensions

The Thurstone method has been widely used for developing differential scales which are utilized to measure attitudes towards varied issues like war, religion, etc. Such scales are considered the most appropriate and reliable when used for measuring a single attitude. But an impotent deterrent to their use is the cost and effort required to develop them. Another weakness of such scales is that the values assigned to various statements by the judges may reflect their own attitudes. The method is not completely objective; it involves ultimately subjective decision process. Critics of this method opine that some other scale designs give more information about the respondent's attitude in comparison to differential scale.

Summated scale (or Likert type scale)

Summated scale (or Likert type scale) are developed by utilizing the item analysis approach wherein a particular item is evaluated on the basis of how well it discriminates between those persons whose total score is high and those whose score is low. Those items or statements that best meet this sort of discrimination test are included in the final instrument.

Thus, summated scales consist of a number of statements which express either a favourable or unfavourable attitude towards the given object to which the respondent is asked to react. The respondent indicates his agreement or disagreement with each statement in the instrument. Each response is given a numerical score, indicating its favourableness or unfavourableness, and the scores are totalled to measure the respondent's attitude. In other words, the overall score represent the respondent's position on the continuum of favourable-unfavourableness towards an issue.

Most frequently used submitted scales in the study social attitudes follow the pattern devised by Likert. For this reason they are often referred to as Likert-type scales. In a Likert scale, the respondent is asked to respond to each of the statements in terms of several degrees, usually five degrees (but at times 3 or 7 may also be used) of agreement or disagreement. For example, when asked to

express opinion whether one considers his job quite pleasant, the respondent may respond in any of the following ways:

- Strongly agree,
- Agree
- Undecided
- Disagree
- Strongly disagree
- We find that these five points constitute the scale. At one extreme of the scale there is strong agreement with the given statement at the other, strong disagreement, and between them lie intermediate points. We may illustrate this as under:

Strongly Agree (1)	Agree (2)	Undecided (3)	Disagree (4)	Strongly Disagree (5)

Each point on the scale carries a score. Response indicating the least favourable degree of jobs satisfaction is given the least score (say 1) and the most favourable is given the highest score (say 5).These score–values are normally not printed on the instrument but are shown here just to indicate the scoring pattern. The Likert scaling technique thus, assigns a scale value to each of the five responses. The same thing is done in respect of each and every statement in the instrument. This way the instrument yields a total score for each respondent, which would then major the respondent's favourableness towards the given point of view. If the instrument consists of, say 30 statement, the following score values would be revealing.

30X5 = 150Most favourable response possible

30X3 = 90A neutral attitude

30X1 = 30Most Unfavourable attitude.

The scores for any individual would fall between 30 and 150.If the score happens to be above 90, it shows favourable opinion to the given point of view, a score below 90 would mean unfavourable opinion and a score of exactly 90 would be suggestive of a neutral attitude.

Procedure: The procedure for developing a Likert type scale is as follows:

- As first step, the researcher collects a large numbers of statement which are relevant to the attitude being studied and each of the statement expresses definite favourable Nasser unfavourableness to a particular point of view or the attitude and that the number favourable and unfavourable statements is approximately equal.
- After the statements have been gathered, a trial test should be administered to a number of subjects. In other words, a small group of people, from

those who are going to be studied finally, are asked to indicate there response to each statement by checking one of the categories of agreement or disagreement using a 5 point scale as stated above.

- The response to various statement are scored in such a way that a response indicative of the most favourable attitude is given the highest score of 5 and that with the most unfavourable attitude is given the lowest score, say, of 1.
- Then the total score of each respondent is obtained by adding his scores that he received for separate statement.
- The next step is to array these total score and find out those statements which have a high discriminatory power for this purpose, the researcher may select some part of the highest and the lowest total scores, say the top 25 per cent and the bottom 25 per cent. These two extreme groups are interpreted to represent the most favourable and the least favourable attitude and are used as criterion groups by which to evaluate individual statements. This way we determine which statements consistently correlate with low favourability and with which with high favourability.
- Only those statements that correlate with the total test should be retained in the final instrument and all others must be discarded from it.

Advantages: The Likert type scale has several advantages. Mention may be made of the important ones:

- It is relatively easy to construct the Likert type scale in comparison to Thurstone type scale because Likert type scale can be perform without a panel of judges.
- Likert type scale is considered more reliable because under it respondents answer each statement included in the instrument. As such it also provides more information and data then does the Thurstone–type scale.
- Each statement, included in the Likert types scale, is given an imperial test for discriminating ability and as such, unlike Thurstone-type scale, the Likert type scale permits the use of statements that are not manifestly related (to have a direct relationship) to the attitude being studied.
- Likert type scale can easily be used in respondent-centered and stimulus-centered studies, through it we can study how responses differ between people and how responses differ between stimuli.
- Likert-type scale takes much less time construct, it is frequently used by the students of opinion research. Moreover, it has been reported in various studies that there is high degree co-relation between Likert type and Thurstone type scale.

Limitations

There are several limitations of the Likert type scale as well. One important limitation is that, with this scale, we can simply examine whether respondents are more or

less favourable to a topic, but we cannot tell how much more or less they are. There is no basis for belief that the five positions indicated on the scale are equally spaced. The interval between 'strongly agree' and 'agree', may not be equal to the interval between 'agree' and 'undecided'. This means that Likert scale does not rise to a stature more than that of an ordinal scale, whereas the designers of Thurstone scale claim the Thurstone scale to be an interval scale. One further disadvantage is that often the total score of an individual has little clear meaning since a given total score can be secured by a variety of answer patterns. It is unlikely that the respondent can validly react to a short statement on a printed form in the absence of real-life qualifying situations. Moreover, there 'remains a possibility that people may answer according to what they think they should feel rather than how they do feel'. This particular weakness of the Likert type scale is met by using a cumulative scale which we shall take up later in this unit.

In spite of all the limitations, the Likert-type summated scales are regarded as the most useful in a situation wherein it is possible to compare the respondent's score with a distribution of scores from some well-defined groups. They are equally useful when we are concerned with a programme of change or improvement in which case we can use the scales to measure attitudes before and after the program of change or improvement in order to assess whether our efforts have had the desired effects. We can as well correlate scores on the scale to other measures without any concern for the absolute value of what is favourable and what is unfavourable. All this accounts for the popularity of Likert-type scales in social studies relating to measuring of attitudes.

Cumulative scales

Cumulative scales or **Louis Guttmann's** scalogram analysis, like other scales, consists of series of statements to which a respondent expresses his agreement or disagreement. The special of this type of scale is that statements in it form a cumulative series. This, in other words, means that the statements are related to one another in such a way that an individual, who replies favourably to say item No. 3, also replies favourably to items No. 2 and 1, and one who replies favourably to item no. 4 also replies favourably to items no. 3, 2, and 1, and so on. This being so an individual whose attitude is at a certain point in a cumulative scale will answer favourably all the items on one side of this point, and answer unfavourably all the items on the other side of this point. The individual's score is worked out by counting the number of points concerning the number of statements he answers favourably. If one knows this total score, one can estimate as to how a respondent has answered individual statements constituting cumulative scales. The major scale of this type of cumulative scale is the Guttmann's scalogram. We attempt a brief description of the same below.

The technique developed by Louis Guttmann is known as scalogram analysis or at times simply 'scale analysis'. Scalogram analysis refers to the procedure for determining whether a set of items forms a unidimensional scale. A scale is said to

be unidimensional if the responses fall into a pattern in which endorsement of the item reflecting the extreme position results also in endorsing all items which are less extreme under this technique, the respondents are asked to indicate in respect of each item whether they agree or disagree with it, and if these items form a unidimensional scales, the response pattern will be as under:

Response Pattern in Scalogram Analysis

Item Number				Respondent Score
4	3	2	1	
X	X	X	X	4
-	X	X	X	3
-	-	X	X	2
-	-	-	X	1
-	-	-	-	0

X = Agree

– = Disagree

A score of 4 means that the respondent is in agreement with all the statements which is indicative of the most favourable attitude. But a score of 3 would mean that the respondent is not agreeable to item 4, but he agrees with all others. In the same way we can interpret other values of the respondents' scores. This pattern reveals that the universe of content is scalable.

Procedure: the procedure for developing a scalogram can be outlined as under:

- The universe of content must be defined first of all. In other words, we must lay down in clear terms the issue we want to deal within our study.
- The next step is to develop a number of items relating the issue and to eliminate by inspection the items that are ambiguous, irrelevant or those that happen to be too extreme items.
- The third step consist in pre-testing the items to determine whether the issue at hand is scalable (The pretest, as suggested by Guttmann, should include 12 or more items, while the final scale may have only 4 to 6 items. Similarly, the number of respondents in a pretest may be small, say 20 or 25 but final scale should involve relatively more respondents, say 100 or more).

  In a pretest the respondents are asked to record their opinions on all selected items using a Likert type 5 –point scale, ranging from 'strongly agree' to 'strongly disagree'. The strongest favourable response is scored as 5, whereas the strongest unfavourable response as 1. The total score can thus range, if there are 15 items in all, from 75 for most favourable to 15 for the least favourable.

  Respondent opinionnaires are then arrayed according to total score for analysis and evaluation. If the responses of an item from a cumulative

scale, it response category scores should decrease in an orderly fashion as indicated in the above table. Failure to show the decreasing pattern means that there is over lapping which shows that the item concerned is not a good cumulative scale item, i.e., the item has more than one meaning. Sometimes the over lapping in category responses can be reduced by combining categories. After analysis the pretest results, a few items, say 5 items, may be chosen.

- The next step is again to total the scores for the various opinionnaires, and to re array them to reflect any shift in order, resulting from reducing the items, say, from 15 in pretest to, say, 5 for the final scale. The final pretest results may be tabulated in the form of a table given in Table 4.2.

Table 4.2 *The Final Pretest Results in Scalogram Analysis*

Scale type	Item 5	12	3	10	7	Errors per case	Number of cases	Number of errors
5 (perfect)	x	x	x	x	x	0	7	0
4 (perfect) –		x	x	x	x	0	3	0
(Nonscale)-		x	-	x	x	1	1	1
(Nonscale)-		x	x	-	x	1	2	2
3 (perfect) -		-	x	x	x	0	5	0
2 (perfect) -		-	-	x	x	0	2	0
1 (perfect) -		-	-	-	x	0	1	0
(nonscale) -		-	x	-	-	2	1	2
(nonscale) -		-	x	-	-	2	1	2
0 (perfect)	-	-	-	-	-	0	2	0

The table shows that five items (numbering 5, 12, 3, 10 and 7) have been selected for the final scale. The number of respondent is 25 whose responses to various items have been tabulated along with the number of errors. Perfect scale types are those in which the respondent's answers fit the pattern that would be reproduce by using the person's total score as a guide. Non –scale types are those in which the category patterns differs from that expected from the respondent's total score, i.e., non-scale cases have deviations from unidimensionality or errors. Whether the items (or series of statements) selected for final scale may be regarded a perfect cumulative (or a unidimensional scale), we have to examine on the bases of the coefficient of the reproducitibility. Guttmann has set 0.9 as the level of the minimum reproducibility in order to say that the scale meets the test of unidimensionality. He has given the following formula for measuring the level of reproducibility:

Guttmann's Coefficient of Reproducibility = $1 - e/n(N)$

Where e = number of errors.

n = number of items

N = number of cases

For the above table figures,

Coefficient of Reproducibility = 1-7/5(25) = 0.94

This shows that items number 5, 12, 3, 10 and 7 in this order constitute the cumulative or unidimensional scale, and with this we can reproduce the responses to each items knowing only the total score of the respondent concerned.

Scalogram analysis, like any other scaling technique, has several advantages as well as limitations. One advantage is that it assures that only a single dimension of attitude is being measured. Researcher's subjective judgment is not allowed to creep in the development of scale since the scale is determined by the replies of respondents. Then, we require only a small number of items that make such a scale easy to administer. Scalogram analysis can appropriately be used for personal, telephone, or mail surveys. The main difficulty in using this scaling technique is that in practice perfect cumulative or unidimensional scales are very rarely found and we have only to use its approximation testing it through coefficient of reproducibility or examining it on the bases of some other criteria. This method is not a frequently used method for the simple reason that its development procedure is tedious and complex. Such scales hardly constitute reliable bases for assessing attitudes of persons toward complex objects for predicting the behavioural responses of individuals towards such objects. Conceptually, this analysis is a bit more difficult in comparison to other scaling methods.

4.2.5 Regression and Prediction

In statistics, regression is a mathematical method of modelling the relationships among three or more variables. It is used to predict the value of one variable given the values of the others. A regression analysis yields an equation that expresses the relationship. Statistical technique used to establish the relationship of a dependent variable, such as the sales of a company, and one or more independent variables, such as family formations, Gross Domestic Product, per capita income, and other economic indicators. By measuring exactly how large and significant each independent variable has historically been in its relation to the dependent variable, the future value of the dependent variable can be predicted. Essentially, regression analysis attempts to measure the degree of correlation between the dependent and independent variables, thereby establishing the latter's predictive value. For example, a manufacturer of baby food might want to determine the relationship between sales and housing starts as part of a sales forecast. Using a technique called a scatter graph, it might plot on the X and Y axes the historical sales for ten years and the historical annual housing starts for the same period. A line connecting the average dots, called the regression line, would reveal the degree of correlation between the two factors by showing the amount of unexplained variation-represented by the dots falling outside the line.

Thus, if the regression line connected all the dots, it would demonstrate a direct relationship between baby food sales and housing starts, meaning that one

could be predicted on the basis of the other. The proportion of dots scattered outside the regression line would indicate, on the other hand, the degree to which the relationship was less direct, a high enough degree of unexplained variation meaning there was no meaningful relationship and that housing starts have no predictive value in terms of baby food sales. This proportion of unexplained variations is termed the *coefficient of determination,* and its square root the Correlation Coefficient. The correlation coefficient is the ultimate yardstick of regression analysis: a correlation coefficient of 1 means the relationship is direct-baby food and housing starts move together; –1 means there is a negative relationship-the more housing starts there are, the less baby food is sold; a coefficient of zero means there is no relationship between the two factors.

Regression analysis is also used in securities' markets analysis and in the risk-return analyses basic to Portfolio Theory.

Prediction (forecasting) is a planning tool which helps management in its attempts to cope with the uncertainty of the future. It starts with certain assumptions based on the management's experience, knowledge, and judgment. These estimates are projected into the coming months or years using one or more techniques such as Box-Jenkins models, Delphi method, exponential smoothing, moving averages, regression analysis, and trend projection. Since any error in the assumptions will result in a similar or magnified error in forecasting, the technique of sensitivity analysis is used which assigns a range of values to the uncertain factors (variables). A forecast (which indicates what 'might' happen) should not be confused with a budget (which shows what 'ought' to happen). Prediction is a basic management function involving formulation of one or more detailed plans to achieve optimum balance of needs or demands with the available resources. The prediction process: (i) identifies the goals or objectives to be achieved, (ii) formulates strategies to achieve them, (iii) arranges or creates the means required, and (iv) implements, directs, and monitors all steps in their proper sequence.

Linear Regression Analysis

When it is estimated by using the methods of correlation that two variables (or data series) are correlated with other and it is also tested that expression of such relationship between the considered variables are theoretical permissible, then the next step in the process of analysis is of predicting and/or estimating the value of one variable from the known value of the other variable. This task, in econometrics literature is called as 'regression analyses'. Literary, the word 'regression' means a backward movement. In general sense, 'regression' means the estimation and/or prediction of the unknown value of one variable from the known value of the other variable. Hence, it is a study of the dependence of one variable on other variable(s).

Regression as a tool

Econometricians use regression analysis to make quantitative estimates of various theoretical relationships exist in the literature of social sciences and management, which previously have been completely theoretical in nature.

Linear regression model

The simplest single-equation linear regression model can be written as:

$$Y = \alpha + \beta X$$

Errors in the regression model

The relationship that expressed in equation above assumes that the value of Y only depends on the given value of *X*. In other words, it is only *X* variables that influence *Y* variable. But this is not the reality. This type of relationship may be a rare phenomenon, particularly when we are talking of business and economic relationship. Rather the truth is that there is almost always a variation that comes from other sources also. This additional variation comes in part from the left or omitted variable. Again, if these extra variables are added to the equation, still there are some variables which may cause variation to the dependent variable (*Y*), which is not explained in the equation. These variations may come from the sources such as omitted influence of the variables, some degree of measurement error, and selection of incorrect functional forms or may be such sources which are quite unpredictable.

Generalizing the regression model

But this equation can be generalized by increasing the numbers of observations and by allowing more and more independent variable. Now by including the observations, the equation can be written as:

$$Y_i = \alpha + \beta X_i + \mu_i, \text{ given that } (i = 1, 2, 3, \ldots, n)$$

where Yi = the ith observation of the dependent variable

Xi = the ith observation of the independent variables of the relationship

μ_i = the ith observation of the error or the stochastic error of the relationship

a and b = regression coefficients or the parameters to be estimated and

n = total numbers of observations

Methods of Constructing Regression Equations

Broadly speaking, regression can be studied in two ways:

1. Graphical method or
2. Constructing a regression line

Both the methods are analysed in detail below.

1. **Graphical method:** By drawing a scatter diagram, one can get the regression results. A scatter diagram (as discussed earlier) contains one point for one pair of values for two series of observations, X and Y. All the related pairs of observations are plotted on the graph paper and two regression lines can be drawn to predict the value of variable X and variable Y. That regression line which is drawn to predict the values of Y for a value of X is called as 'regression line of Y on X'. Similarly, the regression line that is used to predict a value of X for a value of Y is called as regression line of X on Y.
2. **Constructing a regression line:** Elhance and Aggarwal (1999) has suggested two ways for drawing a regression line, viz., (a) free hand curve method or (b) methods of least squares.
   - **Freehand curve method:** Here, the pairs of the values of both the variables X and Y (in the form of scatter diagram) are identified and then plotted on the graph paper. Then, two free hand regression lines are drawn by joining the identified values. One of these lines is drawn in such a way that the positive deviations of Y-series from its mean are cancelled by the negative deviations. With this, the sum of the deviations on one side of the line is equal to the sum of deviations on the other side of the series. This gives the regression line of Y on X. Similarly, the other regression line can also be drawn in such a way that the positive deviations of the series X from its mean value can be cancelled by its negative deviations. This line is called as regression line of X on Y. But in reality, it is very difficult to draw regression lines by using such a method. The greatest difficulty that arises while using this method is that of adjusting the negative and positive deviations to cancel each other in the scatter diagram.
   - **Method of least squares estimates:** In this method, the unknown parameters (á and â) are estimated in such a manner that the sum of squared deviations between the numerical data available and the model should be as minimum as possible. By simplifying the model in the above derived way, we can derive the two normal equations of *X on Y* as:

     $$\sum X = n\alpha + \beta \sum Y$$

     $$\sum XY = \alpha \sum Y + \beta \sum Y^2$$

Overview of Forecasting Methods

An overview of all the forecasting methods are provided here:

A. Extrapolative methods

1. Simple moving average

This method averages the last *n* observations of a time series. It is appropriate only for very short or very irregular data sets, where features like trend and seasonality cannot be meaningfully determined, and where the mean changes slowly.

2. Exponential smoothing, such as the Holt-Winters method

A more complex moving average method, involving parameters reflecting the level, trend and seasonality of historical data, usually giving more weight to recent data. Widely used in general business because of its simplicity, accuracy and ease of use. This method's robustness makes it useful even when historic data are few or volatile. It is a frequent winner in forecasting competitions.

3. Auto regressive moving average (ARMA)

An even more complex class of moving average models, capable of reflecting autocorrelations inherent in data. It can outperform exponential smoothing when the historical data period is long and data are non-volatile. But it does not perform as well when the data are statistically 'messy.'

B. Explanatory variable methods

1. Regression analysis

Fitting a curve to historical data using a formula based on independent variables (explanatory variables) and an error term. Although these methods are relatively simple, and are helpful both in analysing patterns of historical data and for correlation analysis, they are not generally recommended for forecasting. They have performed poorly in forecasting competitions.

2. Predictive modelling

An area of statistical analysis and data mining, that deals with extracting information from data and using it to predict future behaviour patterns or other results. A predictive model is made up of a number of predictors, variables that are likely to influence future behaviour.

3. Artificial neural networks

Patterned after the neural architecture of the brain, these methods allow for nonlinear connections between input and output variables, and for learning patterns in data.

4. Econometric modelling

Systems of simultaneous equations to represent economic relationships.

C. Simulation modelling methods

1. Cell-based modelling

Modelling of individual homogeneous units (cells) over time, such as age/sex cells in pension forecasting. These models are usually deterministic, but may be stochastic. They are useful to model large systems.

2. System dynamics simulation

Simulation of a system as a whole over time, incorporating feedback loops as well as stocks and flows. Such methods are useful for complex systems.

3. Multi-agent simulation

A computer representation that employs multiple interacting agents and behavioural rules to mimic the behaviour of a real system. This method is especially useful for modelling complex adaptive systems.

D. Judgmental methods

These methods rely on expertise and intuition, rather than on statistical analysis of historical data. Such methods are particularly useful when historical data is scarce. Many of the methods of 'futurism'—such as the Delphi method, visioning and scenario building—fall under this category.

E. Composite methods

1. Bayesian forecasting

This family of methods combines statistical methodology with structured integration of human judgment: new evidence is used to update a statistical forecast, based on application of Bayes' theorem. These methods are good for highly seasonal data with short history.

2. Other

Combinations of forecasting methods usually perform better in forecasting competitions. The use of composite methods will increase as decision makers are increasingly called on to combine their intuitions with data-based decision making from forecasting models.

4.2.6 Hypothesis Testing

A number of steps are involved in testing a hypothesis:

- Formulate a hypothesis: Let us discuss about drug. The drug is tested on a few patients and based on the response from patients, a decision has to be taken whether the drug should be introduced or not. We make certain assumptions about the parameter to be tested – these assumptions are known as hypothesis.

We start with a null hypothesis Ho: =100. This is a claim or hypothesis about the values or population parameters.

This tested against alternate hypothesis Ha: =100. The null hypothesis is tested with available evidence and a decision is made whether to accept this hypothesis or reject it. If the null hypothesis is rejected, we accept the alternate hypothesis.

- Setting up a suitable significance level: there are two types of errors that can be committed in making decisions accepting or rejecting the null hypothesis:
  (1) Type I error: An error made in rejecting the null hypothesis, when infect it is true.
  (2) Type II error: An error made in accepting the null hypothesis, when infect it is not true.

  The level of significance signifies the probability of committing Type I error and is generally taken an equal to 5 per cent. This means that even after testing the hypothesis, when a decision is made, we may still be committing 5 per cent error in rejecting the null hypothesis when it is actually true. Sometimes the value of is taken as 0.01 but it is the discretion of investigator, depending upon the sensitivity of the study.
- **Choose a test criteria:** This means selection of a suitable test statistical that can be used along with the available information carrying out the test. The different test statistical that are normally used are:
  - Normal Distribution: Z-Test, this is most often used, when the sample size is more than 30
  - T-Test is used for small samples only
  - F-Test
  - Chi-Square test
- **Compute the test characteristics:** This involve the actual collection and consideration, we have to find the sample mean (X) and then compute the calculated 'Z'. This calculated value (absolute) is compared with tabulated value obtained from normal distribution table against the decided criteria (Value of X and one tail or two tail).
- **Make a decision:** If the calculated value of the test characteristics is greater than the tabulated value, the null hypothesis is rejected and the alternate hypothesis is accepted. Talking in terms of critical region, the value of calculated characteristics falls outside the acceptance region.

CHECK YOUR PROGRESS

1. List two of the most widely used measurement scales.
2. List two possible sources of error in measurement.
3. Which three important criteria must the sound measurement meet?
4. List two bases according to which the scaling procedures may be classified.

4.3 BASIC CONCEPTS OF COMPUTERS

The word 'compute' means 'to calculate'. We all are familiar with calculations in our day-to-day life. We apply mathematical operations like addition, subtraction, multiplication, etc., and many other formulae for calculations. Simpler calculations take less time while complex calculations take much longer. Accuracy is another important factor to be kept in mind while calculating something. So man explored with the idea to develop a machine which can perform this type of arithmetic calculation faster and with more accuracy. Hence, the device or machine called **computer** was developed.

The computer we see today is quite different from the one made in the beginning. The number of applications of a computer has increased, the speed and accuracy of calculation has increased. We must appreciate the impact of computers in our day-to-day life. Reservation of tickets in airlines and railways, payment of telephone and electricity bills, deposits and withdrawals of money from banks, business data processing, medical diagnosis, weather forecasting, etc., are some of the areas where computer has become extremely useful.

However, there is one limitation of the computer. Human beings do the calculations on their own, but the computer is just a machine which has to be given proper instructions to carry out its calculation. This is why we should know how a computer works.

A computer is an electronic device which not only does arithmetic calculations quickly, but many other things. It serves different purposes for different people. A computer cannot only act as a calculator, but can also solve various problems and manipulate data. Figure 4.1 illustrates a personal computer.

Fig. 4.1 *Personal Computer*

The following are the characteristics of a computer:

- It operates at a very fast pace.
- It gives very accurate results.
- Computers are very diligent and can work for hours without getting tired.
- They have the capacity to store large amounts of data for a very long time.
- They have no IQ or feelings.

4.3.1 History of Computer

This section discusses the history of computers.

Calculating machines

It took many centuries for early man to build mechanical devices for counting large numbers. The first calculating device called the ABACUS was developed by the Egyptian and Chinese people.

The word ABACUS means a calculating board. It consisted of sticks in horizontal positions on which were inserted sets of pebbles. A modern form of ABACUS is given in Figure. 4.2. It has a number of horizontal bars, each having ten beads. Horizontal bars represent units, tens, hundreds, etc.

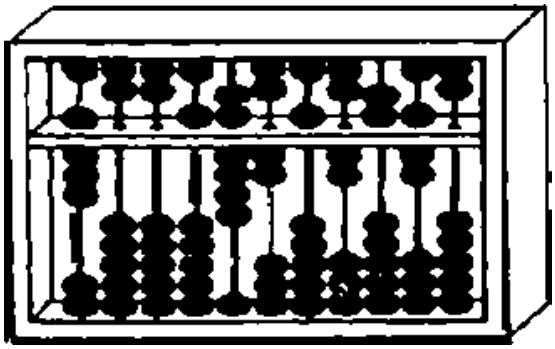


Fig. 4.2 *An Abacus*

NOTES

Napier's bones

English mathematician John Napier built a mechanical device for the purpose of multiplication in 1617. The device was known as Napier's bones.

Slide rule

English mathematician Edmund Gunter developed the slide rule. This machine could perform operations like addition, subtraction, multiplication, and division. It was widely used in Europe in the 16th century.

Pascal's adding and subtractory machine

You might have heard of Blaise Pascal. He developed a machine at the age of 19 that could add and subtract. The machine consisted of wheels, gears and cylinders.

Leibniz's multiplication and dividing machine

Around 1673, the German philosopher and mathematician Gottfried Leibniz built a mechanical device that could both multiply and divide.

Babbage's analytical engine

It was in the year 1823 that a famous Englishman Charles Babbage built a mechanical machine to do complex mathematical calculations. It was called *difference engine*. Later he developed a general-purpose calculating machine called *analytical engine*. You should know that Charles Babbage is called the *father of computer*.

Mechanical and electrical calculator

In the beginning of the 19th century the mechanical calculator was developed to perform all sorts of mathematical calculations. Up to the 1960s it was used widely. Later the rotating part of mechanical calculator was replaced by an electric motor and so it was called the electrical calculator.

Modern electronic calculator

The electronic calculator used in the 1960s was run with electron tubes, which was quite bulky. Later it was replaced with *transistors* and as a result the size of calculators became too small.

The modern electronic calculator can compute all kinds of mathematical computations and mathematical functions. It can also be used to store some data permanently. Some calculators have in-built programmes to perform some complicated calculations.

Figure 4.3 illustrates vacuum tube, a transistor and an integrated circuit chip.

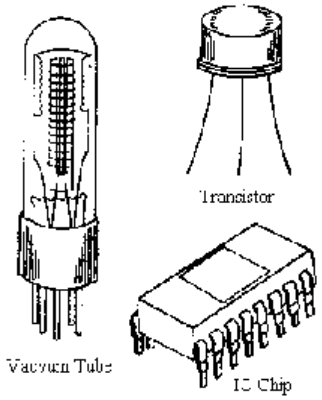


Fig. 4.3 *Vacuum Tube, Transistor, IC*

4.3.2 Computer Generations

You know that the evolution of computer started from 16th century and resulted in the form that we see today. The present day computer, however, has also undergone rapid change during the last fifty years. This period, during which the evolution of computer took place, can be divided into five distinct phases known as *Generations of Computers*. Each phase is distinguished from others on the basis of the type of *switching circuits* used.

(i) First Generation Computers

First generation computers used *Thermion valves*. These computers were large in size and writing programmes on them was difficult. Some of the computers of this generation were:

ENIAC: It was the first electronic computer built in 1946 at University of Pennsylvania, USA by John Eckert and John Mauchly. It was named Electronic Numerical Integrator and Calculator (ENIAC). The ENIAC was 30´50 feet long, weighed 30 tons, contained 18,000 vacuum tubes, 70,000 registers, 10,000 capacitors and required 150,000 watts of electricity. Today your favourite computer is many times as powerful as ENIAC, but still its size is very small.

EDVAC: It stands for Electronic Discrete Variable Automatic Computer and was developed in 1950. The concept of storing data and instructions inside the computer was introduced here. This allowed much faster operation since the computer had rapid access to both data and instructions. The other advantage of storing instruction was that computer could make logical decisions internally.

Other important first generation computers

EDSAC: It stands for Electronic Delay Storage Automatic Computer and was developed by M.V. Wilkes at Cambridge University in 1949.

UNIVAC-1: Ecker and Mauchly produced it in 1951 by Universal Accounting Computer setup.

Limitations of first generation computers

The following are the major drawbacks of First generation computers:

- The operating speed was quite slow.
- Power consumption was very high.
- It required large space for installation.
- The programming capability was quite low.

(ii) Second Generation Computers

Around 1955 a device called the *Transistor* replaced the bulky electric tubes in the first generation computer. Transistors are smaller than electric tubes and have higher operating speed. They have no filament and require no heating. The manufacturing cost was also very low. Thus the size of the computer got reduced considerably.

It is in the second generation that the concept of Central Processing Chapter (CPU), memory, programming language and input and output units were developed. The programming languages such as COBOL, FORTRAN were developed during this period. Some of the computers of the Second Generation were:

- **IBM 1620:** Its size was smaller as compared to first generation computers and was mostly used for scientific purpose.
- **IBM 1401:** Its size ranged from small to medium and was used for business applications.
- **CDC 3600:** Its size was large and was used for scientific purposes.

(iii) Third Generation Computers

The third generation computers were introduced in 1964. They used *Integrated Circuits* (ICs). These ICs are popularly known as *Chips*. A single IC has many transistors, registers and capacitors built on a single thin slice of silicon. So it is quite obvious that the size of the computer got further reduced. Some of the computers developed during this period were IBM-360, ICL-1900, IBM-370, and VAX-750. Higher level language such as BASIC (Beginners All-purpose Symbolic Instruction Code) was developed during this period.

Computers of this generation were small in size, low cost, large memory and processing speed is very high.

(iv) Fourth Generation Computers

The present day computers that you see today are the fourth generation computers that started around 1975. They use *large scale Integrated Circuits* (LSIC) built on a single silicon chip called microprocessors. Due to the development of microprocessor it is possible to place computer's *central processing chapter* (CPU) on single chip. These computers are called microcomputers. Later *very large scale Integrated Circuits* (VLSIC) replaced LSICs.

Thus, the computer which used to occupy a very large room in earlier days can now be placed on a table. The personal computer (PC) that you see these days is a fourth generation computer.

(v) Fifth Generation Computer

The computers of the 1990s are said to be fifth generation computers. The speed is extremely high in fifth generation computers. Apart from this, it can perform *parallel processing*. The concept of *artificial intelligence* has been introduced to allow the computer to take its own decisions. It is still in a developmental stage.

4.3.3 Types of Computers

Now let us discuss the varieties of computers that we see today. Although they belong to the fifth generation they can be divided into different categories depending upon the size, efficiency, memory and number of users. Broadly they can be divided into the following categories:

- **Microcomputer:** Microcomputer is at the lowest end of the computer range in terms of speed and storage capacity. Its CPU is a microprocessor. The first microcomputers were built of 8-bit microprocessor chips. The most common application of personal computers (PC) is in this category. The PC supports a number of input and output devices. An improvement of 8-bit chip is 16-bit and 32-bit chips. Examples of microcomputer are IBM PC, PC-AT.
- **Mini computer**: This is designed to support more than one user at a time. It possesses large storage capacity and operates at a higher speed. The mini computer is used in multi-user system in which various users can work at the same time. This type of computer is generally used for processing large volumes of data in an organization. They are also used as servers in Local Area Networks (LAN).
- **Mainframes:** These types of computers are generally 32-bit microprocessors. They operate at a very high speed, have very large storage capacity and can handle the workload of many users. They are generally used in centralized databases. They are also used as controlling nodes in Wide Area Networks (WAN). Examples of mainframes are DEC, ICL and IBM 3000 series.
- **Supercomputer:** They are the fastest and most expensive machines. They have a high processing speed as compared to other computers. They have also multiprocessing technique. One of the ways in which supercomputers are built is by interconnecting hundreds of microprocessors. Supercomputers are mainly being used for weather forecasting, biomedical research, remote sensing, aircraft design and other areas of science and technology. Examples of supercomputers are CRAY YMP, CRAY2, NEC SX-3, CRAY XMP and PARAM from India.

4.3.4 Various Input Devices of a Computer

The following are the various input devices of a computer:

CD-ROM

CD-ROM (Compact Disc, read-only-memory) is an adaptation of the CD that is designed to store computer data in the form of text and graphics, as well as hi-fi stereo sound. A standard CD is 120 mm (4.75 inches) in diameter and 1.2 mm (0.05 inches) thick and is composed of a polycarbonate plastic substrate (under layer - this is the main body of the disc), one or more thin reflective metal (usually aluminium) layers, and a lacquer coating. The CD-ROM, like other CD adaptations, has data encoded in a spiral track beginning at the centre and ending at the outermost edge of the disc. The spiral track holds approximately 650 MB of data. That's about 5.5 billion bits.

Storage capacity: BYTE

In most computer systems, a byte is a chapter of data that is eight binary digits (Bit) long. The amount of information a computer can store is measured in byte. A byte is the chapter most computers use to represent a character such as a letter, number, or typographic symbol (for example, 'g', '5', or '?'). e.g. to store CAT, it takes 3 bytes.

Kilobyte (KB or kb): 1 kb is equal to 1000 bytes (1024 bytes to be precise). Megabyte (MB): 1 MB is equal to 1000 KB (1024 Kilobytes to be precise). Gigabyte (GB): 1 GB is equal to 1000 MB (1024 MB to be precise).

Keyboard

A keyboard is the primary text input device in most computers. The mouse is also a primary input device but lacks the ability to easily transmit textual information. The keyboard also contains certain standard function keys, such as the Escape key, tab and cursor movement keys, shift and control keys, and sometimes other manufacturer-customized keys. The computer keyboard uses the same key arrangement as the mechanical and electronic typewriter keyboards that preceded the computer.

Mouse

A mouse is a small device that a computer user pushes across a desk surface in order to point to a place on a display screen and to select one or more actions to take from that position. The mouse first became a widely-used computer tool when Apple Computer made it a standard part of the Apple Macintosh. Today, the mouse is an integral part of the graphical user interface (GUI) of any personal computer. The mouse apparently got its name by being about the same size and colour as a toy mouse.

The most conventional kind of mouse has two buttons on top: the left one is used most frequently. In the Windows operating systems, it lets the user click once to send a 'Select' indication that provides the user with feedback that a particular position has been selected for further action. The next click on a selected position or two quick clicks on it causes a particular action to take place on the selected object. For example, in Windows operating systems, it causes a program associated with that object to be started. The second button, on the right, usually provides some less-frequently needed capability. For example, when viewing a web page, you can click on an image to get a popup menu that, among other things, lets you save the image on your hard disk. Some mouse have a third button for additional capabilities. Some mouse manufacturers also provide a version for left-handed people. Some people use a mouse pad to improve traction for the mouse ball.

4.3.5 Computers in Research

The following are some of the uses of computers in research:

- **Word processing:** Replaces type writer as a tool for writing and typing any type of document, be it a letter, memos, legal documents, reports, article, books and so on.
- **Graphics and desktop publishing:** Replaces the pen and r-ACTO knife to simplify creating everything from corporate logos to business graphics, new letters, books, advertising catalogs etc.
- **Database management:** Replaces card, file and file cabinet, making it much easier to manage large collections of data and information, such as mailing lists, accounts payable and receivable, invoices, appointments, students' enrollment, inventories and many more.
- **Spreadsheets:** Replaces the calculator and ledger sheet to speed up any type of statistical and mathematics calculations, from financial analysis to scientific data analysis.
- **Communication:** Replaces the library card, file and book stocks providing instant access to massive volumes of information stored on computer around the world all from one's own desktop. E-mail through computers has emerged as cheaper and faster means of communication. Internet has brought wealth of information on our fingertips. Multimedia techniques have also emerged as important tools of training and communication.
- **Education:** Complementing classroom materials by providing activities in geography, physics, chemistry, biology, mathematics, logic and other subjects. Distant learning through 'Net Varsity' is becoming a reality.

CHECK YOUR PROGRESS

5. Name the first calculating device of the world.
6. What are the five distinct phases of the evolution of computers known as?
7. List two examples of third generation computers.
8. What is the full form of CD-ROM?

4.4. INTERPRETATION AND REPORT WRITING

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for the broader meaning of research finding. The task of interpretation has two aspects viz., (i) The effort to establish continuity in research through linking the result of the given study with those of another, and (ii) The establishment of some explanatory concepts.

In one sense, interpretation is concerned with the relationship within the collected data. Its is also partially analysis of data. Interpretation also extends beyond the data of the study to include the results of other researches, theories and hypotheses. Thus, interpretation is the device through which the factors that seem to explain what has been observed by the researcher in the course of the study can be better understood and it also provides the theoretical conception which can serve as a guide for further researches.

Need for interpretation

The usefulness, success and utility of research findings lie in correct interpretation. Therefore, interpretation is very essential. Interpretation is considered to be the basic component of a research process due to the following reasons:

- It is through interpretation that the researcher can well understand the abstract principle that works as basis of his findings. Through this he can link his findings with those of other studies which have the same abstract principle, and hence, can make predictions. Fresh enquires can test these predictions later on. This way the continuity in research can be maintained.
- Interpretation generates the establishment of explanatory concepts that can serve as a guide for future research studies. It opens new avenues of intellectual adventure and stimulates the quest for more knowledge.
- A researcher can appreciate his findings only through interpretation. He can clearly understand what they are and can make others understand the significance of his research findings.
- The interpretation of the findings of exploratory research study often results in hypothesis for experimental research and as such interpretation is involved

in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such a study have to be interpreted on a post-facto basis, in which case the interpretation is described as a 'post-facto' interpretation.

Precautions in interpretation

It should be kept in mind that even after collecting proper data and correct analysis, wrong interpretation would lead to wrong decision-making. Therefore, it is very important that job of interpretation be done with patience in an impartial manner and also in the correct perspective.

4.4.1 Significance of Report Writing

One of the most important parts of the research work is a research report. Without a research report, the research work is considered incomplete. A well-designed research study, correct hypothesis and striking generalizations and findings are not of much importance till the time they are not presented properly. Research result must add to knowledge. All this shows the importance of writing the research report. Some people are of the opinion that a research report is not a part of research process, but the general opinion is that research report is an essential part of the research process. The last step in a research process is writing of the report. In this step, the researcher should take this step carefully and for this purpose he can take guidance and assistance from experts of same research field.

4.4.2 Steps in Report Writing

The following are the steps involved in writing a research report:

- **Logical analysis of subject matter:** There are two ways to develop a subject, i.e., *logical* and *chronological*. The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. When we develop some matter from simplest possible structure to the most complex structure, a logical connection between these structures is essential. Chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.
- **Preparation of the final outline:** Outlines are the framework upon which long written matter is developed. Outlining helps in logical organizing of written material. It consists of all valid points in a proper sequence.
- **Preparation of the rough draft:** When the researcher conducts a research study, he writes down everything in a sequential manner so that others can understand it better. He writes down the procedure adopted by him to carry on the research study, i.e., the procedure for data collection and analysis. He will also mention all the limitations faced during the study. He will write down the major findings, various suggestions he wants to offer regarding the problem concerned.

- **Rewriting and polishing of rough draft:** More time is required in this step than the writing of the draft. Careful revision of written text is essential because it is this revision that makes the difference between ordinary and excellent writing. The researcher must ensure that there is unity and cohesion of material to be presented. He should check if the rough draft is consistent or not. Everything, including the usage, spelling and grammar, should be checked.
- **Preparation of the final bibliography:** The bibliography which is generally appended to the research report is list of books which are pertinent to research which has been done. It contains all the books and works consulted by researcher for his research study. The arrangement of these should be in an alphabetical order and can be broken down into two parts. While the book and pamphlet names may be in the first part, the names of newspapers, magazines, journals, articles, etc. are given in the second part. Even though this is not the only way by which bibliography is presented, it is the easiest for the readers to understand.
  - **Books and magazines**
    (i) Author's name, last name first
    (ii) Title, underlined to indicate italics
    (iii) Place, publisher and date of publication
    (iv) Number of volumes

  Example: Nag, A. 2010. *International Business Strategy.* New Delhi: Vikas Publishing House.
  - **Articles in magazines and newspapers**
    (i) Author's name, last name first
    (ii) Title of article in quotation marks
    (iii) Name of periodical, underlined to indicate italics
    (iv) Volume and number
    (v) Date of issue
    (vi) Pagination

  Example: Rossa, Robert V. 'Coping with Short Term International Money Flows.' *The Banker,* London. September 1971, p. 995.
- **Writing the final draft:** This is the last step in writing a report. The final draft should be written in a concise and objective style. The language should be simple, avoiding vague expressions such as 'it seems', 'there may be' and the like. While writing the final draft, the researcher should avoid the use of jargons and abstract terminology.

  The final draft should also contain various examples and illustrations based on common experience as these are the most effective in communicating the research findings to other people. We must never forget that the main aim of every report is to try to find a solution to a particular logical problem.

The report must contribute to the solution of the problem and should also increase the knowledge of both, the reader and the researcher.

4.4.3 Format of a Research Report

NOTES

Every reader who is reading a research report should be conveyed enough about the research study so the he can obtain scientific knowledge and can check if the methods used are adequate or not. This will help the reader understand whether the methods used are adequate or not and on the basis of this, they can develop an opinion of how the findings are to be taken. A proper format of the report is required for this purpose. Format of report means the contents of a research report. A format of the research report should comprise of the following:

(i) **Preliminary pages:** The report should carry a title and the date, followed by an acknowledgement in the form of a 'preface' or 'foreword'. To help the reader easily locate the required information in the report, we must also provide them with a table of contents and a list of tables and illustrations.

(ii) **Main text:** The main text not only gives an outline of what the research report is about, but also provides the reader will all the necessary details. The first page of the main text should contain the title of the research study. This should then be followed by other details, starting with the second page. Every new section of the report should begin from a new page. The main text of the research report have the following sections:

- Introduction
- Statement of findings and recommendation
- Results
- Implication drawn from the results
- Summary

Introduction

The main objective of the introduction is to familiarize the reader with the project report. It contains the clear objectives or purpose of the research. Enough information is given to describe why a problem is worth studying or investigating. The definitions of the important concepts used in the study and the hypothesis should be properly explained and discussed in the introduction of the report.

The methodology adopted for conducting the study should be clearly defined and explained for the knowledge of scientific reader, as he is generally interested in knowing about the basic design of study. In case the study is an experimental one, the experimental manipulations should be clearly stated.

If questionnaires or interviews have been used, then exactly what questions were asked under what circumstances should be mentioned. The questionnaire or interview schedule is usually given in an appendix. The reader should be informed about the sample chapter, sample size, etc. These questions help in estimating the possible limits of generalization of the findings. All limitations for conducting the study should be clearly mentioned.

Statement of findings and recommendations

The findings and recommendations should be given in simple words so that every concerned person can understand them easily. If findings are very detailed, they should be put in summarized form at this point.

Results

The next step in the writing of the main text of the report is making a detailed presentation of the findings of the study. This should be supported with tables, charts and a validation of those results. The results form a major chunk of the research report and can extend over several chapters. Statistically, reductions and summaries of the data, rather than the raw data are presented in the result section. The sequence of all results should be scientific and fitted into identifiable sections. All required results should be there in the report. Ultimately the researcher must really use his own judgment to decide the outline of his report. The clear narration of the problem by the researcher with which he was concerned should be there in report. All the procedures adopted to solve the problem and the conclusion derived from these procedures should be mentioned in this result section.

Implications of the result

After finding the results, research should state all the implications of the results. After knowing the results, every reader would be interested in knowing the implication of same result as he wants to have an exact understanding of results related to concerned phenomenon. Such implications must have the following three aspects:

- Inferences from the same study which can be applied in similar conditions in the future.
- Conditions in which the study is applicable: These conditions can limit the extent of legitimate generalization of the inferences.
- Provide answers for questions which are still unanswered or new questions which are raised because of the study.

We should always end the report with a short conclusion, which will enable proper understanding to the reader about the current problem and their implications.

- **Summary:** These days it is a general practice to end a report with a short summary. This summary gives the details of major happenings, results and conclusions of the research study.
- **End matter:** The appendix should be given at the end of the report to mention all technical data such as questionnaires, sample information, mathematical derivations and the like ones. Bibliography of sources consulted should be given. Index, an alphabetical listing of names, places and topics along with the number of pages in the book or report on which they are mentioned or discussed should invariably be given at the end of report.

4.4.4 Final Presentation of Research Report

When the research study is over, all data have been analysed, results have been developed, all implications have been made, we can present the research report in two ways

(a) Oral presentation

(b) Written presentation

Oral Presentation: In many cases oral presentations are considered very effective, particularly when policy formulation is based on recommendations of study. The merit of oral presentation is mutual discussion. Better understanding is generated amongst all concerned persons. Researcher gets the chance to deal with human behaviour.

The main demerit of presentation is the lack of any permanent record related to research study. It may be possible that people are unable to retain in their memory, what has been presented to make oral presentation more effective, many visual-aids are used. These visual aids support the oral presentation and create better impact in understanding complex phenomenon, like use of slides, wall charts, blackboards, ppt, OHP slides are quite helpful in generating better impact and understanding. In conferences, seminars, academic institutions, researchers are using oral presentations with the help of visual aids to explain their research work. But in business field where policy implications are important, final report is presented in written format. Clients cannot be expected to digest a range of quantitative and qualitative data given verbally.

Written Presentation: As client is unable to remember all qualitative data and quantitative data through oral presentation, therefore it is better for a researcher to prepare a written report. There are definite set of considerations to be followed in making a written presentation, which are given below:

- **Size and physical design:** The research report should be written on unruled paper 8.5"*11" in size. The black or blue ink should be used if the report is to be written by hand. A margin of at least one and one-half inches should be allowed at the left hand and at least half an inch at right hand of the paper. There should also be one-inch margins, top and bottom. If research report is to be typed, then all typing should be double spaced on one side of the page only except for the insertion of the long quotations.
- **Procedure:** Various steps should be followed.
- **Format/layout:** Looking into the objective and nature of the problem, format of the report should be decided and accordingly adopted.
- **Treatment of quotations:** Quotation should be placed in quotation marks and double spaced, forming and immediate part of the text. But if a quotation is of four or five written lines then it should be single–space and indented at least half an inch to the write of the normal text margin.

- We should keep the following things in mind while writing footnotes:
  - o Footnotes are meant for cross references, citation of authorities and sources, acknowledgement and explanation of a point a view. It should be always understood that footnote is neither an end nor a means of display of scholarship.
  - o Footnotes are customarily separated from the textual material by a space if half an inch and a line about one and half inches long.
  - o Footnotes should be numbered consecutively, usually begins with 1 in each chapter separately.
  - o Footnotes are always typed in single space though they are divided from one another by double space.
- **Documentation style:** About documentation, the first footnote reference to any given work should be complete in its documentation, giving all the essential facts about the edition used. Such documentary footnoted followed a general sequence.
- **Punctuation and abbreviations in footnotes:** Researcher needs to ensure the correct use of punctuations in footnotes, to define the clear understanding of text. Certain English and Latin abbreviations are used in bibliographies and footnotes to eliminate tedious repetition.
- **Use of statistics, charts and graphs:** The use of statistics is very common in research study, as it provides the scientific learning and understanding to researcher and reader both. Use of statistics simplifies the complex issues in research and gives logic to relationships of variables. Statistics are usually presented in the form of tables, charts, bars, line-graphs and pictograms.
- **The final draft:** In final draft research should take care of final writing of report. Language should be simple and easy to understand. Use of jargons should be avoided if the problem is concerned with every human being.
- **Bibliography**
- **Preparation of index**

CHECK YOUR PROGRESS

9. List the steps involved in research report writing.
10. What the various sections of the main text of a research report?
11. What are the two ways in which a research report can be presented?

NOTES

4.5 SUMMARY

- Measurement is a relatively complex and demanding task, especially when it concerns qualitative or abstract phenomena. By measurement we mean the process of assigning numbers to objects or observation, the level of measurement being a function of the rules under which the numbers are assigned.
- Nominal scale is simply a system of assigning number symbols to events in order to label them.
- The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule.
- In case of interval scale, the intervals are adjusted in terms of some rule that has been established as bases for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero but it is not possible to determine for them, what may be called an absolute zero or the unique origin.
- Ratio scale has absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be can conceive zero of length and similarly we can conceive of an absolute zero of time. Ratio scale represents the actual amounts of variables.
- Scaling describes the procedures of assigning numbers to various degrees, of opinion, attitude and other concepts.
- In statistics, regression is a mathematical method of modeling the relationships among three or more variables. It is used to predict the value of one variable given the values of the others. A regression analysis yields an equation that expresses the relationship.
- Prediction (forecasting) is a planning tool which helps management in its attempts to cope with the uncertainty of the future. It starts with certain assumptions based on the management's experience, knowledge, and judgment. These estimates are projected into the coming months or years using one or more techniques such as Box-Jenkins models, Delphi method, exponential smoothing, moving averages, regression analysis, and trend projection.
- A computer is an electronic device which not only does arithmetic calculations quickly, but many other things. It serves different purposes for different people. A computer can not only act as a calculator, but can also solve various problems and manipulate data.

- First generation computers used *thermion valves*. These computers were large in size and writing programmes on them was difficult.
- Around 1955 a device called the *transistor* replaced the bulky electric tubes in the first generation computer. Transistors are smaller than electric tubes and have higher operating speed. They have no filament and require no heating. The manufacturing cost was also very low. Thus the size of the computer got reduced considerably.
- The third generation computers were introduced in 1964. They used *Integrated Circuits* (ICs). These ICs are popularly known as *Chips*. A single IC has many transistors, registers and capacitors built on a single thin slice of silicon. So it is quite obvious that the size of the computer got further reduced.
- The present day computers that you see today are the fourth generation computers that started around 1975. They use *large scale Integrated Circuits* (LSIC) built on a single silicon chip called microprocessors. Due to the development of microprocessor it is possible to place computer's *central processing unit* (CPU) on single chip. These computers are called microcomputers. Later *very large scale Integrated Circuits* (VLSIC) replaced LSICs.
- The computers of the 1990s are said to be fifth generation computers. The speed is extremely high in fifth generation computers. Apart from this, it can perform *parallel processing*. The concept of *artificial intelligence* has been introduced to allow the computer to take its own decisions.
- Research report is a means of communicating research experience to others.
- The purpose of the research report is to communicate to interested persons the methodology and the results of the study in such a manner as to enable them to understand the research process and to determine its validity.
- Research report is a narrative and authoritative document on the outcome of a research effort. It represents highly specific information for a clearly designated audience. It helps in presenting the problem studied, methods and techniques used for data collection and analysis, findings, results and recommendations. It provides the information's which can be utilized for future purposes. It is a way of analysing the researcher's abilities and competency.
- Various steps have to be taken care of by the researcher in presenting the report to client. This presentation could be oral presentation or written presentation.

4.6 KEY TERMS

- **Measurement:** It is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence.
- **Nominal scale:** It is a system of assigning number symbols to events in order to label them.
- **Validity:** It is an extent to which a test measures what we actually wish to measure.
- **Regression:** It is a mathematical method of modelling the relationships among three or more variables.
- **Prediction:** It is a planning tool which helps management in its attempts to cope with the uncertainty of the future.
- **CD-ROM:** It is an adaptation of the CD that is designed to store computer data in the form of text and graphics, as well as hi-fi stereo sound.
- **Keyboard:** It is the primary text input device in most computers.
- **Mouse:** It is a small device that a computer user pushes across a desk surface in order to point to a place on a display screen and to select one or more actions to take from that position.
- **Interpretation:** It is the task of drawing inferences from the collected facts after an analytical and/or experimental study.
- **Research report:** It is a means of communicating research experience to others.

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Nominal scale and ordinal scale are two of the most widely used measurement scales.
2. The situation and the instrument are two possible sources of error in measurement.
3. The sound measurement must meet the tests of validity, reliability and practicality.
4. The scaling procedures may be broadly classified on the basis of subject orientation and response form.
5. ABACUS was the first calculating device of the world.
6. The five distinct phases of the evolution of computers are known as generations of computers.
7. IBM-360 and ICL-1900 are two examples of third generation computers.
8. The full form of CD-ROM is compact disc, read-only-memory.

9. Logical analysis of subject matter, preparation of the final outline, preparation of the rough draft, rewriting and polishing of rough draft, preparation of the final bibliography and writing the final draft are the steps involved in the writing of a research report.
10. Introduction, statement of findings and recommendation, results, implication drawn from the results and the summary are the various sections of the main text of a research report.
11. A research report can be either presented orally or in a written form.

4.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write a short note on measurement in research.
2. What is scaling?
3. State three advantages and disadvantages of the Likert type scale.
4. Define regression and prediction.
5. Write a short note on the fifth generation of computer. How is it different from the fourth generation?
6. State the limitations of the first generation computers.
7. Distinguish between microcomputer and mainframe computer.
8. What do you understand by interpretation?
9. Write a short note on the need for interpretation.
10. Write a short note on the significance of report writing.

Long-Answer Questions

1. Explain the various scales of measurement.
2. Discuss the test of validity, test of reliability and the test of practicality.
3. Explain the various scaling techniques.
4. Discuss the linear regression analysis in detail.
5. What are the methods of constructing regression equation? Explain in detail.
6. Write a note on forecasting methods.
7. Explain the various generations in computer technology.
8. Explain the various types of computers.
9. What is the role of computer in research?
10. Explain the various steps in the process of report writing.
11. Explain the format of the research report.
12. Discuss the final presentation of the research report in detail.

4.9 FURTHER READING

Booth, Wayne. 2008. *The Craft of Research*, Third edition. Illinois: University of Chicago Press.

Creswall, John W. 2008. *Research Designs: Quantitative, Qualitative and Mixed Methods Approaches*. London: Sage Publications.

Christenson, Larry B. *et al.* 2010. *Research Methods, Design and Analysis*, Eleventh edition. New Jersey: Allyn and Bacon.

Kothari, C. R. 2008. *Research Methodology: Methods and Techniques*. New Delhi: New Age International.

Kumar, Ranjit. 2010. *Research Methodology: A Step-by-Step Guide for Beginners*, Third edition. New Delhi: Sage Publications.

NOTES