SYLLABI-BOOK MAPPING TABLE

Communication Research

Syllabi	Mapping in Book
Unit I Introduction to mass communication research, meaning of research, scientific method, characteristics, steps in research identification and formulation of research problem in communication research.	**Unit 1:** Introduction to Mass Communication Research **(Pages: 3–15)**
Unit II Basic elements of research, concepts, definitions, variables, hypothesis and causation. Hypothesis: Type of hypothesis, characteristics of good hypothesis, hypothesis testing, survey research, content analysis and historical method. Measurement: Meaning levels and types of measurement.	**Unit 2:** Basic Elements of Research **(Pages: 17–40)**
Unit III Sampling in communications research, types, their applications and limitations. Methods of Data Collection: Interview, questionnaire, observation and case study, applications and limitations of different methods.	**Unit 3:** Sampling in Communication Research **(Pages: 41–61)**
Unit IV Use of statistics in communication research, basic statistical tools, measures of central tendency (mean, mode and median), measures of dispersion (standard deviation), correlation and chi square.	**Unit 4:** Statistics in Communication Research **(Pages: 63–86)**
Unit V Data processing, analysis, presentation and interpretation of data, use of graphics in data presentation. Writing a research proposal, writing research report, components and style.	**Unit 5:** Data Processing **(Pages: 87–102)**

CONTENTS

INTRODUCTION 1

UNIT 1 INTRODUCTION TO MASS COMMUNICATION RESEARCH 3-15

1.0 Introduction
1.1 Unit Objectives
1.2 Mass Communication Research
1.3 Research—Meaning and Scientific Definition
1.4 Research Methods
1.5 Characteristics of Communication Research
1.6 Steps in Identification of Research
1.7 Summary
1.8 Key Terms
1.9 Answers to 'Check Your Progress'
1.10 Questions and Exercises
1.11 Further Reading

UNIT 2 BASIC ELEMENTS OF RESEARCH 17-40

2.0 Introduction
2.1 Unit Objectives
2.2 Basic Elements of Research
2.3 Concept and Definition of Research
2.4 Variables
2.5 Hypothesis and Causations
  2.5.1 Hypothesis
  2.5.2 Hypothesis Testing
  2.5.3 Types of Hypothesis
  2.5.4 Characteristics of a Good Hypothesis
  2.5.5 Causation
2.6 Survey Research
2.7 Content Analysis and Historical Method
2.8 Measurements — Meaning, Levels and Types
2.9 Summary
2.10 Key Terms
2.11 Answers to 'Check Your Progress'
2.12 Questions and Exercises
2.13 Further Reading

UNIT 3 SAMPLING IN COMMUNICATION RESEARCH 41-61

3.0 Introduction
3.1 Unit Objectives
3.2 Sampling in Communication Research
  3.2.1 Types of Sampling
  3.2.2 Application and Limitations of Sampling

3.3 Methods of Data Collection
3.3.1 Interviews, Questionnaires and Observation
3.4 Case Study Application
3.5 Limitations of Different Sampling Methods
3.6 Summary
3.7 Key Terms
3.8 Answers to 'Check Your Progress'
3.9 Questions and Exercises
3.10 Further Reading

UNIT 4 STATISTICS IN COMMUNICATION RESEARCH **63-86**

4.0 Introduction
4.1 Unit Objectives
4.2 Use of Statistics in Communication Research
4.2.1 Basic Statistical Tools
4.3 Measures of Central Tendency
4.4 Measures of Dispersion
4.4.1 Standard Deviation
4.4.2 Chi-square Test
4.4.3 Correlation
4.5 Summary
4.6 Key Terms
4.7 Answers to 'Check Your Progress'
4.8 Questions and Exercises
4.9 Further Reading

UNIT 5 DATA PROCESSING **87-102**

5.0 Introduction
5.1 Unit Objectives
5.2 Data Processing
5.3 Data Analysis
5.4 Presenting and Interpreting Data
5.4.1 Data Presentation
5.4.2 Role of Graphics in Presenting Data
5.4.3 Data Interpretation
5.5 How to Write a Research Proposal
5.6 Writing Research Report-Component and Style
5.7 Summary
5.8 Key Terms
5.9 Answers to 'Check Your Progress'
5.10 Questions and Exercises
5.11 Further Reading

INTRODUCTION

NOTES

This book, '*Communication Research*', provides know-how about the means to gather research evidence, build up arguments pertaining to research and gives rise to critical thoughts about them. This book is valuable for those who have little or no research background. It also discusses and elaborates on fundamental research issues and provides building blocks for enhanced study. It offers the reader with both, comfort and knowledge. The contents of this book are designed to provide learning on how to begin and progress with researches. This is done by making reasoned cases and offering research conclusions.

Communication Research aligns methods which give an overview of the concepts, methods and tools by which communication research is designed, conducted, interpreted and critically evaluated. This book contains a detailed discussion of fundamental research issues and it also provides building blocks for further study, blended with both, clarity and knowledge.

Each unit begins with an Introduction to the main topic, followed by an outline of Unit Objectives. The topic is then explained in detail, in a way that is easy to understand. The units comprise of 'Check Your Progress' questions to test the understanding of the reader. Each unit has a Summary, a glossary of Key Terms, Answers to 'Check Your Progress' and Questions and Exercises. At the end of each unit, Further Reading lists the names of other books and websites which are related to the topic.

UNIT 1 INTRODUCTION TO MASS COMMUNICATION RESEARCH

NOTES

Structure

1.0 Introduction
1.1 Unit Objectives
1.2 Mass Communication Research
1.3 Research—Meaning and Scientific Definition
1.4 Research Methods
1.5 Characteristics of Communication Research
1.6 Steps in Identification of Research
1.7 Summary
1.8 Key Terms
1.9 Answers to 'Check Your Progress'
1.10 Questions and Exercises
1.11 Further Reading

1.0 INTRODUCTION

This unit provides a basic understanding of mass communication research. Research is the interpretation of a topic by means of thorough study and analysis. Scientific research of a topic or an entity helps us to understand the same in a better way and from a scientific point of view. It is a combination of many observations and experiments. This unit offers an introduction to key research methods and mass media communication processes. We see how various steps are involved in identification and formulation of research problem in communication research.

1.1 UNIT OBJECTIVES

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

- Describe mass communication research
- Explain the meaning and scientific definition of research
- List the various methods of research
- Define the characteristic of communication research
- List the various steps involved in the identification of research

1.2 MASS COMMUNICATION RESEARCH

NOTES

Mass communication can be defined as the process of broadcasting and transferring information to a large group or section of people. Since the message has to be broadcasted to a large section of people at the same time, i.e., a town, city or a country, therefore it requires some form of media. In fact, media is only means of relaying information to the entire, targeted, desired or intended mass at the same time. One of the characteristic of communication research everywhere (whether good or bad) is that it tends to respond to circumstances and events of the 'real world'. As a result, this is the central feature of the media which is continuously into reporting back to the society about the reality and happenings around us. Another reason for this is that the news media frequently becomes implicated in events that they report. Communication research is an indication of the power that the media exercises over their audiences. Some media have identified film stars as manifestations of 'pseudo-individuality'. This is maintained through shallow variations of physical appearance, which seek to mask the inherent standardization of life and conformity in industrial societies (Horkheimer and Adorno 1972).

In brief, mass communication can be defined as the process of producing a large number of messages to be sent to a large number of people. The Internet, television and magazines and newspapers are all the means of communication that are used to transmit messages to large groups or mass of people.

In earlier days, when a message was required to be broadcasted to the entire population, a messenger was sent throughout the kingdom. The ancient means of communicating for sending information or messages to a large section of the population were bugles, fire towers, smoke or drums. These were used to send information to people staying in areas spread across vast geographies. In the more recent times (after invention of the printing press), information was published in books or newspapers. These practices were also carried out to communicate information to the masses. Nowadays, there are three main avenues of mass communication, i.e., publishing, broadcasting and digital communication.

It is very important to research or invent highly efficient means of communication which are used by a large number of people. Basically, such researches are conducted through surveys. The data which is collected by means of a survey can be analyzed by researchers in terms of quantity and quality, before using or publishing the results. Research also affects marketing and advertising people from Madison Avenue. In modern times, as social media has emerged for use in mass communications and as a way of collecting research data, many methodologies are available at different rates for mass communication research.

Mass communication research provides the ability to understand various aspects of research and different theories, in every sector of the field of communication. It also plays the role of an international forum which focuses on the people who are academically or professionally inclined towards the current

NOTES

research in communication and its related fields. Mass communication has a structure that comprises of many levels. This is because this field consists of bureaucratic businesses and groups that are into the business of production of diverse messages for many consumers. This mass production of messages affects independent families as well as social systems. The structure of the process of mass communication attracts attention at various levels. There is complete absence of cross level integration in mass communication research. Researchers from varying intellectual backgrounds end up working at one particular level. The theories and methodologies used by them are most common at that level and rarely cross their level of existence. This is due to the existence and emergence of real-life limitations. These limitations prevent their theories and experiments from moving either vertically or horizontally, away from their present level, to consider both production and consumption. The basis of the development of this field is the existence of a wider vision to take mass communication into account, as a process from production to consumption and also vertical multilevel perspectives which interconnect different theoretical maps. Methodological pluralism comprises of the combination of hard and soft data. This data comes from different sources and is required for prospects at various levels of mass communication research. One of its requirements also focuses on identifying and defining the units of analysis at various levels. The key challenge here is the application of available research options for creating new research strategies. The study of multilevel research strategies and methodological tools always serve as an important support to be applied to mass communication research (Price, 1988).

Why mass communication research?

The following reasons support the advent and success of mass communication research:

- The study and theories which are applied in all areas of communication, provide a detailed overview of the field of mass communication.
- Rigorous and empirical analysis results in a research which is dependable and of superior quality, with very scarce possibility of errors.
- Experienced and knowledgeable editorial assistance indicates a wide range of interests which may be within or external to the conventional boundary of communication discipline. This provides an insight into the proceedings and the expected results.
- The consideration of more than one discipline provides clarity of the communication processes and their result.
- Guest-edited, special issues bring about in-depth examination of a specific area of importance and its relevance.

Since mass communication is a complicated field, hence it requires in-depth research to achieve more information on its characteristics and existence in the society.

1.3 RESEARCH — MEANING AND SCIENTIFIC DEFINITION

NOTES

Research is an organized method of gathering, processing and studying information for better understanding of topic of study, or the subject of research. The very important thing about it is that it contributes to the understanding of the phenomenon and communicates that understanding to others. It enables one to move further to find out what development is going on in this field and whether or not it needs further study or research.

The definition of research in a broader perspective includes the collection of data, information and facts for the advancement of knowledge, after further analysis and interpretation. Watching news on TV, surfing the Internet or reading any type of research book is also a kind of research. The world of science prefers to use the word 'review' instead of research to describe the process of learning, which is one of the major principles of defining a scientific research.

Scientific definition

Research is a process of performing a methodical study to find an answer to a specific question. The prime goal of any experimental process is to find a definite answer. The conduction of research process must be structured and should follow a strict and standard protocol (a set of rules). These rules are more or less similar, but may vary slightly across different fields of science, based on issue and subjects.

A research should be planned well, such that it must include various reviews of articles. The Internet may be checked for literature of past researches that are available in the market. This helps in evaluating the kind of questions that need to be answered. Any research, whether it is scientific, economic, historical or relating to mass media research requires some kind of interpretation and an opinion from different researchers who are involved in it.

1.4 RESEARCH METHODS

A research always begins with planning. The effectiveness and efficiency of a research increases with the focus and clarity on it. The more focused a researcher is, the lesser will be the time taken to complete the research. This will ultimately lower the overall cost of the research. This process broadly includes planning, selecting the method of research, focus group, data collection and finally analyzing, interpreting and reporting results.

Planning

Research plans are formulated on information which may be required, depending on the availability of resources. Some times an organization wants to know everything about its products, services, programs, etc. The planning of a research

is based on the information that is required to be gathered to make major decisions about a product, service, program, etc. Those who have very little or limited resources and are planning to start a research, can take help from different methods and combine a unique blend of details and information. They can understand more about certain areas of their products, services and programs (Lemert, 1981).

Further planning needs to be considered to design a research approach. While designing a research plan, the researcher should address the following questions:

- Who are the target population to benefit from the research information?
- What is the purpose of the research?
- What decisions are anticipated from the result of the research?
- What sources should be used to collect information (employees, customers, groups of employees or customers, certain documentation, etc.)?
- What is the deadline for collecting information (gives a limited time to collect it)?
- What resources in hand can be used for gathering information?
- Which are the most reasonable devices that can be used for collecting data (e.g., interviews, questionnaires, observing staff and/or clients in the program, examining documentation, dividing the workforce and customers into focus groups, etc.)?

Selecting the research method

Selecting the research method is a very tough task that needs minute observation and clear cut decision-making. If the selected method is not appropriate then the desired result would not be achieved. Selection of the most fundamental techniques for business research implies the strategy of extracting vital information required by the main decision-makers, in an economic and lucrative way (Munoz, 2003). Before finalizing the research method, the following questions must be considered:

- What kind of information is needed for making current decisions?
- What is the amount of information that can be collected and analyzed practically and cost-effectively?
- What will be the level of accuracy of the information?
- Will the employed methods be successful in getting all the information that is needed?
- In case added information is required, what additional methods may and must be used?
- Will decision makers accept the credibility of the information?
- Will the target audience agree with the methods that are used (e.g., will they exercise care in filling out questionnaires, participate in interviews or focus groups, etc.)?

- Is training required to administer the methods? Who is the right person to administer the method?
- How is the analysis of information to be done?

NOTES

A researcher uses a combination of methods in his research, for example, a questionnaire is used to collect maximum information from a large number of people, quickly. The respondents are then interviewed to get more in-depth information. Further, case studies can also be used for more in-depth analysis of notable and unique cases. Usually, four levels of information are gathered, by means of reactions and feelings, learning, changes in skills and effectiveness. So the usefulness of the results of the research depends on how far the results are effective.

Focus group

Focus groups are very efficient means for evaluating new services or testing new ideas. Essentially, focus groups are collective and simultaneous interviews of 10 to 20 people of one common group. They can be very useful. One can get ample amount of information during a focus group session. The planning broadly depends on the focus group which decides the results and the intention of the research. Focus groups comprise of 6 to 10 members who have certain similarities, e.g., parallel age group, status in a program, participative and reflective members, etc. People who do not know each other are given the preference to be members, otherwise the outcome of the result will not be as desired.

Data collection

Data collection is a combined process of gathering and preparing the data for use. Recording is the main purpose of collecting data. The recorded data is used for decision-making that pertains to vital issues. This data can also be used as information that can be passed on to others. Primarily, data collection is carried out with the purpose of providing information pertaining to a particular topic. The process of data collection needs proper planning that includes activities like, agreement of goals, target data, definitions, methods, etc. Precollection is the most critical step in the activity of data collection. Once precollection is done, data can be collected from the field by interviewing or other methods. This activity can be carried out in an organized, planned and methodical way. It is usually collected by two means, namely, personal interviews and mail questionnaires.

- Census

Census is the activity of collecting the data which pertains to every person or thing that forms part of a particular group or section of the population. The advantages of a census are that it is very accurate and detailed. Its disadvantages are that it is expensive and time-consuming.

- Sample survey

A sample survey focuses on a core section of the target population. Its advantages are that it is cost-effective and consumes less time. On the other hand, its

disadvantages are that it is not accurate and lacks details. Sometimes, surveys are biased. This results in increased chances of error due to manipulated data.

Research method finally concludes with analyzing, interpreting and reporting results for which various statistical tools like SPSS (a computer program which is used for statistical analysis) are used. A data file is used to enter the data which is collected. When data on the questionnaires is converted into a format that can be read by a computer, the process is known as data entry.

Variable values which are in the form of open answers, should be identified and categorized. This operation is known as coding. The data thus achieved, needs to be edited for its identification and correcting the errors. The errors can be checked by using appropriate measures, for instance, repetitive checking of the source of information. Any changes that appear after these checks should be updated. The processing function comprises of: data entry, coding, editing, checking and update/correction. Together, all these functions are defined as the data preparation step of the survey process.

Check Your Progress

1. Define mass communication.
2. Which is the only means of relaying information to a large population, at the same time?
3. How was a message communicated in early days?

1.5 CHARACTERISTICS OF COMMUNICATION RESEARCH

The basic characteristic of a communication research is that it follows a structured procedure. It is employed because of its systematic nature. A communication research follows certain preestablished rules and regulations, from its inception stage to its final stage. Some of its other characteristics are as follows:

Research is logical: Without doctoring or structuring ideas logically, a scientific research cannot progress in any investigation.

Research is reductive: As a rule, being reductive refers to reducing a researcher's responsibility. This means that the findings of one researcher are transferred to other researchers to prevent them from repeating the same research.

Research is replicable: In situations, where other researchers are interested in previously investigated questions, they may want to confirm the findings of a previous research in new environment, with a new group of subjects, or at a different point in time.

NOTES

Research is generative: This is the key feature of a research, because answering one question leads to the emergence of many new questions which call for further investigation and findings. In other words, it can be said that the characteristic of a research are constant. On a broader perspective, research can be characterized such that:

- The aim is the conclusion
- The procedures are public
- The inferences are not certain
- The substance is the strategy

Understanding the characteristics of research help one to design one's own study better. It also helps in reading, evaluating and carrying out research systematically. These characteristics are:

- A realistic concentration
- The educator-cum-researcher's self-performance
- Teamwork
- Continuous development
- A plan of action
- Keeping the research and its findings accessible

A few of the common mistakes which may be usually committed at the time of carrying out any research are:

- The hypothesis is mistaken for an explanation (which fails to perform an experimental test)
- Ignoring or ruling out data that does not support the hypothesis
- Failure to estimate all errors quantitatively

1.6 STEPS IN IDENTIFICATION OF RESEARCH

To carry out a good research, a researcher must be specific about the topic. He must know what is to be said and it must be said effectively. At the time of selecting a topic, a subject which is most interesting for the researcher should be chosen. The researcher should also be able to collect ample information about the subject of research. In case, the content of the subject to be researched is extensive, it will be difficult to complete the research. The following points also need to be considered before conducting a research:

- Identification of the problem
- Formulation of research problem and objectives
- Development of a theoretical framework
- Designing of a research methodology
- Collection, analysis and interpretation of data

NOTES

- Presentation of conclusion and recommendation
- Planning a course of action

Apart from this, information needs to be located from a variety of reference sources. These sources include encyclopedias, almanacs, scholarly journals, books, magazines and newspapers. These sources should be secured in a printed form, on CD-ROMs and on the Internet. This is helpful in getting related information about the research which is being carried out.

An outline for the paper needs to be prepared by organizing notes from note cards into topics, subtopics, details and subdetails. A typical and organized format for the same is as follows:

(i) (topic)

a. (subtopic)

b. (detail)

c. (sub detail)

A researcher must prepare note cards and outline to write a rough draft of his research. As a draft is written, numbered footnotes may be used to credit sources from which quotations or major ideas have been taken. Any changes that are needed to be ensure that the ideas are clearly expressed and the written matter has accurate spelling and grammar may be incorporated. At the end of the research, it is important to maintain a record of the means used to collect information for the research. This helps in preventing last minutes tension and errors. This information may be provided by bibliography cards. As a research approaches its final stage, the researcher may require ample references for information about various topics, to locate facts and answer questions.

Formulation of research problems

Research topic: Discussion and study on the formulation of research problems begins with the source of research topics. The researchers come up with ideas of a research project when they experience practical problems in the field. There is direct involvement of researchers in the implementation of health, social and human service programmes. Because of their involvement, they get several ideas which are based on what they observe around them. Other people who work with (or survey) the researchers also come up with various inputs, though most of these may seem meaningless to anyone who is not part of the research.

The literature review: Literature is also a good source of research ideas in the specific area. Researchers also get inputs from reading literature or trying to extend or refine a previous research. RFPs (request for proposals) act as a good source of research ideas. These RFPs are circulated by government organizations and other companies. The agency may want a research on some of the problems which are described in these RFPs.

NOTES

Review of literature is a very critical initial step of any research project. This review also shows that whatever idea a researcher has selected, has already been thought of, at least to some extent.

Following are a few of the tips to conduct a literature review:

- All attempts should focus on the scientific literature. To start a research, the most reliable research journals should be targeted. It is important to promote research journals that use a blind review system.
- The review needs to be done early in the research process. A literature review provides a significant amount of learning to the researcher. This helps him in making tradeoffs that may arise later.

A researcher may come across a study which is similar to the one he is planning to do. The review of literature of other research studies can also be referred, to have a quick and easy start. Several important constructs need to be included in a research study. Previous research work helps to ensure that a researcher has included all major and relevant constructs in his study. After reviewing the literature, he may observe that many studies constantly look for an outcome, which has not been included in his study. Without that construct the study might not be considered relevant or credible.

A review of literature of similar studies also gives an idea about what measurement instruments may be appropriate for the present study. Finally, the review of literature also helps one to learn from the experiences of others, to anticipate common problems and avoid shortcomings if any.

Feasibility of study: One major question that may arise in the researcher's mind is about the feasibility of the study. A few major decisions are made between rigor and practicality. When a study is carried out scientifically, a researcher is forced to take decisions and perform activities, which he may not normally opt for. The implementation of the program needs to be controlled more carefully. If unlimited resources are available and there is unbridled control over circumstances, a researcher would be able to conduct the best quality research. However, since it is difficult for the circumstances to be ideal, so almost every time researchers are more likely to opt for profitable exchanges where they find to get the desired rigor.

When it is checked whether a research project is feasible or not, the following practical considerations should always be considered:

- Completion time of the research
- Ethical constraints, if any
- The cooperation that is required completion of the project
- Cost of conducting the research

If the researcher fails to consider any of these factors, it will lead to disaster later.

CHECK YOUR PROGRESS

4. What is the basic characteristic of a communication research?
5. What is important for a researcher, to carry out a good research?
6. Name some sources of reference that are used in a research.

1.7 SUMMARY

- Mass communication is the process where messages are transmitted from a sender to a large number of recipients, through various types of media, for instance, newspapers, television, magazines, etc. The medium which is chosen for mass communication has the ability to reach a huge audience.
- Many methods of research and approaches are used for the study of media and the process of mass communication.
- Research is an organized method to investigate and establish the facts about a particular topic or subject.
- Any process that includes the collection of data, information and facts for the enhancement of knowledge is scientifically known as a research.
- This unit has provided an organized introduction to the main methods of research, which are used in the study of mass communication and media.
- Mass communication research is the study of the means of transfer of information, through media, to a large group or section of the population

1.8 KEY TERMS

- **Scientific research:** A research which is based on scientific investigation and theories
- **Broadcasting:** The process of transmitting information to audiences through a blend of audio and video signals
- **Media:** The key means of transferring information and entertainment to a large number of people
- **Protocol:** A system of fixed rules and formal behaviour used at official meetings
- **Cost-effective:** A most economic outcome or product for the users/end consumers
- **Census:** The process of officially counting something, especially a country's population and recording various facts

NOTES

- **Survey:** An investigation of the opinions, behaviour, etc., of a particular group of people
- **Collaboration:** The act of working with another person or group of people to create or produce something
- **Bibliography:** A list of books or articles about a particular subject or by a particular author

1.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Mass communication can be defined as the process of broadcasting and transferring information to a large group or section of people.
2. Media is only means of relaying information to a large population at the same time.
3. In earlier days, when a king was required to broadcast a message, a messenger was sent throughout the kingdom.
4. The basic characteristic of a communication research is that it follows a structured procedure.
5. To carry out a good research, it is important for a researcher to focus his facts specific to the topic.
6. Some sources of reference that are used in a research include encyclopedias, almanacs, scholarly journals, books, magazines and newspapers.

1.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is scientific research?
2. Define media.
3. What is one main characteristic of communication research?
4. List a few common mistakes that may be committed at the time of carrying out a research.
5. What points need to be considered before carrying out a research?

Long-Answer Questions

1. Explain the concept of mass communication research.
2. Discuss the meaning and scientific definition of research.
3. Analyse the parameters for selecting a method of research.
4. Write a note on the characteristics of communication research.
5. How are the steps within a research identified?

1.11 FURTHER READING

Baley, Kenneth. 1987. *Method of Social Research.* New York: Free Press.

Horkheimer, Max and T. W. Adorno. 2002. *Dialectic of Enlightenment.* Stanford: Stanford University Press.

Lemert, J. 1981. *Communication Poll.* Chicago: Nelson-Hall.

Trochim, William. 2006. *The Research Methods Knowledge Base Ohio:* Atomic Dog Publications.

NOTES

UNIT 2 BASIC ELEMENTS OF RESEARCH

NOTES

Structure

2.0 Introduction
2.1 Unit Objectives
2.2 Basic Elements of Research
2.3 Concept and Definition of Research
2.4 Variables
2.5 Hypothesis and Causations
  2.5.1 Hypothesis
  2.5.2 Hypothesis Testing
  2.5.3 Types of Hypothesis
  2.5.4 Characteristics of a Good Hypothesis
  2.5.5 Causation
2.6 Survey Research
2.7 Content Analysis and Historical Method
2.8 Measurements — Meaning, Levels and Types
2.9 Summary
2.10 Key Terms
2.11 Answers to 'Check Your Progress'
2.12 Questions and Exercises
2.13 Further Reading

2.0 INTRODUCTION

Every research should have the following basic elements:

- It should be systematic
- It should be logical
- It should be transferable
- It should be replicable
- It should be generative

This unit discusses about basic elements of research, concept and definition of research, variables, hypothesis, measurement and levels of measurement.

2.1 UNIT OBJECTIVES

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

- Explain the basic elements of research
- Explain the concept and definition of research

NOTES

- Define variables, hypothesis and causation
- List the characteristics of a good hypothesis
- Discuss survey research and content analysis
- Enumerate the levels and types of measurements

2.2 BASIC ELEMENTS OF RESEARCH

The methodological process of collecting, processing and analyzing information to increase our understanding of the phenomenon under study is called research. The most important thing about research is to contribute to the understanding of the event or phenomenon and to communicate that understanding to others. It enables us to move further in knowing the development that is going around in this field and the need for further research or study. Towards a larger perspective, the activities of research include gathering of data, facts and information for the advancement of knowledge. After further analysis and interpretation, the research process has been divided into four major categories:

(i) Specification, measurement and manipulation of theoretical variables
(ii) Selection of a research design
(iii) Analysis of data obtained from study
(iv) Interpretation and robustness of the findings

Research is the process of defining the interrelationships of conceptual, methodological and substantive fields. Conceptual field refers to the concepts and ideas in abstract form, methodological knowledge refers to the designs, strategies, measuring devices, and analytic techniques used to study a phenomenon or theory and the substantive domain refers to the events/processes studied.

2.3 CONCEPT AND DEFINITION OF RESEARCH

Research is a combination of the words, 're' and 'search'. This implies searching again and again to thoroughly understand the issue and find an apt or effective solution or a strategy to deal with the issue. Such a solution or strategy adds to the understanding of the subject. In common parlance, research refers to a search for knowledge. In simple words, research is an endeavour to discover answers to problems (intellectual or practical) through the application of scientific methods. The Webster's international dictionary gives a very inclusive definition of research as a 'careful, critical inquiry or examination in seeking facts or principles, diligent investigation in order to ascertain something'. Research is essentially a systematic enquiry that seeks facts through objective and verifiable methods in order to discover the relationship among them and to deduce from them, broad principles or laws. Research is really a method of critical thinking. It comprises defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data, making deductions and reaching conclusions to

NOTES

determine whether they fit the formulated hypothesis. Thus, the term research refers to a critical, careful and exhaustive investigation, or enquiry, or experimentation or examination. It is aimed at the revision of accepted conclusions, in the light of newly discovered facts. Francis G. Cornel defines research as, 'the activity of collecting information in an orderly and systematic fashion'. John Best defines research as, 'the systematic and objective analysis and recording of controlled observations that may lead to the development of generalizations, principles, or theories, resulting in prediction and possibly ultimate control of event'.

2.4 VARIABLES

A variable is usually anything that is changeable. For example, it can be a characteristic or a value. Generally, variables are used in psychological experiments. This is done to determine if changes to one thing result in changes to another.

Independent variable

In a psychological experiment, the variable that is controlled and manipulated by the experimenter is known as an independent variable. For instance, in case a research is conducted to see how lack of sleep affects test performance, the deprived sleep will act as an independent variable.

Dependent variable

A variable that changes according to the independent variable is known as a dependent variable. In some cases, dependent variables would be the scores measured on the test performance.

Extraneous and confounding variables

Apart from independent and dependent variables there is another category of variables. This third category of variables also exists in many experiments and is known as extraneous variables. These type of variables may have an impact on the relationship between independent and dependent variables. When we discuss about a research on how the scarcity of sleep affects test performance, it needs to be realized that other aspects like age, gender and academic background, are also likely to influence the outcome. In these circumstances, the researcher may record the values of these extraneous variables, to control their effects on the results.

Types of extraneous variables

- **Participant variables**

  Variables that are associated with the individual traits of each participant are known as participant variables. Participant variables tend to influence the response of a participant. These variables comprise of factors like, upbringing dissimilarities, frame of mind, concerns, aptitude, consciousness and other characteristics that are unique to each individual.

NOTES

- **Situational variables**

  Variables that are related to environmental characteristics which influence the response of each participant are known as situational variables. In case a participant is undergoing a test in a cold environment, then temperature may be considered as a situational variable. Although, temperature is a situational variable, not all participants are affected by the cold.

  There are a large number of instances where the researcher controls extraneous variables. When participant variables are considered, the researcher may choose participants from similar background and with matching temperaments. This is an attempt to make sure that these characteristics do not influence the outcome. However, in a few cases, it is not possible to control the variables. These variables are known as confounding variables. Confounding variables are likely to influence dependent variables. As a result of this influence, it is not easy to identify whether the outcome is due to the impact of the independent variable, the confounding variable or a combination of both.

Operational definition of a variable

The creation of operational definitions for independent and dependent variables is important. An operational definition provides a description of the process of measuring and defining variables, in a study. Instances from the experiment of how scarce sleep affects test performance may be considered. This requires exact operational definitions for both the variables. If it is hypothesized that students who are not allowed to sleep display drastically low scores in a test, a few different concepts will need to be defined. In this experiment, students may be defined as participants who are registered in a preliminary, university-level course of psychology.

The second step is to operationally define the variable 'sleep deprivation'. In this scenario, participants who were allowed less than five hours of sleep before the test, fall in the category of 'sleep deprivation'. Here, in the preliminary course of psychology, the marks scored by a student in a unit exam are known as the test variable.

Frequently, problems pertaining to identification of independent and variables are registered by students in an experiment. As both, the complexity of the task and that of the experiment increases, a few questions arise for identification of a variable. A likely question can be, 'what is the experimenter manipulating?'. Independent variables have the tendency to vary either naturally, or they can be directly manipulated by the experimenter. Another question would be, 'what is being measured?'. A variable that is being measured by the experimenter is known as the dependent variable.

2.5 HYPOTHESIS AND CAUSATIONS

NOTES

A hypothesis is an assumption about the following:

- Relationship between/among variables
- Level of influence of independent variables on the dependent variable
- Value of population parameter

A hypothesis is an idea or an explanation of something that is observed to occur. This term derives itself from the Greek language, which means 'to put under' or 'to suppose.' To present a hypothesis scientifically, it is important to have a scientific method which can be tested. Usually scientists establish their scientific hypotheses on earlier inferences that they are not able to explain on the basis of the available scientific theories. The words 'hypothesis' and 'theory' are often interchangeable in their use, either commonly or formally. However, a hypothesis which is based on science is not synonymous with a scientific theory. A hypothesis that works is also that which is provisionally accepted.

In case a probable relation between phenomena is looked into, for instance, it is not possible to investigate the hypothesis that a relation exists in the same way one might investigate a proposed new law of nature. There are some instances in these types of examinations, which display no effect. These render the hypothesis false. Statistical tests are alternatively used to know the probability of the overall effect in the absence of any real relation. If this probability is very low (e.g., less than 1%), it can be assumed that a relation exists. Any other observed effect is likely to be purely coincidental.

Within statistical testing of hypothesis, the null hypothesis and the alternative hypothesis are compared. The null hypothesis can be referred to as the hypothesis of no difference. An alternative hypothesis presents itself as an alternative to null hypothesis: it indicates the presence of a relationship between the variables under study. Alternative hypothesis may assume many forms, according to the type of the hypothesized relation. This hypothesized relation may be two-sided or one-sided.

The most popular levels for testing the hypotheses are .10, .05 and .01. The decision to accept an alternate hypothesis or reject a null hypothesis should be taken before collection and inspection of the observations. If this decision is taken after the data to be tested is already known, the test will not be valid.

A significant factor to be noted here is that this procedure is dependent on the size of the sample within the study. If the size of the sample is very small, it would be insufficient to reject a null hypothesis. Hence the size of the sample should be specified beforehand. The size of each statistical test needs to be specified to be used to test the hypothesis.

A hypothesis is a theory that explains an occurrence, that can validated in a way to prove or disprove the hypothesis. The hypothesis is regarded as acceptable

NOTES

for the entire testing period. The aim of the experimenter is to thoroughly test the conditions of the hypothesis. Hypothesis is a very important part of the scientific method. This also holds true for other methods. It is not required for a hypothesis to be correct or incorrect. However, one who formulates the hypothesis should be capable of testing it thoroughly, for instance, if a researcher has proposed hypothesis which says that exposure to X causes Y in laboratory rats, then he or she must also be able to list other factors that which cause Y. Whenever researchers establish the outcome to support a hypothesis, they also publish the steps to approve as well as disprove the it. This renders strength to their case.

There are cases where a hypothesis does not deliver. Such cases are also acceptable since they too support the cause of science. In the instance where exposure to X causes Y, the researcher can also prove that exposure to X does not cause Y. Moreover, a researcher can recommend the need for further research on Y. In this example, the hypothesis proving wrong does not render the substance X safe.

2.5.1 Hypothesis

A number of different approaches are used by a researcher for studying a number of topics or subjects. The process of research may either be informal, simple, formal, or somewhat sophisticated. Every research begins with a general idea which is in the shape of a hypothesis that is irrespective of the type of process. Generally, a research problem or question is framed in the beginning of a research. A research question may either look like a basic question about some issue or phenomena or a question pertaining to the relationship between two or more variables, for example, a typical research question may look like, 'Does the flexibility of work hours improve the productivity of employees?' Another question may look like, 'How do flexible hours influence employees' work?'

A hypothesis is not same as a research question. It is more specific and proposes a predictable occurrence. It is a provisional clause that describes the relation between two or more variables. The key point of distinction between a research question and a hypothesis is that a hypothesis predicts an outcome that is experimental, for instance, a hypothesis may state, 'There is a positive relationship between the availability of flexible work hours and employee productivity.'

The hypothesis is beneficial in the following ways:

- The direction and focus for a research effort is determined by the hypothesis.
- It forces the researcher to give a clear view of the aim of the research.
- Hypothesis decides on the variables that a study will consider and those that will be rejected by the study.
- It forces the researcher define the variables of interest operationally.

Since a research study is based on the hypothesis, a lot of thinking and observation goes into its development. The researcher's skill determines the value of the hypothesis.

2.5.2 Hypothesis Testing

NOTES

Testing of hypothesis is a systematic method which is important for validating the information and helps in decision-making. The steps involved in hypothesis testing are:

- Mention the hypotheses of interest
- Ascertain a suitable test statistic
- State a significant level of statistic
- Specify the decisive rule for rejecting or accepting the null hypothesis
- Gather data and carry out the required calculations
- Decide to reject or accept the null hypothesis

Stating the hypotheses

A researcher usually utilizes two hypotheses in a research study — the null and the alternative hypothesis. The hypothesis which is being tested is known as null hypothesis and it is labeled as H. The null hypothesis is also known as the hypothesis of no difference. It should comprise of a statement of equality (=, ≥, or £). The alternative hypothesis provides an alternative to the null hypothesis and consists of a statement of inequality (≠). The null and the alternative hypotheses complement each other.

It is a belief that the null hypothesis is correct all through the analysis. The analysis is based on null hypothesis, for example, the null hypothesis may indicate the average age of entering college freshmen as 21 years.

H_0 = The average age of entering college freshman = 21 years

If the data that is collected and analyzed shows that the average age of entering college freshmen is more than or less than 21 years, the null hypothesis is not accepted. According to this scenario, there are three ways to specify the alternative hypothesis, they are: (i) the average age of entering college freshman is not 21 years (the average age of entering college freshmen ≠ 21); (ii) the average age of entering college freshman is less than 21 years (the average age of entering college freshmen < 21); or (iii) the average age of entering college freshman is greater than 21 years (the average age of entering college freshmen > 21 years).

The aim of the research determines which alternative hypothesis is to be selected. The preceding second and third examples of alternative hypotheses involve the use of a 'one-tailed' statistical test. This is known as 'one-tailed' since the statement implies a direction (greater than '>' or less than '<'). The first example represents a 'two-tailed' test. There is inequality expressed (age ≠ 21 years), but the inequality does not imply direction. One-tailed tests are used more frequently in management and marketing research. The reason for this is a need that specifies a specific direction of the outcome. For instance, a researcher would be curious to know whether Product A performed better than Product B (Product A's

NOTES

performance > Product B's performance), or vice versa (Product A's performance < Product B's performance), rather than whether Product A performed differently than Product B (Product A performance ≠ Product B performance). In addition, information can be obtained by knowing that employees who work from 7:00 am to 4:00 pm exhibit higher level of productivity than those who work from 3:00 pm to 12:00 am (early shift employee production > late shift employee production), rather than simply knowing that these employees have different levels of productivity (early shift employee production ≠ late shift employee production).

Determining the appropriate test statistic

The most suitable tool for statistics (statistical hypothesis testing) depends on the different features of the sample population of interest. This includes the size and distribution of the sample. The test statistic is capable of assuming more than one arithmetic values. The value of the test statistic significantly impacts the decision. It is important to use the most suitable statistics for achieving a logical outcome.

Specifying the statistical significance level

A null hypothesis may or may not be rejected. A rejected null hypothesis may be true or false in its actual sense. In addition, a null hypothesis which is not rejected, actually, could either be true or false. Nevertheless, the possibility of rejecting a true hypothesis or failing to reject a false hypothesis always exists.

When a null hypothesis that is true is rejected, it is called a Type I error. On the other hand, a failure to reject a false null hypothesis is known as a Type II error. The probability of committing a Type I error is termed α and the probability of committing a Type II error is termed β. As the value of α increases, the probability of committing a Type I error increases. As the value of β increases, the probability of committing a Type II error increases. While one would like to decrease the probability of committing of both types of errors, the reduction of α results in the increase of β and vice versa. The best way to decrease both types of error is to increase sample size.

The probability of committing a Type I error, α, is called the level of significance. Before data is collected one must specify a level of significance, or the probability of committing a Type I error (rejecting a true null hypothesis). There is an inverse relationship between a researcher's desire to avoid making a Type I error and the selected value of α; if not making the error is particularly important, a low probability of making the error is sought. The greater the desire is to not reject a true null hypothesis, the lower the selected value of α. In theory, the value of α can be any value between 0 and 1. However, the most common values used in social science research are .05, .01, and .001, which respectively correspond to the levels of 95 per cent, 99 per cent, and 99.9 per cent likelihood that a Type I error is not being made. The tradeoff for choosing a higher level of certainty (significance) is that it will take much stronger statistical evidence to ever reject the null hypothesis.

NOTES

Determining the decision rule

sIt is important to identify the circumstances which affect the rejection or failure to reject the null hypothesis. This is to be determined before data are collected and analyzed. The calculated statistic or the probability terminology can state the decision rule. Irrespective of the method used, the decision will be the same.

Data collection and performing calculations

No other research techniques, than the ones which have already been approved should be used for collection and analysis of data. These techniques are finalized initially in the research process. After identifying the topic of research, the type of data required and the methods of collection may be decided. This deciding factor determines the method for analysing data.

Deciding whether to reject the null hypothesis

Application of the decision rule is involved in this step. The decision rule facilitates the rejection or failure to reject null hypothesis. If one rejects the null hypothesis, the alternative hypothesis can be accepted. However, as discussed earlier, if one fails to reject the null hypothesis, he or she can only suggest that the null may be true.

Example

XYZ Corporation is a company that is focused on a stable workforce which has a small turnover. XYZ has been in business for 50 years and has more than 10,000 employees. The company has always promoted the idea that its employees stay with it for a very long time and it has used the following line in its recruitment brochures: 'The average tenure of our employees is 20 years.' Since XYZ is not very sure if the statement is still true, a random sample of 100 employees is taken and the average age turns out to be 19 years with a standard deviation of 2 years. Can XYZ continue to make its claim, or does it need to make a change?

(i) State the hypotheses.

$H_0 = 20$ years

$H_1 \neq 20$ years

(ii) Specify the significance level. Since the firm would like to keep its present message applicable to new recruits, it selects a fairly weak significance level ($\alpha = .05$). Since this is a two-tailed test, half of the alpha will be assigned to each tail of the distribution. In this situation the critical values of $Z = +1.96$ and -1.96.

(iii) State the decision rule. If the computed value of Z is greater than or equal to $+1.96$, or less than or equal to -1.96, the null hypothesis is rejected.

(iv) Reject or fail to reject the null. Since 2.5 is greater than 1.96, the null is rejected. The mean tenure is not 20 years; therefore XYZ needs to change its statement.

NOTES

2.5.3 Types of Hypothesis

A research hypothesis is a set of assumptions or suppositions that are suggested to explain the correlation between phenomena or seem to explain an observed pattern in a phenomenon. This observed pattern can be tested either through collection of data or experimentation. A hypothesis does not necessarily constitute a valid theory, unless it is validated with observations and can be accepted as a practical theory only after close scrutiny. There are the following four types of hypotheses:

(i) **Inductive hypothesis** is a generalization which is based on specific observations. The logic that converts specific observations into general theory forms the basis of inductive research. In this case, the researcher's specific outlook is transformed into a specific one. When a hypothesis is framed, researches and surveys are carried out for its verification and a conclusion is thus reached.

(ii) **Deductive hypothesis** is derived from theory and provides evidence that supports, expands, or contradicts the theory. Moreover, deductive ideas form the basis of a deductive research. This research converts general theory into a particular hypothesis, which is fit for testing. Here, the researcher's general outlook is transformed into a specific one. There is a possibility of him, having some idea about human behavior or some other social phenomenon. On the basis of this knowledge, information is gathered to support or disprove the hypothesis. On the basis of deductions, the theory is presented in the form of a hypothesis. This is followed by a technique to test it. If the hypothesis is supported by the data, then it can be concluded that the theory is right.

(iii) **Directional hypothesis** states the expected direction of the relation or difference. In a directional hypothesis, the direction is predictable (it may be either positive or negative) for instance, the performance of girls is better than boys ('better than' indicates a predictable direction).

(iv) **Non directional hypothesis** states that relation or difference between variables exists.

Hypotheses that only indicate the relation between two variables without predicting their effect are known as non directional hypotheses, for instance, in the case of the performance of girls and boys, a non directional hypothesis states how the performance differs without indicating the kind of difference.

(v) **Null hypothesis** states that there is no significant difference or relation between variables. Null hypothesis is the hypothesis of no difference. In a large number of researches, proving a hypothesis is difficult and at times impossible. Therefore, a large number of researchers establish hypothesis which indicates their thoughts about what is going to happen. Any hypothesis has two or more contradictory possibilities, only one of the possibilities is true. A null hypothesis is considered to be true and any other hypotheses besides it are known as alternative hypotheses.

Scientific method

The way of acquiring knowledge through experiment is called a scientific method. It nullifies the standard human bias of reasoning by encouraging replicability and cross-checking. Scientists form hypotheses, about aspects of the world, then test them. These experiments must have the capability to be easily reproduced, so that other scientists can cross-check the data. After thorough testing, a hypothesis may be supported or contradicted by the data.

NOTES

2.5.4 Characteristics of a Good Hypothesis

Following are the characteristics of a good hypothesis:

- It should be based on sound reasoning
- It should be based on previous research
- It should follow the most likely outcome, not the exceptional outcome
- It must provide a reasonable explanation for the predicted outcome
- It must clearly state the relationship between the defined variables
- It must define the variables in easy-to-measure terms
- It should be capable of undergoing any tests within a reasonable duration of time.

It is important to follow some fundamental principles for making sure that the requirements of both, the research and the researcher, are fulfilled by the hypothesis. It is important for an ideal hypothesis to:

- Be logically consistent: It should be aligned with the current research literature and knowledge base. The hypothesis should have a meaning.
- Coincide with the present writing or offer a sound base for any variations. It is not important that the hypothesis supports the present literature but then, there should be a rational logic to disprove it.
- Have the capability to undergo tests. The hypothesis has no meaning if the means for carrying out a research are not established.
- Have clarity and simple terminology and should not be confusing.

2.5.5 Causation

The belief that events occur in a predictable manner, such that one event leads to another, is called causation. It is the 'causal relationship between conduct and result'. In other words, causation offers a way of linking conduct with a resulting effect, characteristically in the form of an injury. In criminal law, *actus reus* (an action) is the terminology which is used to define causation, which gives rise to specific injury or similar effects. This blends with *mens rea* (a state of mind) to contain the components of guilt. Causation is valid and can be applied only where a result has been attained. Hence it does not matter in cases where there is no clarity of the offence.

NOTES

The majority of legal systems are more or less related to the theories of being fair and just. The idea of penalizing a person for incurring losses to another, or framing compensation charges, emerges from the tenet that one who causes injury to another should be held responsible for his deeds. Though, every legal system has clauses for stringent liabilities Some parts of any legal system have qualities of strict liability, in which the *mens rea* is not concerned with the result and its consequent liability. On the other hand, many parts try to establish liability by projecting the defendant as the cause (by application of various legal tests to determine this) of a specific injury or loss. The different degrees of probability and consequences that route from physical activities and behaviour, are easily picked up even by young children. The higher is the predictability of the result, the more likely it is for the actor to cause an intentional injury or loss. The law has many means to secure the simple rule of practical experience.

Check Your Progress

1. What does the word 'research' mean?
2. How does Webster's international dictionary define research?
3. Define a variable.
4. What is an independent variable?
5. What is a dependent variable?

2.6 SURVEY RESEARCH

Survey research is one of the most important areas of measurement in applied social research. Broadly, a survey research comprises of any process of analysis, where respondents are asked questions. A survey may either be a short paper-and-pencil feedback form or a detailed one-on-one interview. These are approximately divided into the following:

Questionnaires and interviews

The survey can begin only after the method of survey has been selected. The various methods that emerge as an instrument of survey result from, decision-making in terms of, questions, content of the questions, phrasing of questions, format of the response and the phrasing and order of questions.

Types of surveys

Surveys can be divided into two categories: questionnaires or interviews. Usually, questionnaires are paper-pencil instruments which are answered by respondents. On the basis of the respondent's replies, an interviewer finalizes an interview. There are instances where it is not easy to differ between a questionnaire and an

interview. For instance, it is commonly believed that questionnaires comprise of short and closed-ended questions, whereas interviews are concerned with broad and open-ended ones. However, there are questionnaires that have open-ended questions and interviews that are based on an array of closed-ended questions.

In the last ten years, survey research has changed drastically. Nowadays, there are telephonic surveys which are computerized and use arbitrary dialing techniques. Public places have automated stalls that seek inputs from the general public. A totally revolutionary form of group interview has emerged as focus group methodology. More and more, survey researches are blended well with the delivery of service. Generally, hotels have a survey questionnaire at the reception. In restaurants, waiters present the customers with a short customer satisfaction survey forms along with the bill.

Disadvantage of survey methods

A large number of the people relate questionnaires to mail surveys. Almost everyone has, sometime or the other, received an email questionnaire. Surveys on emails have a large number of advantages. The most important feature is that they are practically inexpensive to administer. The same mail can be sent across to a large number or people. They facilitate the respondent to fill out the form according to their convenience. However, emailed forms also have a few disadvantages. The responses are usually very scarce. Secondly, mails are not the preferred means of detailed survey.

A questionnaire that is handed out to a group is another way of conducting a survey. Out of a large group, a typical sample of respondents is selected and asked to respond to an organized series of questions. To expedite the process, questionnaires are distributed in groups. In case of any confusion among respondents, pertaining to the questions, they can always seek clarification from the researcher or whoever is conducting the survey. The business can decide the venue for the survey, which may be the company or the business premise.

How does one differentiate between a group-administered questionnaire and a group interview or focus group? A group administered questionnaire is an activity that occurs in a specific area, where every respondent is given a questionnaire and requested to fill it, within that area. One respondent fills one questionnaire. In the group interview or focus group, the progress of the session is controlled by the interviewer. Everyone is a dedicated part of a group. They share comments and respond to queries. Some of them are specifically assigned to note the proceedings of the whole group. None of the respondents in the group complete an interview individually.

Another category of questionnaire is the household drop-off survey. Household in this case may be the respondent's home or workplace (business premises). The researcher provides the questionnaire to the respondent at his household. Usually, the respondent is requested to mail or post the completed questionnaire, which is attached to an envelope with the company's address on it.

NOTES

However, at times the interviewer comes back to pick the completed survey document. This is an effort to combine both, the mail survey and the group administered questionnaire approaches. Similar to a mail survey, the respondent can fill the questionnaire in the privacy of his domain. And like a group administered questionnaire, the interviewer is personally in touch with the respondent and a remote survey instrument is not sent. The respondent has the facility to clarify any doubts that come to his mind, about the questionnaire. By and large, this form of a survey is likely to provide a higher number of responses.

Interviews

Interviews are more personal means of research, as compared to questionnaires. A personal interview is where an interviewer interacts directly with the respondent. The facility provided to the interviewer in a personal interview is the key difference between an interview and a mail survey. A personal interview gives the interviewer the prospect of probing or asking follow-up questions. For seeking views or impact, interviews are more convenient to the respondent. Interviews take more time and require extensive resources. Interviewers are regarded as part of the equipment to measure reactions. Thus, they need to undergo a thorough training in handling any type of contingency.

Telephonic forms of interview are one of the most common forms of survey. Via a telephonic interview, a researcher is able to get information quickly. It is believed by a large number of people that generally, telephonic interviews are the basis of most of the public opinion polls that are reported. They are similar to personal interviews in a way that they facilitate a fair degree of personal contact between the interviewer and the respondent. The interviewer can ask follow-up questions. However, this system also has a few major drawbacks. The majority of the people do not have publicly-listed telephone numbers. Some do not have telephones. Generally, people consider such phone calls as intrusion into their lives. Due to this reason, these phone calls should be short and to the point.

2.7 CONTENT ANALYSIS AND HISTORICAL METHOD

By the time you get to the analysis of your data, most of the really difficult work has been done. It is much more difficult to: define the research problem; develop and implement a sampling plan; conceptualize, test and make the measures operational and develop a design structure. If you have done this work well, the analysis of the data is usually a fairly straightforward affair.

In most social research the data analysis involves three major steps, done in roughly this order:

- Cleaning and organizing the data for analysis (data preparation)
- Describing the data (descriptive statistics)
- Testing hypotheses and models (inferential statistics)

NOTES

Data preparation involves checking or logging the data in, checking the data for accuracy, entering the data into the computer, transforming the data and developing and documenting a database structure that integrates various measures.

Descriptive statistics are used to describe the basic features of data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. Descriptive statistics simply describes what the data shows.

Inferential statistics looks into queries, paradigms and hypotheses. In a large number of instances, the inferences that result from inferential statistics go beyond the instant data. Inferential statistics is used to judge the thoughts of the population from a sample data. On the other hand, inferential statistics are used to judge the probability of a noticed difference between groups. Thus, inferential statistics are used to reach more generic conclusions from the collected data. Descriptive statistics are used to determine the work on the collected data.

In a large number of researches, the analysis section is divided into three stages. The first stage describes the method of preparation of data and focuses on the fact that it should be brief and to the point. The descriptive statistics that is envisaged may be divided into huge volumes. In the majority of the articles these structured and divided into summary, tables and graphs. This structure only illustrates information that is has prime relevance. It is normal for a researcher to correlate every inferential analysis with particular research questions or hypotheses mentioned in the introduction, or notes. In a large number of analyses it is essential to focus on all factors and not be selective. Too many details may divert the reader from the crux of the write-up. The details are usually reserved for the appendices, keeping the critical information in the body of the report.

Historical research

Historical research comprises examining the content and validating the data statistically. This data may have been originally collected for some other purpose. This technique of historical research is referred to by social scientists as 'secondary analysis. A researcher views secondary analysis very importantly because it facilitates him with the capability to assemble research designs at minimum costs. It is also less time-consuming for him. Secondary analysis works on the following types of data sources: newspapers, diaries, census data, vital statistics, church records and opinion polls. Researches on trends are carried out by sociologists. These help them assess attitude and behavioral changes over a period of time.

An organized analysis of existing documents, help in investigating changes in variable(s). Let us take an instance where a sociologist is studying the grief experienced by children who have lost a parent. It is his hypothesis that as time passes by, the extent of the grief will subside. He illustrates this by a 'J' curve. To save time, a researcher zeroes on one best way to collect this information. He advertises personal experiences from dairies that have been written by children (ages 8 through 15) who have lost a parent. This is done by placing advertisements

NOTES

in the most popular newspapers. This may help the researcher in obtaining a number of significant cases. Most of these are likely to approximate a longitudinal design, because there are some children are in the habit of writing diaries. The diaries are studied in an organized way by focusing on the content to get clarity on the behaviour of these children.

Till date, the analysis of this information has been limited to studying and categorizing the inferences. The next step is to establish whether there is sufficient information supporting the observation. In the present scenario, use of developed and automated text editing programs can help in content analysis (see MacTavish, 1985). Computers provide the convenience of typing in accurate details (such as newspaper articles, entries from diaries, speeches, etc.). They also simplify the process of categorizing the text. However, there is a likelihood of the researcher being subjectively biased in interpreting the secondary resources.

Going back to the example of content analysis of children's diaries, the historical method has some drawbacks. The most common disadvantage is that the diary containing this experience may not be specific to this feature. It may be a general diary which has all sorts of experiences. As a result, the major part of the information may not be of any use to the researcher, who is focused on a particular hypothesis.

A second major obstacle in this study is the likelihood of the researcher's bias in interpreting the material. If one child writes 'Today would have been daddy's birthday,' is this an expression of bereavement or a statement of fact? It is for the researcher to take a decision on this without going through the thought process of the subject. This raises doubts about how reliable and valid are the indicators of grief. Another drawback of content analysis is that it takes a lot of time to analyse historical materials. For instance, if a 100 diaries are received from children, it would take very long to read and interpret the information in them. (this excludes the additional time required to enter this data into a computer).

Disadvantages of the historical method

(i) Researchers are likely to be biased in interpreting historical sources.

(ii) Interpreting sources is very time consuming.

(iii) Computerized content analysis is costly to quantitatively analyze—programs of this type take large blocks of computer core time and make analysis much more expensive than standard statistical procedures, which are used in evaluating survey data.

(iv) The sources of historical materials may well be problematic—for example women are more likely than men to keep diaries, not all records are kept in consistent patterns, original authors bring their own perspectives and biases to the interpretation of events.

(v) Due to the lack of control over external variables, historical research is very weak with regard to the demands of internal validity.

Some of the advantages of historical analysis include the following:

(i) The historical analysis is not interfering—it has no impact on the outcome of the research.

(ii) It is designed to support trend analysis.

(iii) In comparison to longitudinal designs, content analysis is generally more cost-effective.

(iv) There exists no likelihood of interaction between the researcher and the subject.

NOTES

Historical methods are more used in sociology and sociological researches. The reason for their wide use is their cost-effectiveness and availability. It is also to be kept in mind that it is possible to blend content analysis with other research designs. These analytical techniques are also useful in interpreting open-ended responses to questionnaires, or in understanding qualitative data that results from face-to-face interviews.

2.8 MEASUREMENTS—MEANING, LEVELS AND TYPES

The task of managing analyses and records conclusions as part of a research methodology. Here, let us consider two major concepts of measurement. In the level of measurement, we first explain the meaning of the four major levels of measurement: nominal, ordinal, interval and ratio. Then we move on to the reliability of measurement, including consideration of true score theory and a variety of reliability estimators.

The various measures used in a social research need to be understood. The four main types of measurements are:

(i) **Survey research** is the design and use of interviews and questionnaires.

(ii) **Scaling** takes into consideration the key methods of development and implementation of a scale.

(iii) **Qualitative research** provides a general idea about the various types of non-numeric measurement approaches.

(iv) **Unobtrusive measures** offer a number of methods for measuring. These measurement methods have no impact on the purpose of the research.

Levels of measurement

Levels of measurement are concerned with the relation between the values that are assigned to the attributes for a variable. Let us take an example where the 'party affiliation' variable has many attributes. It is assumed that in this specific election context there are three meaningful attributes, these are: 'republican',

NOTES

'democrat' and 'independent'. To analyse the results of this variable conveniently, arbitrary values of 1, 2 and 3 are assigned to these three attributes. The level of measurement determines the relation between these three attributes. Here, numbers substitute text terms. This, however does not mean that higher values indicate 'more' and lower values indicate 'less'. Value is only used as abbreviation for the attribute. The level of measurement can be considered as nominal in this case.

Importance of the level of measurement

The level of measurement helps us decide how to interpret the data from a variable. When we know that a measure is nominal (like the one just described), then we know that the numeric values are just short codes for the longer names. Secondly, knowing the level of measurement helps us to decide the appropriate statistical analysis, on the values that were assigned. On nominal measure, we can never average the data values or do a t-test on the data. There are four broad levels of measurement, they are as follows: nominal, ordinal interval and ratio.

On a nominal scale, the numerical values just 'name' the attribute uniquely. No ordering of the cases is implied. For example, jersey numbers in basketball are measures at the nominal level. A player with number 30 is not more in any way than a player with number 15.

On the ordinal scale, the attributes can be rank-ordered. Here, distances between attributes have no meaning. For example, on a survey you might code educational attainment as 0 = less than H.S.; 1 = some H.S.; 2 = H.S. degree; 3 = some college; 4 = college degree; 5 = post college. In this measure, higher numbers mean more education. But is the distance from 0 to 1 same as 3 to 4? Of course not. The interval between values can not be interpreted in an ordinal measure.

On the interval scale, the distance between attributes *does* have meaning. For example, when we measure temperature (in Fahrenheit), the distance from 30–40 is the same as that from 70–80. The interval between values is interpretable. Because of this, it makes sense to compute an average of an interval variable, whereas it does not make sense to do so for ordinal scales. But note that in interval measurement ratios do not make any sense—80 degrees is not twice as hot as 40 degrees (although the attribute value is twice as large).

Finally, in ratio scale, there is always an absolute zero that is meaningful. This means that you can construct a meaningful fraction (or ratio) with a ratio variable. Weight is a ratio variable. In applied social research most 'count' variables are ratio, for example, the number of clients in past six months. Why? Because one can have zero clients and because it is meaningful to say that '...we had twice as many clients in the past six months as we did in the previous six months.'

In the level of measurement idea it is important to recognize that there is a hierarchy implied. Assumptions tend to be less restrictive and data analyses tend to be less sensitive at lower levels of measurement. At each level up the hierarchy, the current level includes all the qualities of the one below it and adds something new. In general, it is desirable to have a higher level of measurement (e.g., interval or ratio) rather than a lower one (nominal or ordinal).

Qualitative measures

Qualitative research is a vast and complex area of methodology that can easily take up whole textbooks on its own. It is important to consider a number of queries before undertaking qualitative research.

NOTES

For generating new theories or hypotheses

One of the key reasons for carrying out a qualitative research is to gain more experience in the subject of interest. It is frequently seen that in applied social research (particularly economics and psychology) graduating students switch from a literature review on a topic of interest to conducting a research proposal with theories and hypotheses. This research is based on current thinking. The direct experience of the phenomenon is omitted. It is important for every student, who is beginning a research, to execute a deep study into the phenomenon of research. Prior to beginning a multivariate analysis of gender-based differences in wages, one needs to investigate various work contexts and observe how gender is perceived and its effect the distribution of wages. Before recognizing the effects of a new psychotropic drug for the mentally incapacitated, one should visit several mental hospitals to observe what goes on. Once this is accomplished, it will give rise to a fresh perspective which stems out of direct experience. One may possibly devise one's own ideas about the causes. The majority of the new and valuable theories and hypotheses originate from these sources. Obviously, a balance needs to be struck here. Direct experience should be based on qualitative research. It is also important to realize the right time and way to formulate some tentative theories and hypotheses that can undergo explicit tests.

For better understanding of the issues

Qualitative research is specially used to investigate complicated and sensitive issues. For instance, a quantitative research, on the other hand, would help in understanding how people view topics like God and religion, human sexuality, the death penalty, gun control, etc. It would help more in summarizing some main points pertaining to these issues.

To trade detail for generalization

Qualitative research proves to be an excellent tool for providing detailed information. Although there are quantitative studies that offer details of numeric data, there is a difference. In a detailed quantitative research, the data itself gives both, shape and limit to the analysis. For instance, if a simple interval-level quantitative measure is gathered, it will provide fairly delimited analysis (e.g., descriptive statistics, use in correlation, regression or multivariate models, etc.). Typically, attempts to generalize are likely to be rather direct, in most quantitative researches. Since the same variable is gathered from everyone who is part of the sample, generalization gets simpler. It only requires calculation of some aggregate statistics like a mean or median.

However, the process is a bit more complex in a large number of qualitative researches. In this case, the data is more unprocessed and is rarely organized. As a result, one needs to be all set to arrange the unprocessed data. There are

NOTES

innumerable techniques to do this. Still, generalizations based on mock interviews or written documents becomes a difficult task.

In the majority of qualitative researches, elaborate information has both, advantages and drawbacks. The advantages are that it facilitates one to elaborate on the subject of the research in the original language of those who are participating in the research. Actually, a few of the popular 'qualitative' researches are often published in book form, often in a style that almost approaches a narrative story. On the negative side, in such details, it is difficult to determine the generalized themes. In fact, many qualitative researchers do not even care about generalizing —they are content to generate rich descriptions of their phenomena. This is the reason for high values which are a result of mixing qualitative research with quantitative one. Quantitative research proves excellent at converting huge volumes of data into summaries and rendering it more generic, on the basis of statistical projections. Qualitative research again proves to be excellent at giving explanations from the participant's point of view. It offers plentiful and detailed descriptions that transform quantitative results into human context.

Is funding available for this research?

Most researches have to encounter payment issues. There is little information to propose any research that would be difficult to carry out for lack of funds. For qualitative research, this is often an especially challenging issue, because much qualitative research takes an enormous amount of time. It is very labor intensive and yields results that may not be generalized for policy-making or decision-making, many funding sources view it as a 'frill' or as simply too expensive. There is a lot that one can (and should) do in proposing qualitative research that will often enhance its fundability.

Unobtrusive measures

Unobtrusive measures work without intervention of the researcher in the context of the research. On the other hand, for direct and participant observation, physical presence of the researcher is required. This may get the respondents to change their behaviour to seem supportive to the researcher. A questionnaire interrupts the normal behavioural stream. There are possibilities of respondents shying away from filling out a survey questionnaire or resenting the questions that are asked.

It is believed that unobtrusive measurement causes a reduction in the bias that is caused due to the intrusion of the researcher, or measurement instrument. Moreover, unobtrusive measures lessen the researcher's control on the type of data that is collected. A few of the constructs may not have any available unobtrusive measures.

Three types of unobtrusive measurement are discussed here.

Indirect measures

The occurrence of unobtrusive or indirect measure in a research context is natural. The researcher does not establish any formal measurement procedure to collect

NOTES

data. The creativity and thought processes of the researcher restrict the types of indirect measures that are possibly available, for example, let us consider the measurement of the popularity of different artifacts in a museum. One can establish a system of measuring that is mechanical and hidden from the clientele of the museum. In a research that was conducted, the technique was plain. In this research, new tiles were installed in front of each exhibit. The subject of measurement was the time affecting the wear and tear of the tiles. This feature was an indirect indicator of the number of patrons and their interest. An electronic system of measurement can bring about a major degree of improvement to this approach. Another way is by placing concealed scanners, the recordings of which may be used to translate the patron's level of interest.

In a research that was carried out on the listening preferences of radio audiences showed the presence of favorite indirect measures. Instead of carrying out an obtrusive survey or interview pertaining to favorite radio stations, the researchers approached regional auto traders and garages. They examined the tuning of the radios in all cars that were brought for servicing. Similarly, if one needed to find out about magazine preferences, one may go through discarded magazines or carry out a door-to-door magazine recycling effort.

This instances exhibit the most important features about indirect measures—one needs to exercise care in following the ethical method for this type of measurement. As the name sounds, indirect measure involves the collection of information without letting the respondent know about it. There is a possibility of their right to privacy being violated in the process. However, some parts of the information that is collected may not be private or not an intrusion of their privacy.

Sometimes an indirect measure is more suitable, easy to obtain and follows ethics. Just like all other measurements, these measures should be readable and valid, for example, radio station preferences are collected at different times and their results are correlated. These results might be useful in knowing how reliable the tests are. Another way is to combine indirect measures with other direct measures of the same construct (supposedly in a pilot study) to help establish construct validity.

Check Your Progress

6. What is a survey?
7. Into which two categories can a survey be divided?
8. Which is a less familiar type of questionnaire?
9. What does data preparation involve?
10. What does historical research consist of?

NOTES

2.9 SUMMARY

- A research process can be divided into the following major categories: specification/measurement and manipulation of theoretical variables, selection of research design, analysis of data obtained from study and interpretation and robustness of the findings.
- A research is made up of the following two syllables: 're' and 'search', where 're' means again and search means to examine, scrutinize or investigate deeply. It is an organized way of asking questions.
- In a research, variables are both, important and deceptive.
- An independent variable can be manipulated by the researcher. A dependent variable is that which is dependent on the independent variable and its changes. Certain other variables which are extra, have a tendency to offer explanations or create doubts on the conclusions.
- The entire research revolves around the process of writing a hypothesis. Hypothesis is a summary or conclusion of a research.
- Causation is a chain of events wherein, one event acts as a cause to create another.
- Survey research is one of the most essential area of measurement of applied research.
- Content analysis is a tool which is used to establish the presence of certain concepts or words in a research.
- The historical method of carrying out a research is done by the analysis of historical development and origin of the topic of research.
- Measurements are the means to quantify a research and its components.

2.10 KEY TERMS

- **Generalization:** A general statement that is based on only a few facts or examples
- **Variable:** An entity which has the capability to keep on changing
- **Independent variable:** A variable that is controlled and manipulated by the experimenter
- **Dependent variable:** A variable that changes according to the independent variable
- **Extraneous variables:** Type of variables that may have an impact on the relationship between independent and dependent variables
- **Participant variables:** Variables that are related to the individual characteristics of each participant
- **Situational variables:** Variables that are related to factors in the environment that may impact the manner in which each participant responds

- **Causation:** The process of one event causing or producing another event
- **Actus reus:** A latin term that means the deed of crime
- **Mens rea:** A latin term that means the intention to commit a crime
- **Questionnaires:** A written list of questions that are answered by a number of people for the purpose of collecting information
- **Bereavement:** The state of having lost a relative or close friend

NOTES

2.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The word 'research' is a combination of the words, 're' and 'search'. This implies searching again and again to thoroughly understand the issue and find apt or effective solution or strategy to deal with the issue.
2. The Webster's international dictionary gives a very inclusive definition of research, as a 'careful, critical inquiry or examination in seeking facts or principles, diligent investigation in order to ascertain something'.
3. A variable is usually anything that is changeable.
4. In a psychological experiment, the variable that is controlled and manipulated by the experimenter is known as an independent variable.
5. A variable that changes according to the independent variable is known as a dependent variable.
6. A survey can range from a short paper-and-pencil feedback form to an intensive one-on-one in-depth interview.
7. A survey can be divided into the following two categories: questionnaire and interview.
8. A less familiar type of questionnaire is the household drop-off survey.
9. Data preparation involves checking or logging the data in, checking the data for accuracy, entering the data into the computer; transforming the data and developing and documenting a database structure that integrates various measures.
10. A historical research consists of content analysis and statistical evaluation of the data that was originally collected for some other purpose.

2.12 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Why is research a methodological process?
2. How did John Best define research?
3. What are extraneous and confounding variables?

NOTES

4. What are the assumptions of a hypothesis?
5. How has survey research changed in the last ten years?

Long-Answer Questions

1. Enumerate the basic elements of research.
2. Explain the concept and definition of research.
3. Write a note on hypothesis and causations.
4. Describe the significance of a survey research.
5. What role does measurement play in a research?

2.13 FURTHER READING

Boyer, Carl. 1959. *The History of Calculus and its Conceptual Development.* New York: Dover Publications Inc.

Hempel, C. G. 1952. *Fundamentals of Concept Formation in Empirical Science.* Chicago: The University of Chicago Press.

Stuart, Mill, John. 2009. *A System of Logic.* Middlesex: The Echo Library.

UNIT 3 SAMPLING IN COMMUNICATION RESEARCH

NOTES

Structure

3.0 Introduction
3.1 Unit Objectives
3.2 Sampling in Communication Research
  3.2.1 Types of Sampling
  3.2.2 Application and Limitations of Sampling
3.3 Methods of Data Collection
  3.3.1 Interviews, Questionnaires and Observation
3.4 Case Study Application
3.5 Limitations of Different Sampling Methods
3.6 Summary
3.7 Key Terms
3.8 Answers to 'Check Your Progress'
3.9 Questions and Exercises
3.10 Further Reading

3.0 INTRODUCTION

The process of sampling involves random signaling of research subjects or topics and coming out with reports on their nature and quality. Besides communication, the method of sampling has been applied to a large number of problems in medicine, social sciences, etc. This unit presents the importance of sampling in communication research, types of sampling, methods of data collection, etc.

3.1 UNIT OBJECTIVES

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

- Explain the role of sampling in communication research
- Discuss the various types of sampling
- List the applications and limitations of sampling
- Analyse the methods of data collection
- Describe the significance of interview, questionnaire and observation
- Explain sampling with the help of a case study application
- Give an overview of the limitations of different methods of sampling

3.2 SAMPLING IN COMMUNICATION RESEARCH

NOTES

The activity of the sample of collection cannot be observed on the basis of the type of the hypotheses. A sample is drawn and studied by the researcher, from larger group in question. The outcome of the research is shared with the participants of the larger collectivity. To be more specific, if it is required to know the reaction of senior citizens towards the planned changes in income tax benefits, it is almost impractical to ask every senior citizen for their view on the new policy. Generally, researchers are not able to access the entire population for a research. Hence, they usually depend on samples which are likely to represent the total population. The sample should represent the total population in a fair way. Researchers use different statistics to identify sampling margins of error which they also refer to as confidence intervals.

The two fundamental techniques used for sampling are: probability sampling and non probability sampling. In a probability sampling, every factor of the general population has an equal or known probability of being part of the sample. On the other hand, if non probability sampling is practiced, the chances of every factor of the general population being present in the sample, cannot be determined. For this reason, probability sampling is more preferred by researchers and they are critical towards the flaws of non probability sampling. There are several circumstances where it is not possible to use the techniques of probability sampling. In such cases, non probability sampling is the selected technique of sampling that is used, though it may not be a technique that is preferred.

3.2.1 Types of Sampling

(i) Probability sampling

Following are the sub categories of samples which result from probability sampling:

a. Simple random sample

A simple random sample is a kind of sample that is gathered in a way such that there is equal probability of the presence of every factor of the population, in the sample. Every factor of the population is listed and given an independent number. The sample is selected on the basis of a table that generates random digits. Occasionally, instead of the random number table, the 'Monte Carlo' method is adopted. In this method, all the elements are shuffled and drawn at random until the required elements are selected.

Researchers frequently assert the use of a 'simple random sample'; nevertheless, they are rarely stringent in using samples of this type. This reason for this can be attributed to the fact that it is rarely possible to identify every element within a total population, for instance the possibility of identifying all residents of a nation, state, or even a community is very low. The second reason for not using a simple random sample is the possibility of failure (such as failure to complete a

questionnaire, subjects drop out, etc.). No response usually produces a biased or non-random sample in the final analysis – although the researcher had intended to use a probability method.

b. Stratified random sample

NOTES

This method of sampling divides the population into multiple groups or subgroups. These groups or subgroups, which are also known as strata, have elements that are required by researcher to make sure that the population is represented adequately. If it was desired to represent male and female perspectives, the population listing (or sampling frame) would be divided into two subgroups. This would be followed by selecting a simple random sample from every subgroup, with the help of a table of random numbers or the 'Monte Carlo' method. It is possible to divide sampling frames according to one or more social features like, gender, educational attainment, religious affiliation, or age.

Each stratum has elements that have been selected by the use of simple random method. Researchers can have an equal or proportionate number of individuals in each subgroup. Let us take an instance where 60 per cent of the students at a given university are females and the researcher desires to take a sample of 100 students by employing the proportionate stratified sample method. For this, the researcher would, at random, choose 60 female students and 40 male students. For the disproportionate stratified sample, 50 males and females would be selected, at random, from the stratified sampling frame.

c. Cluster sampling

Researchers have developed the cluster sample as a means of alleviating some of the costs of time and money. In this type of probability sample, the total population is identified as having clusters of elements. These clusters are then randomly selected via either simple random, or stratified random techniques.

For example, if you were asked to sample a population of random households in the City of Chennai, you could list all possible households and randomly select each of them, on the basis of specific cases. You would then drive to the designated household and conduct the interview. But because these households are likely to be widely scattered and because it would be a tedious and boring task to use census maps to identify, list and randomly select households, a cluster sample would be a more efficient means of collecting representative information.

In the cluster sample, one would create a sampling frame by identifying all neighborhoods or residential areas and then take a random sample of these clusters. These clusters might be randomly selected — hence a type of simple random selection. This depends on some characteristic that are deemed to be important, then a stratified random technique would be employed.

NOTES

Drawbacks of probability samples

a. Probability samples are expensive.

b. Non-response is a typical problem.

c. If one cannot list the elements in the population, they are impossible.

d. Probability samples are much more time consuming.

(ii) Non probability sampling

Non probability samples are not expensive to conduct and employ. However, the non probability sample is not generally considered to be very valid. There are several types of non probability samples, some of them are listed below:

a. Accidental samples

These samples comprise of those units, the cases of which are easily obtained. When an accidental sample is created (also known as an availability sample), the desired size of the sample is defined by the researcher. Data is collected on the basis of a number of individuals who are to be part of the study.

A large number of researches are conducted by using accidental sampling, For instance, the data obtained from 350 students in an introductory psychology class is a type of accidental sample. Media programs on television and radio also use accidental samples, when they make telephonic calls to know the views of individuals. It is important to know that the size of the sample does not indicate whether it is a probability or non probability sample.

One key drawback of accidental sampling is that elements which are selected may not be ideal representations of the larger population.

b. Quota samples

Quota sample is an effort to estimate stratified random sampling technique in a non-random manner. The researcher begins by determining the factors, which according to him represent the population. This is followed by determining a sample size for each category. Lastly, individuals are selected on the basis of their availability. Here, we can visualize an instance where equal numbers of women and men are interviewed, to know how they viewed municipal laws that govern wages for jobs. By means of utilizing a quota sample, willing and available individuals may be interviewed. This process continues till the desired number of individuals in each subgroup, have been interviewed.

c. Purposive samples

Purposive samples are at times referred to as judgment samples. Researchers use these samples to determine the cluster sample by using a non probability sample. A typical group of individuals is selected by the researcher. This group may reflect the larger population. Let us visualize a researcher who desires to survey the attitudes of freshman college students at a particular university. He is likely to select the students from one or more freshman English classes. This may be due to

the assumption that since all students are sure to take freshman English classes, those in any class represent the entire freshmen class.

The judgment sample is also likely to consider an individual (instead of a group) as a sampling unit. By considering an individual as a sampling unit, the researcher subjectively defines a 'typical' case and then tries to select those individuals who ideally represent this definition. However, there is no way of guaranteeing that the researcher has been able to identify the most representative cases, neither can his or her definitions be considered as typical or accurate. The purposive sample exhibits a clear non probability approach and it is likely that the sample is biased since the selection process is not random.

NOTES

d. Snowball sampling

In this method of sampling, available respondents to be included in the sample are selected by the researcher. Once the subject to be researched has been reviewed, the researcher attempts to identify other individuals, who would be part of the sample. For example, in a research that is carried out on rich people in Chicago, it is likely that the list of millionaire Chicago residents is not adequate. Many residents have been left out. The list may start with the rich people that are known to the researcher. The researcher may interview them, and they can possibly refer other wealthy people to him. Thus this snowball referral method may provide a sample of the desired size. However, it is doubtful for this sample to be unbiased and a perfect representation. Actually speaking, often more than one sampling strategy is used by researchers for a given project.

3.2.2 Application and Limitations of Sampling

One cannot underestimate the importance of sampling for the success of a research. Even though the concepts are successfully executed, the interviewers carefully trained, the statistical measures applied and interpreted correctly; an unrepresentative and biased sample can render the results of any study invalid.

Limitations of Sampling - Sometimes the information about each and every unit of the population is required. This is possible only through the complete enumeration because the sample will not serve the purpose. The following are some examples where sampling is not allowed:

- To conduct elections, we need a complete list of voters. The candidates participating in the election will not accept the results prepared from a sample. With increase in literacy, the people may become statistical minded and they may be willing to accept the results prepared from a sample. In advanced countries, opinion polls are frequently conducted and people accept the results of sample survey unofficially.
- Tax is collected from all tax payers. A complete list of all the tax payers is required. The telephone, gas and electricity bills are sent to all the consumers. A complete list of the owners of land and property is always prepared to maintain records. The position of stock in factories requires complete entries of all the items in the stock.

NOTES

However, probability samples have the following drawbacks:

- They are expensive
- Non-response is a special problem
- If one cannot list the elements in a population, they are impossible
- They are very time consuming

CHECK YOUR PROGRESS

1. What are confidence intervals?
2. Name the two types of sampling techniques.
3. Define a simple random sample.

3.3 METHODS OF DATA COLLECTION

A key activity in any research study is the collection of data. Data which is not accurate can affect the outcome of a research and finally exhibit false results.

On a given scale, data collection methods for impact evaluation tend to vary across a range. Quantitative and qualitative methods are located at either ends of the range.

Quantitative data collection methods are dependent on random sampling and structured data collection instruments. These translate different experiences into response categories that have been determined earlier. They give results that can be easily summarized, compared and generalized.

Quantitative research is inclined towards testing hypotheses which arises as a result of theory and/or being able to estimate the size of a phenomenon of interest. On the basis of the subject to be researched, the participants may be given different treatments at random. If this is not practical, the researcher may collect data on participant and situational characteristics. If it is intended to generalize from research participants to a larger population, probability sampling is used to select participants.

Typical quantitative data gathering strategies include:

- Experiments/clinical trials
- Observing and recording well-defined events (e.g., counting the number of patients waiting in emergency at specified times of the day)
- Obtaining relevant data from management information systems
- Administering surveys with closed-ended questions (e.g., face-to face and telephone interviews, questionnaires etc). (http://www.achrn.org/quantitative_methods.htm)

3.3.1 Interviews, Questionnaires and Observation

NOTES

Interviews

In quantitative research (survey research), interviews are more structured than in those in qualitative research. A structured interview comprises of researcher who asks a typical set of questions. (Leedy and Ormrod, 2001)

Face-to-face interviews enable the researcher to establish a rapport with potential participants and gain their cooperation. In a research, these interviews exhibit the highest number of responses. They also facilitate the researcher in clarifying unclear responses and assist in follow-up activity. These interviews are not feasible in case of large samples as they tend to prove expensive and time-consuming. (Leedy and Ormrod, 2001)

Telephonic interviews consume less time and are cost effective. Anyone who has a telephone can be easily accessed. The drawback of this interview is its low response rate. However, the response rate is much higher than that of a mailed questionnaire. This method may prove biased in the sense that the researcher cannot conduct this survey on people who do not have telephones.

Computer-assisted personal interviewing is in the form of an electronic questionnaire. The interview is conducted on a laptop or hand-held computer. The advantages of this system are that it saves time and does not require the interviewer to carry paper questionnaires. Nevertheless, this system is expensive to establish and requires interviewers with computer and typing skills.

The commonly used qualitative methods can be divided into the following three categories: in-depth interview, observation methods and document review.

In-depth interview

In the in-depth interviewing method (unstructured interviewing), researchers elicit information in order to achieve a holistic understanding of the interviewee's point of view or situation. This method is also capable to identify interesting areas for further analysis. In this type of interview informants are asked open-ended questions. Wherever necessary, they are probed, to obtain the data desired by the researcher. In-depth interviewing is also referred to as qualitative interviewing because it comprises of qualitative data. Patton (1987:113) recommended the following fundamental approaches for conducting qualitative interviewing:

(i) The informal conversational interview

This interview is similar to a chat conversation. In this process, informants often do not tend to remember that they are undergoing an interview. Major part of the questions is related to the immediate context. Informal conversational interviews prove to be valuable for investigating interesting subjects of investigation

NOTES

(ii) The general interview guide approach (commonly called guided interview)

In this method, a basic checklist ensures that none of the relevant topics has been missed out. However, the interviewer has the advantage to explore, probe and ask questions that are of concern to the researcher. This method proves useful to extract information pertaining to specific topics. Wenden (1982) established a checklist to form the base for interviewing her informants, in a research that led her PhD studies. She recommended this method useful as it provides the freedom of probing, as part of the research, without the researcher having to cross the boundaries of the aim of the research.

(iii) The standardized open-ended interview

Researchers using this approach prepare a set of open-ended questions which are carefully worded and arranged for the purpose of minimizing variation in the questions posed to the interviewees. Due to this advantage, this is the most preferred method to collect interviewing data and it involves two or more researchers. Though this technique is not flexible, it still makes probing possible, depending on the nature of the interview and the skills of the interviewers (Patton 1987:112).

Questionnaires

Paper-pencil questionnaires are capable of focusing on and hence can cover large populations. They are cost-effective as well as they take less time. The opinions obtained through such questionnaires, even on issues that spark controversies, are more genuine. This is because their responses bear anonymity. However, this method too has its own disadvantages. Majority of the people who receive questionnaires don't return them and those who do might not be representative of the originally selected sample. (Leedy and Ormrod, 2001)

Web based questionnaires is the latest technique which is rapidly developing. It uses the Internet for its research. In this method, an email is sent to the respondent. This email has the address of a website which, when clicked, opens a secure website to fill in a questionnaire. This technique of research is less detailed and saves time. A drawback of this technique is that it only includes those respondents who have access to computers. Another drawback is the possibility of the responses not being accurate, due to the hurry with which they are completed.

Questionnaires often make use of checklist and rating scales. These devices help simplify and quantify people's behaviors and attitudes. A checklist is a list of behaviors, characteristics, or other entities that the researcher is looking for. Either the researcher or survey participant simply checks whether each item on the list is observed, present or true or vice versa. A rating scale is more useful when a behavior needs to be evaluated on a continuum. They are also known as likert scales. (Leedy and Ormrod, 2001)

Qualitative data collection methods play an important role in impact evaluation by providing information that is useful to understand the processes of

NOTES

observed results and assess changes in people's perceptions of their well-being. Furthermore, qualitative methods can be used to improve the quality of survey-based quantitative evaluations, by helping generate evaluation hypothesis; strengthening the design of survey questionnaires and expanding or clarifying quantitative evaluation findings. These methods are characterized by the following attributes:

- They tend to be open-ended and have less structured protocols (i.e., researchers may change the strategy of data collection by adding, refining, or dropping techniques or informants)
- They rely more heavily on interactive interviews; respondents may be interviewed several times to follow up on a particular issue, clarify concepts or check the reliability of data
- They use triangulation to increase the credibility of their findings (i.e., researchers rely on multiple data collection methods to check the authenticity of their results)
- Generally, their findings are not generalized to any specific population, rather each case study produces a single piece of evidence that can be used to seek general patterns among different studies of the same issue

Regardless of the kind of data involved, data collection in a qualitative study takes a great deal of time. The researcher needs to record any potentially useful data thoroughly, accurately and systematically, using field notes, sketches, audiotapes, photographs and other suitable means. The data collection methods must observe the ethical principles of research.

Questioning techniques

All individuals are not equally capable to express their thinking and ideas. Researchers who are more articulate with their questions are able to extract the maximum quality information from respondents. The latest techniques of questioning may be listed as follows:

(i) Clarity in questions

Cicourel (1964) reflects that 'many of the meanings which are clear to one will be relatively opaque to the other, even when the intention is genuine communication.' Hence, the language used for communicating with respondents should be simple for them to understand and comprehend. To enhance their comprehensibility to the interviewees, questions should be easy to understand, short, and devoid of jargon (Kvale 1996).

(ii) One question at a time

Interviewers should not bombard the interviewee with many questions at the same time. Questions should be put forward, one at a time. This will help the respondent concentrate on single questions without getting confused or fluttered.

NOTES

(iii) Ask truly open-ended questions

Truly open-ended questions call for answers which provide the interviewees with the freedom to give their independent views without being affected by other factors. Ideal instances of such questions are: 'What do you think about your English?' 'How do you feel about the method of teaching English in your home country?' 'What is your opinion of English lessons in the UK?'

(iv) Questions pertaining to opinion/feeling questions should follow those pertaining to experience/behaviour

It is always important to begin with questions which get the interviewee to talk about his experience or behaviour, before asking him about his views or ideas. This helps in streamlining the process of getting the respondents to express both clearly with a correlation between the two. An ideal instance of this is asking, 'What happened?' followed by, 'How do you feel now?'

(v) Sequence the questions

Funnelling is a specialized method of questioning. The questions in this technique may be generic, precise, extensive or restricted. Cohen and Manion cited an instance from a research which was conducted by Sears, Maccoby and Levin: All babies cry, of course. Some mothers feel that if you pick up a baby every time it cries, you will spoil it. Others think you should never let a baby cry for very long. How do you feel about this? What did you do about it? How about the middle of the night? (Sears, Maccoby and Levin, 1957, cited in Cohen & Manion 1994:277)

(vi) Probe and follow-up questions

A probe is done to get more in-depth answers to questions. It is inclined to enhance the quality of data which is collected. It also enables the respondent to know the level of details that are required. Direct questions prove to be the easiest way of probing, for instance, questions like, 'Could you say something more about that?', 'Can you give a more detailed description of what happened?, 'Do you have further examples of this?', etc. Body language exhibited by the interviewer, that may be a simple nod or a verbal 'hmm', hints the interviewee to elaborate on his answer. Repeating significant words of an answer can lead to further elaboration (Kvale 1996:).

(vii) Interpret questions

During the entire round of the interview, the researchers should prompt the interviewee to elaborate on his answers and where necessary, the researcher should himself elucidate the statements for avoiding any misinterpretations. Kvale (1996) recommends using questions like, 'Is it correct that you feel that.....?'; 'Does the expression..... cover what you have just expressed?'. This facilitates the interviewees to verify or refute what the researchers have interpreted.

NOTES

(viii) Avoid sensitive questions

Researchers should not ask sensitive questions that might provoke the informants. Such questions are likely to hinder the process of interview. Cicourel (1964) holds the opinion that 'the respondent may well feel uneasy and adopt avoidance tactics if the questioning is too deep.'

(ix) Encourage a free rein but maintain control

The researchers should allow the interviewees to be adventurous and touch the topics which are adjacent to the research. However, there should be a tentative list of ideas or areas which may be helpful for the interviewee to know. Palmer (1928:) recommends that skillful interviewers must be capable to direct the course of an interview to benefit the interest of the research.

(x) Establish rapport

A rapport can be established by having a sense of respect for the views of the interviewee, being supportive towards their feelings and giving recognition to their statements. The researcher can exhibit these feelings by the tone of his voice and even his body language. Additionally, Kvale (1996) suggests that 'a good contact is established by attentive listening, with the interviewer showing interest, understanding, and respect for what the subjects say.' An ideal interview allows the interviewees to speak all that they desire, without interruptions. It gives them the freedom to speak and think at their own pace.

Observation method in sample collection

Observational research techniques draw the sole involvement of researchers in the process of observation. An observation method has many advantages. Typically, the observations display a flexibility where they are not bound by a hypothesis (a hypothesis is an idea that expresses what is expected in an observation). For instance, a researcher may carry out observations prior to conducting a more structured research. These observations would be helpful in preparing questions. This is called descriptive research. The outcome of an observational research is considered highly valid. Trochim states that validity is the best available approximation to the truth of a given proposition, inference, or conclusion. The results of an observational research are rated high in terms of validity due to the ability of the researcher to collect detailed information pertaining to a specific mode of behavior. However, there are disadvantages in the form of problems associated with reliability and generalization. Reliability defines the degree of simulation, an observation can have. Observing repeated behaviour may consume a lot of time. Trochim defines the ability to generalize, or external validity, as the extent that the study's findings would also be true for other people, in other places and at other times. In an observational research, the outcome may be a result of the focus on a specific section of the population. This type of an outcome cannot be regarded as general. A researcher's prejudiced approach may also pose a problem. A frequent assumption is that the researcher only focuses on what he

NOTES

prefers. This problem of prejudice can be tackled by coaching and monitoring or using electronic means to record the inferences. Thus, on a broad level, observations prove to be of value to researchers.

Types of observations

(i) Direct (reactive) observation

Direct observations allow the people to know that they are being watched. The sensitive factor is their reaction. A common concern here is the possibility of individuals resorting to opinions and views that are not genuine. They may exhibit fabricated views. These artificial behaviour tactics may result out of their preference to be socially desired or as an attempt to uphold their privacy in relationships. However, unnatural behaviour cannot be continued for a long time. Any observation that extends over a long period can be capable to spot natural tendencies. Another question that arises, points to the possibility of the findings being generic. It is likely that the sample of individuals is more specific to individuals, instead of representing the entire population. It is also possible that this sample may not represent the actual behaviour of individuals either (a person may have been in a foul mood on the day of recording the observation). However, observational studies that span across lengthy periods usually tend to counter issues related to external validity

Two popular methods of direct observations are:

a. **Continuous monitoring** is the activity where a subject or subjects are observed and their behaviour and activities are recorded (either manually, electronically, or both). Continuous monitoring is a popular tool that is used by business groups to evaluate the performance of their workforce. However, it may have problems due to the Hawthorne Effect. The Hawthorne Effect proposes the theory which says that the reaction of workers to the attention they get from the researchers causes a direct increase in their productivity. This reaction should be known to observers. Continuous monitoring is also important in the field of academics, where it is used to observe the interaction between teachers and students. It is also important in dietary and nutrition, where it helps researchers in recording an individual's food intake. Though continuous monitoring is comparatively a simple process, but it consumes a lot of time. It is a sure shot means to get ample amount of information.

b. **Time allocation** is the activity in which a venue and a timeslot are fixed to record the activities of individuals, who are not aware that their activities are being recorded. It is widely used to determine what percentage of the time is devoted by people towards specific activities (i.e. playing with their kids, working, eating, etc.). Following are the disadvantages of this approach:

   i A larger sample is required to find a general representation of the time spent by people in various activities.

ii Probing into peoples' private lives may result in unkind reaction.

iii Additionally, questions like when, how often and where should the observation focus are the key areas of concern.

NOTES

A large number of researchers have resolved these problems by means of random visits to planned locations at different times.

(ii) Unobtrusive observation

All techniques used for observation that comprise of the study of behaviour of individuals without their knowledge, fall in the category of unobtrusive measures. In this method, there is no risk of the observer changing the behaviour of the subject. One should address the problems pertaining to validity when carrying out unobtrusive observations. A large number of observations are required to be carried out on a particular representative sample, for the outcome to be more generic. This is not easy while a particular group is the focus of the observation. Several groups exhibit features that are exclusive to them and these features increase the level of interest in them for the research. Therefore, these findings are not very valid externally. In addition to this, it is not an easy task to replicate them by resorting to techniques that are not conventional (non-conventional meaning unobtrusive observation). It becomes practically even more complicated to reproduce observations of specific behaviors in researches, in case the researcher is a group participant. Unobtrusive methods may fail on the platform of ethics. Concerns pertaining to informed consent and assault on privacy are most dominant here. An institutional review board is likely to react strongly towards such a research if the subjects have not been taken into confidence for it.

(iii) Behavior trace studies

As the name suggests, behavior trace studies comprise of the activities of finding things that have been left behind by people and working on their interpretations. These studies may revolve around topics like vandalism or even garbage. One of the most popular trace-study was the University of Arizona Garbage Project. To study features like choice of food, habits of waste disposal, liquor intake, etc., of the people of Arizona, anthropologists and students went through the domestic garbage from houses. It needs to be kept in mind that the highlight of an unobtrusive research is that the individuals are not aware that they are the object of study. Astonishingly, the inhabitants of Tucson (the area of research) were supportive towards the study, provided their identities were not revealed. It can be easily visualized that trace studies are likely to give out large quantities of data.

(iv) Disguised field observations

In disguised field analysis the researcher gives an impression, or actually becomes a member of a group for recording data pertaining to that group. The group is not aware of the fact that it is being monitored for a research. In this case, the observer probably plays different roles. The initial role of the observer is to be active in participating in all activities of the group. Let us understand this better with an

NOTES

example; If the researcher is a team member of a sorority and is involved in studying female conflicts within sororities, he/she would be regarded as complete-participant observer. However, the researcher may choose to keep his participation informal, by limiting it to the activity of collecting observations. Here, the only reason for the researcher to get in touch with members of the group is to get to know them. In this case, the researcher is an observer-participant. If the researcher identifies himself with other members of the group without participating in its core procedures, he is seen as a participant-observer. An instance of this type of behaviour can be exhibited when one joins a cult but abstains from engaging in any of its important customs. Nevertheless, one is still recognized as a part of the cult and has the trust of all other members. In ethical terms, the majority of problems are endured by participant-observers. Two of the important factors that need more attention in this case are: sensitivity of the subject of researched and the extent of secrecy that is to be maintained. To sum it up, studies that are carried out in the field and are camouflaged and tend to give a genuine output. However, ethical dilemmas continue to exist.

CHECK YOUR PROGRESS

4. How can inaccurate data damage a research?
5. What kind of results do quantitative data collection methods produce?
6. What is computer-assisted personal interviewing?

3.4 CASE STUDY APPLICATION

The history of a case study research is alternated by timeslots of intense utilization and those where the utilization does not exist. The most initial type of this research was conducted in Europe and extended to France. This research captured all issues related to immigrants. These issues comprised of poverty, unemployment, etc. and were ideal to conduct a case study methodology. Zonabend (1992) suggested that a case study should focus on complete observation, reconstruction and analysis of the concerned cases. He also held that a case study should reflect the opinions of the people who form part of the case which is being studied.

Sociology is that discipline which is very closely linked to case study research. The era around the year 1935, witnessed many issues that were highlighted by researchers in other fields. This era corresponded to a revolution in the discipline of sociology, by causing it to incline scientifically. This was done by measuring research design analysis quantitatively. Hamel (Hamel et al., 1993) cautiously dismissed the disapproval of case study, regarding them as inferior in their findings. He believed that the disadvantages of case study had not been tackled in the right manner. On the other hand, sociology had been projected as an immature discipline.

NOTES

The advancement of quantitative methods resulted in a quicker fall in the case study. However, the 1960s saw researchers getting worried over the short comings of quantitative methods. This rejuvenated their preference for case study. Strauss and Glaser (1967) established a theory of 'grounded theory.' Accompanied by other recognized forms of research, this enhanced the use of case studies.

Continuous criticisms of case studies are causing them to lack capability in reaching more generalized inferences. Yin (1993) held Giddens' view, which believed case methodology to be microscopic since it did not have enough cases. Hamel (Hamel et al., 1993) and Yin (1984, 1989a, 1989b, 1993, 1994) strongly disputed that the relative size of the sample (whether 2, 10, or 100 cases are used) is not able to convert a multiple case to a macroscopic study. The aim of the study should be to ascertain parameters which should then be applicable to all researches. This would increase the level of acceptability for every case, if the set objectives are met.

Designing case studies

Yin (1994) identified the importance of the following five components of research design for case studies:

(i) A study's questions

(ii) Its propositions, if any

(iii) Its unit(s) of analysis

(iv) The logic linking the data to the propositions

(v) The criteria for interpreting the findings

The most common questions of a study begin with 'how' and 'why'. Defining them is the initial task that is to be taken up by the researcher. These questions are important in concentrating on the goal of the study. It is not important for every study to have propositions. The success of an exploratory study is based on a stated purpose or criteria, rather than on propositions. The unit of analysis defines the case. This primary unit of analysis may comprise of groups, organizations or countries. The most under developed characteristics of case studies link data to propositions and the criteria for interpreting the findings (Yin, 1994).

Stake (1995) and Yin (1994) defined a minimum of six sources of evidence in case studies. These six sources reflect the researches of both, Yin (1994) and Stake (1995), they are: documents, archival records, interviews, direct observation, participant-observation, physical artifacts.

Documents may comprise of letters, memoranda, agendas, administrative documents, newspaper articles, or any document that may fulfill the purpose of the study. To support the research, the documents confirm the evidence that is obtained from other sources. Documents also help in reaching conclusive decisions pertaining to events. Documentation may go haywire, when handled by researchers who lack experience, thus attracting strong criticism. Documents are the means of

NOTES

communication between participants of the study. If the observations of the researcher are vivid, they will help in preventing the investigator from being misled.

Documents that are archived may comprise of service records, organizational records, lists of names, survey data, etc. The investigator needs to exercise care in validating the integrity of the records before putting them to use. Although they may be quantitative, they may still be inaccurate.

Interviews are one of the most important sources of case study information. The following are some of the forms of interviews: Open-ended, focused and structured or survey. An open-ended interview is where key respondents are called upon for commenting about certain events. This helps in rendering solutions or providing clarity about events. It may also focus on evidence that is achieved from other sources. The researcher should not depend only on one informant or one source, for data. It is important for him to check other sources for the same data. This will serve as a verification to check the reliability of the source.

A focused interview is conducted when the duration is short and the interviewee is given questions that are designed to suit the time frame. Focused interviews are generally conducted to confirm the information from other sources. This interview is like a survey and can be used in the same way, to collect data in activities like neighborhood studies. Similar to a survey, the questions are more subjective and developed in advance.

Direct observation is required when a field visit is carried out in a case study. It may be an easy process of collecting casual data, or formal protocol for measuring and recording behaviors. This technique is important as it provides more information pertaining to the topic that is being studied. It is even more reliable when the number of observers, involved in the task, is more than one. Glesne and Peshkin (1992) recommended that researchers should be as unobtrusive as the wallpaper.

Participant-observation converts a researcher into an active participant in events that are being researched. This is a common phenomenon in studies of neighborhoods or groups. The technique offers a few extraordinary means of gathering data, but it also has its disadvantages. It is possible for the researcher to change the course of events even though he is within the group. However, this may prove detrimental to the study.

Physical artifact is a collective term used for tools, instruments, or other physical evidence that may be collected during the study, as part of a field visit. A discovery can widen the perspective of the researcher. It should be noted that all sources do not serve importance for every case study It is important to keep in mind that not all sources are relevant for all case studies (Yin, 1994). The investigator should have the capability to deal with all of them, as required. However, every case presents varying opportunities for collecting data.

There are some conditions that arise when a case researcher must start data collection before the study questions have been defined and finalized (Yin,

1994). The chances of success in this case depend on the skill of the investigator. Another factor to be considered here is the advantage of putting rival hypotheses into practice and working on concepts that enhance the quality of the case study. This provides support to a fair approach and innovative thinking of the researcher.

NOTES

3.5 LIMITATIONS OF DIFFERENT SAMPLING METHODS

The Table 3.1 gives limitations of different methods of sampling:

***Table 3.1** Sampling Limitations*

Technique	Application/Advantage	Disadvantage
Simple random	Exceedingly descriptive when it involves the participation of all subjects; the ideal	Not feasible without a complete list of population members; Expensive to administer; may result in disorder when members are required to be isolated from a group; the duration involved may be too long, possibility of the data/sample undergoing a change
Stratified random	Can ensure that specific groups are represented, even proportionally, in the sample(s) (e.g., by gender), by selecting individuals from strata list	More complex, requires greater effort than simple random; strata must be carefully defined
Cluster	Possible to select randomly when no single list of population members exists, but local lists do; data collected on groups may avoid introduction of confounding by isolating members	Clusters in a level must be equivalent and some natural ones are not for essential characteristics (e.g., geographic: numbers equal, but unemployment rates differ)
Stage	Combination of cluster (randomly selecting clusters) and random or stratified random Sampling of individuals	Complex, combines limitations of cluster and stratified random Sampling
Purposive	Ensures balance of group sizes when multiple groups are to be selected	Samples are not easily defensible as being representative of populations due to potential subjectivity of researcher
Quota	Ensures selection of adequate numbers of subjects with appropriate characteristics	Not possible to prove that the sample is representative of designated population
Snowball	Possible to include members of groups where no lists or identifiable clusters even exist (e.g., drug abusers, criminals)	No way of knowing whether the sample is representative of the population
Volunteer, accidental, convenience	Inexpensive way of ensuring sufficient numbers of a study	Can be highly unrepresentative

CHECK YOUR PROGRESS

7. What marks the history of case study research?
8. How many components of research design were identified by Yin, as important for case studies?
9. When does direct observation occur?

NOTES

3.6 SUMMARY

- Sampling as a process is chosen in communication research because it is too expensive and time-consuming to conduct a research on the entire population of people.
- The main sampling methods are: probability sampling and non probability sampling.
- Sometimes there is much variation in each and every unit of the population. In such cases, sampling does not help because sampling takes the most obvious units of a population. For sampling, a certain degree of uniformity within the population is important.
- There are two primary methods of data collection: qualitative and quantitative methods.
- Interviews are the process of data collection where the participants provide specific, and detailed descriptive information, pertaining to the topic of interview. They are expensive and time-consuming.
- Questionnaires are collection of questions that are phrased in a way that the answers to them provide detailed information on the topic of research.
- The observation technique of research is the method of directly examining a phenomena or the topic of research in its natural environment.
- The common types of observation techniques are: direct observation, unobtrusive observation, behaviour trace studies and disguised field observations.
- The method of case study in a research involves a deep study of a single individual, group or event.
- This unit also discusses the limitations of different methods of sampling in communication research.

3.7 KEY TERMS

- **Probability sampling:** A sampling procedure in which each element of a population has a known chance of being selected as a sample
- **Simple random sample:** A sample which is selected on the basis of chance
- **Stratified random sample:** A random sample of a specified size which is drawn from each stratum of the population
- **Cluster sampling:** A random sampling plan in which the population is subdivided into groups called clusters
- **Accidental sampling:** A type of non probability sampling where the sample is drawn from that part of the population which is the most easily accessible

NOTES

- **Quota sampling:** Sampling that involves a group of people, based on their demographic distribution
- **Purposive sampling:** A type of sampling which is carried out with a definite purpose
- **Snowball sampling:** A sampling method where samples are selected on the basis of subjects that have already been researched upon, for the same subject
- **Time allocation:** The allocation of time to different tasks and activities
- **Unobtrusive observation:** The method of observation that does not require the researcher's intrusion in the researched information

3.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Researchers have developed statistical means for determining sampling margins of error which they call confidence intervals.
2. There are two basic types of sampling techniques: probability sampling and non probability sampling.
3. A simple random sample refers to those cases that are selected so that each element in the population has an equal or a known chance of being included in the sample.
4. Inaccurate data collection can impact the results of a study and ultimately lead to invalid results.
5. Quantitative data collection methods produce results that are easy to summarize, compare and generalize.
6. Computer-assisted personal interviewing is a form of personal interviewing, where, instead of completing a questionnaire, the interviewer brings along a laptop or hand-held computer to enter the information directly into the database.
7. The history of case study research is marked by periods of intense use and periods of disuse.
8. Yin (1994) identified five components of research design, as important for case studies.
9. Direct observation occurs when a field visit is conducted during the case study.

NOTES

3.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define sampling in communication research.
2. Which are the typical quantitative data gathering strategies?
3. Which five components of research design were identified by Yin, as important for case studies?
4. What improves the perception of the fairness and serious thinking of the researcher?
5. What is the advantage of the purposive method of sampling?

Long-Answer Questions

1. Trace the significance of sampling in communication research.
2. Discuss the various methods of data collection.
3. How are case studies designed?
4. List and describe the limitations of different sampling methods.

3.10 FURTHER READING

Black, T. R. 1999. *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics.* London: Sage Publications Ltd. p. 118.

Cicourel, A. U. 1964. *Method and Measurement in Sociology.* New York: Free Press.

Cohen, L., Lawrence Manion and Keith Morrison. 2007. *Research Methods in Education.* London: Routledge

Glesne, C. 1992. *Becoming Qualitative Researchers: An Introduction.* New York: Longman.

Hamel, J., Stephanie Dufour and Dominic Fortin. 1993. *Case Study Methods.* California: Sage Publications, Inc.

Kvale, S. 1996. *Interviews: An Introduction to Qualitative Research Interviewing.* London: Sage Publications.

Leedy, D. Paul. 1974. *Practical Research: Planning and Design.* Michigan: Macmillan.

Mason, J. 1996. *Qualitative Researching.* London: Sage Publications.

Palmer, V. M. 1929. *Field Studies in Sociology: A Students' Manual.* Chicago: University of Chicago Press.

Patton, M. Q. 1987. *How to Use Qualitative Methods in Evaluation.* California: Sage Publications, Inc.

Stake, R. 1995. *The Art of Case Study Research.* California: Sage Publications, Inc.

Strauss, A. and Barney Glaser. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Chicago: Transaction Publishers.

Yin, R. 2009. *Case Study Research: Design and Methods.* California: Sage Publications, Inc.

NOTES

UNIT 4 STATISTICS IN COMMUNICATION RESEARCH

NOTES

Structure

4.0 Introduction
4.1 Unit Objectives
4.2 Use of Statistics in Communication Research
  4.2.1 Basic Statistical Tools
4.3 Measures of Central Tendency
4.4 Measures of Dispersion
  4.4.1 Standard Deviation
  4.4.2 Chi-square Test
  4.4.3 Correlation
4.5 Summary
4.6 Key Terms
4.7 Answers to 'Check Your Progress'
4.8 Questions and Exercises
4.9 Further Reading

4.0 INTRODUCTION

In this unit, we will study how statistics are used in communication research. This unit also discusses the basic statistical tools, measures of central tendency and measures of dispersion.

4.1 UNIT OBJECTIVES

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

- Discuss the use of statistics in communication research
- Explain the basic statistical tools
- Define the measures of central tendency (mean, median and mode)
- Define the measures of dispersion (standard deviation, correlation and chi-square)

4.2 USE OF STATISTICS IN COMMUNICATION RESEARCH

Statistics is the science of the collection, organization and interpretation of data. It deals with all aspects of this, including the planning of data collection in terms of

NOTES

the design of surveys and experiments In the discipline of communication research, statistics can be used to extract significant information from a huge mass of data. Statisticians are individuals who are extensively trained and equipped with advanced certifications in the discipline of statistics. They are authorized to deal with, define and draw conclusions from the information that is gathered. They have the skill to interpret the meaning of the collected data to consumers. They make optimum use of statistics, which is therefore significant in communication. Statistics can give a meaningful description to data by describing the population or sample, relating one variable to another, making estimations or predictions, testing hypotheses and deciding wisely.

Communication professionals come across statistical data that is growing in a dynamic way. Their task is to convert this data into meaningful form, to present it to the media in a global scenario that is constantly undergoing changes. This can be understood clearly with the following example: for making sure that their articles are precise and reliable, journalists have to conduct examination of a huge information database that is associated with economic, political and social phenomena. This data also comprise of scientific and academic research reports, public opinion data, academic, commercial and political polls, consumer surveys and countless other numeric data that are broadcasted everyday. Advertising and public relations executives should have the ability to interpret and use updated commercial data that is related to market trends, sales, competitive expenditures, audience ratings, click-throughs (an activity where a visitor clicks on a web advertisement and is taken to the advertiser's website) and other information that is linked to consumers, products, companies and markets. Given the enormous demand for up-to-date, accurate and actionable information, it is reasonable to suggest that statistical thinking and reasoning (Garfield et. al., 2002; Butler 1998; Snee 1990) should be considered a key learning goal for students of media and communication. In a large number of fields, the ability to understand, explain and thoroughly explain research data has emerged as a key (Gal (2002) and Giesbrecht 1996).

4.2.1 Basic Statistical Tools

Basic tools for statistical data analysis are standard deviation, F-statistics, etc. Several statistical tools are available in market. A few of them are fundamental, some are complex and others are customized especially for specific requirements. Actually, in the process of analysis, the most basic activity is the comparing of data to validate its accuracy and balance. By using some fundamental and suitable statistical tools, it is possible to extract a major part of the information which is required for routine laboratory processes (the 't-test, the 'F-test', and regression analysis).

Error

Error can be defined as the deviation of any outcome, from its real value. Errors can be of the following four types, random, systematic, constant and proportional.

NOTES

(i) Random error deviates between the results which are replicated and quantified with standard deviation.

(ii) Systematic errors regularly deviated from the 'true' value, which is quantified as the mean difference (i.e., difference between the true value and the mean of replicate determinations).

(iii) Constant errors are those errors, which are not related to the concentration of the substance that is analyzed (the analyte).

(iv) Proportional errors are those errors, which are concerned with the concentration of the analyte.

The 'true' value of a feature is basically not defined and usually has a comparative definition. However, this does not imply that it is not possible to analyse the content adequately.

Accuracy

Accuracy means the exactness or correctness of the outcome of an analysis, with respect to its real value. It comprises of a blend of random and systematic errors (precision and bias). However, in any circumstances, direct quantification of accuracy is not possible. It is likely for the test result to be an average of many values. Accuracy in determination results in a 'true' quantitative value, i.e., which is exact and has no prejudice.

Precision

Precision is defined as the nearness of an outcome to repeat an analysis of a sample. It is a factor that is used to gauge the dispersion or diffusion of the mean value and is often measured in terms of either standard deviation, standard error or a range (difference between the maximum and the minimum values).

Bias

Bias may be defined as the activity of analytical results, deviating constantly from the 'true' value. It is an outcome of systematic errors in a process. Bias is contrary to 'trueness'. However, it is often used to measure 'trueness', which is the compliance of the mean of analytical results with the true value, i.e., not including the contribution of randomness represented in precision.

Method bias

Method bias is defined as the difference between the average of test results, which is provided by a number of laboratories that use one technique and a reference value that is recognized by all of them. The method bias is likely to depend on the analytical level.

Laboratory bias

Laboratory bias is the difference between the average test result from a particular laboratory and the reference value that is acceptable.

NOTES

Sample bias

The difference between the average of the replicated test results of a sample and the ('true') value of the target population, from which the sample was taken, is known as sample bias. Practically, in case of a laboratory this mainly comprises of a sample preparation, sub sampling and weighing techniques.

It is basically assumed to be a set of data that is obtained by repetitive analysis of the same analyte, in the same sample, in the same conditions. It has a normal or Gaussian distribution. The more skewed is the distribution, the higher is the complexity of the statistical treatment. The primary parameters used are the mean and the standard deviation and the main tools the F-test, the t-test, regression and correlation analysis.

Standard deviation

Standard deviation can be calculated as the square root of variance (V). Variance is referred to, as the sum of the squared deviations from the mean, divided by n–1. In terms of operation, there are a number of ways to calculate variance:

$$S = \sqrt{\frac{\sum (x_1 - \bar{x})^2}{n-1}}$$

or

$$S = \sqrt{\frac{\sum x_1^2 - \left(\sum x_1\right)^2 / n}{n-1}}$$

or

$$S = \sqrt{\frac{\sum x_1^2 - n\bar{x}^2}{n-1}}$$

The mean and the standard deviation can be easily calculated by using a calculator. On a computer, it can be generated easily with the help of a dBASE programme.

Relative standard deviation

For a limited range, the standard deviation of analytical data does not vary much. This variation is directly proportional to the size of such data. The higher are the figures, the higher is the relative standard deviation. Therefore, to compare variations (e.g., precision) it is usually easier to use relative standard deviation rather than the standard deviation itself.

$$\text{RSD} = \frac{S}{\bar{x}} CV = \frac{S}{\bar{x}} \times 100\%$$

If required, if is possible to compute the variance as the square of the standard deviation:

$V = s^2$

NOTES

case is the two-sided (also called two-tailed) test. There exist no defined causes for expecting that the means or the standard deviations of two data sets are not the same. However, in case the mean and or the standard deviation are expected to go in a single direction, for instance, after a change in the analytical process, the one-sided (or one-tailed) test is an ideal choice. In this case the probability that it goes as expected is doubled. Or, more correctly, the uncertainty in the two-way test of 5 per cent (or the probability of 5 per cent that the critical value is exceeded) is divided over the two tails of the Gaussian curve, i.e., 2.5 per cent at the end of each tail beyond *2s*. If we perform a one-sided test with 5 per cent uncertainty, this is actually increased from 2.5 per cent to 5 per cent at the end of one tail. (Note that for the whole Gaussian curve, which is symmetrical, this is equivalent to an uncertainty of 10% in two ways).

This difference in probability in the tests is expressed by using two tables of critical values for both *F* and t. In fact, the one-sided table which is has a confidence level of 95 per cent is equivalent to the two-sided table which is has a confidence level of 90 per cent. This is underscored by the following observation: If in an experiment, the calculated values of F and t are found between the two-sided and one-sided values of F_{tab}, and t_{tab}, the two-sided test does not display much difference. On the other hand, the one-sided test indicates that the result of A is considerably different (more or less) than that of B. Actually, in the first case the 2.5 per cent boundary in the tail just does not exceed and then subsequently, this 2.5 per cent boundary is stretched till 5 per cent. This can then be obviously more easily exceeded. This exhibits that statistical tests vary in severity. To interpret the results in reports correctly, it is necessary to always specify the corresponding statistical techniques (including the confidence limits or probability).

F-test for precision

Since it is likely that the outcome of the *F*-test might be required to make a choice between the student's *t*-test and the Cochran variant, the *F*-test is discussed first. The *F*-test (or *Fisher's test*) is a comparison of the spread of two sets of data to test if both of them belong to the same population, in other words whether the precisions are alike or different.

The test makes use of the ratio of the two variances:

$$F = \frac{s_1^2}{s_2^2}$$

Here the larger s^2 must be the numerator by convention. If the performances are not very different, then the estimates s_1, and s_2, do not differ much and their ratio (and that of their squares) should not deviate much from unity. In practice, the calculated *F* is compared with the applicable *F* value in the F-table (also called the *critical* value,). To read the table it is necessary to know the applicable number of degrees of freedom for s_1, and s_2. These are calculated by:

$df_1 = n_1 - 1$

$df_2 = n_2 - 1$

If $F_{cal} = F_{tab}$ one can conclude with 95% confidence that there is no significant difference in precision (the 'null hypothesis' that $s_1 = s$, is accepted). Thus, there is still a 5% chance that we draw the wrong conclusion. In certain cases more confidence may be needed, then a 99% confidence table can be used, which can be found in statistical textbooks.

NOTES

t-Tests for bias

Depending on the nature of two sets of data (*n, s,* sampling nature), the means of the sets can be compared for bias by several variants of the *t*-test. It can be of following three types:

(i) Student's t-test for comparison of two independent sets of data with very similar standard deviations;

(ii) The Cochran variant of the *t*-test when the standard deviations of the independent sets differ significantly;

(iii) The paired *t*-test for comparison of strongly dependent sets of data.

For the *t*-tests Equation can be written in a different way also:

$$t_{cal} = \frac{|\bar{x} - \mu|}{s/\sqrt{n}}$$

Where,

$\bar{x}$ = mean of test results of a sample

μ = 'true' or reference value

s = standard deviation of test results

n = number of test results of the sample.

To compare the mean of a data set with a reference value normally the 'two-sided t-table of critical values' is used The applicable number of degrees of freedom here is:

$df = n - 1$

If a value for t calculated with the equation does not exceed the critical value in the table, the data are taken to belong to the same population: there is no difference and the 'null hypothesis' is accepted (with the applicable probability, usually 95%).

As with the F-test, when it is expected or suspected that the obtained results are higher or lower than that of the reference value, the one-sided *t*-test can be performed: if $t_{cal} > t_{tab}$, then the results are significantly higher (or lower) than the reference value.

NOTES

More commonly, however, the 'true' value of proper reference samples is accompanied by the associated standard deviation and number of replicates used to determine these parameters. We can then apply the more general case of comparing the means of two data sets: the 'true' value in Equation is then replaced by the mean of a second data set. if two data sets belong to the same population it is tested if the two Gauss curves do sufficiently overlap. In other words, if the difference between the means $\bar{x}_1 - \bar{x}_2$ is small. This is discussed next.

Similarity or non-similarity of standard deviations

When using the t-test for two small sets of data (n_1 and/or n_2<30), a choice of the type of test must be made depending on the similarity (or non-similarity) of the standard deviations of the two sets. If the standard deviations are sufficiently similar they can be 'pooled' and the student t-test can be used. When the standard deviations are not sufficiently similar an alternative procedure for the t-test must be followed in which the standard deviations are not pooled. A convenient alternative is the Cochran variant of the t-test. The criterion for the choice is the passing or non-passing of the F-test (see that is, if the variances do or do not significantly differ. Therefore, for small data sets, the F-test should precede the t-test.

For dealing with large data sets (n_1, n_2,< 30) the 'normal' t-test is used

Student's t-test

(To be applied to small data sets *(n_1, n_2* < 30) where s_1, and s_2 are similar according to *F-test.*

When comparing two sets of data, equation is rewritten as:

$$t_{cal} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Where,

$\bar{x}_1$ = mean of data set 1

$\bar{x}_2$ = mean of data set 2

s_p = 'pooled' standard deviation of the sets

n_1 = number of data in set 1

n_2 = number of data in set 2.

The pooled standard deviation s_p is calculated by:

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

Where,

s_1 = standard deviation of data set 1

s_2 = standard deviation of data set 2

n_1 = number of data in set 1

n_2 = number of data in set 2.

To perform the *t*-test, the critical t_{tab} has to be found in the table the applicable number of degrees of freedom *df* is here calculated by:

NOTES

$$df = n_1 + n_2 - 2$$

Example

The two data sets of Table can be used: With equations t_{cal} is calculated which is lower than the critical value t_{tab} (App. 1, $df = 18$, two-sided), hence the null hypothesis (no difference) is accepted and the two data sets are assumed to belong to the same population: there is no significant difference between the mean results of the two analysts (with 95% confidence).

$$\text{Isd} = \left|\bar{x}_1 - \bar{x}_2\right|_{lsd} = t_{tab} \cdot s_p \cdot \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

The calculation yields *lsd* = 0.69. The measured difference between the means is 10.34 – 9.97 = 0.37, which is smaller than the *lsd,* indicating that there is no significant difference between the performance of the analysts.

Additionally, in this approach 95 per cent of confidence limits of the difference between the means can be calculated as.

Confidence limits = 0.37 ± 0.69 = –0.32 and 1.06

Note that the value 0 for the difference is situated within this confidence interval which agrees with the null hypothesis of $x_1 = x_2$ (no difference) having been accepted.

Cochran's t-test

This test can be applied to small data sets (n_1, n_2, < 30) where s_1 and s_2, are dissimilar according to *F*-test.

Calculate *t* with:

$$t_{cal} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Then determine an 'alternative' critical *t*-value:

$$t_{tab} = \frac{t_1 \frac{s_1^2}{n_1} + t_2 \frac{s_2^2}{n_2}}{t_1 \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

NOTES

Where,

$t_1 = t_{tab}$ at $n_1 - 1$ degrees of freedom

$t_2 = t_{tab}$ at $n_2 - 1$ degrees of freedom

The *t*-test can be performed as usual: if $t_{cal} < t_{tab}$ then the null hypothesis that the means do not significantly differ is accepted.

t-Test for large data sets (n < 30)

In previous example, the conclusion happens to have been the same if the Student's *t*-test with pooled standard deviations had been used. This is caused by the fact that the difference in result of the Student and Cochran variants of the *t*-test is largest when small sets of data are compared and decreases with increasing number of data. More precisely, with increasing number of data, a better estimate of the real distribution of the population is obtained (the flatter t-distribution converges to the standardized normal distribution). When n < 30 for both sets, e.g., when comparing control charts, for all practical purposes, the difference between the Student and Cochran variant is negligible. The procedure is then reduced to the 'normal' *t*-test by simply calculating t_{cal} and comparing this with t_{tab} at $df = n_1 + n_2 - 2$.

Paired t-test

If two sets of data are dependent, the paired t-test proves to be a better instrument to be used to compare, than the regular *t*-test. This can be understood better by an example where two methods are compared by the same analyst, using the same sample(s). It could, in fact, also be applied when the two analysts use the same analytical method at (about) the same time.

The comparison of two methods which use different levels of analyte provides more reliable information as compared to the methods which use only one level. The results can be compared at every level by the *F* and *t*-tests. The paired *t*-test, however, supports different levels if the concentration range is not very broad. According to rule of fist, the range of results should be within the same magnitude. It is wise to consider regression analysis, if a wide range is to be covered, i.e., several powers of ten. However, in intermediate cases, any technique may be used.

The null hypothesis exists when the data sets do not differ, so the test is to see if the mean of the differences between the data deviates significantly from zero or not (two-sided test). If one set is expected to be systematically higher (or lower) than the other, then the one-sided test is appropriate.

Analysis of variance

The ANOVA (analysis of variance) is a robust statistical tool. It is used to compare the results of laboratories or techniques where the influencing factor is more than one. Examples of such factors are: different analysts, samples with different pre-treatments, different analyte levels and different methods within one of the laboratories). A large number of statistical packages can perform this analysis for a PC.

NOTES

Error or uncertainty in the regression line

The 'fitting' of the calibration graph is necessary because the response point y_i, composing the line, does not fall exactly on the line. This implies random errors. These are indicated by an uncertainty about the slope and intercept b and a, defining the line. A quantification can be found in the standard deviation of these parameters. Most computer programmes for regression will automatically produce figures for these. To illustrate the procedure, A practical quantification of the uncertainty is obtained by calculating the standard deviation of the points on the line; the 'residual standard deviation' or 'standard error of the y-estimate', which we assumed to be constant (but which is only approximately.

$$s_y = \sqrt{\frac{\sum(y_1 - \hat{y}_1)}{n-2}}$$

Where,

$\hat{y}$ = 'fitted' y-value for each x_i. Thus, $y_1 - \hat{y}_1$ is the (vertical) deviation of the found y-values from the line.

n = number of calibration points.

Only the y-deviations of the points from the line are considered. It is assumed that deviations in the x-direction are negligible. This is true, only in the case where the standards are very accurately prepared.

Now the standard deviations for the intercept 'a' and slope 'b' can be calculated with:

$$s_a = \sqrt{\frac{\sum x_1^2}{n\sum(x_1 - \bar{x})^2}}$$

and

$$s_b = \sqrt{\frac{S_y}{\sum(x_i - \bar{x})^2}}$$

The uncertainty about the regression line is expressed by the confidence limits of 'a' and 'b' according to Equation $a \pm t.s_a$ and $b \pm t.s_b$.

Check Your Progress

1. What is meant by statistics?
2. Name two common basic tools for statistical data analysis.
3. Name the common types of bias.
4. Define standard deviation.
5. When is the paired t-test a better tool for comparison than the normal t-test?

NOTES

4.3 MEASURES OF CENTRAL TENDENCY

There are three most commonly used techniques to measure central tendency. These are, mean, median and mode.

(i) Mean

The sum of the values divided by the number of values is known as mean. It is often called the 'average.'

The average of a set of n data x_i:

$$\bar{x} = \frac{\sum x_1}{n}$$

- Add all the values together
- Divide by the number of values to obtain the mean

Example: The mean of 7, 12, 24, 20, 19 is (7 + 12 + 24 + 20 + 19) / 5 = 16.4.

(ii) Median

The value which divides a range into two equal halves, such that one half of the values are lower than the median and the other half are higher than the median.

- Sort the values into ascending order. If you have an odd number of values, the median is the middle value.
- If you have an even number of values, the median is the arithmetic mean of the two middle values.

Example: The median of the same five numbers (7, 12, 24, 20, 19) is 19.

(iii) Mode

The most frequently-occurring value (or values) is known as the mode.

- Calculate the frequencies for all of the values in the data.
- The mode is the value (or values) with the highest frequency.

Example: For individuals for the following ages—18, 18, 19, 20, 20, 20, 21 and 23, the mode is 20.

4.4 MEASURES OF DISPERSION

While measures of central tendency are used to estimate 'normal' values of a dataset, measures of dispersion are important to describe the distribution of data, or its variation around a central value. It is possible for two distinct samples to have the same mean or median, but completely different levels of variability, or vice versa. An ideally descriptive set of data is one that includes both of these characteristics. Numerous techniques can be used for measuring the dispersion of a dataset.

NOTES

Range

- The difference between the highest and the lowest sample value
- One of the simplest measures of variability to calculate
- Depends only on extreme values and provides no information about the distribution of the remaining data

4.4.1 Standard Deviation

(i) Standard deviation is the square root of the sample variance.

(ii) Defined so that it can be used to make inferences about the population variance.

(iii) Calculated using the formula: $S_{N-1} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_1 - \bar{x})^2}$

(iv) The values computed in the squared term, $x_1 - \bar{x}$ are anomalies.

(v) Not restricted to large sample datasets, compared to the root mean square anomaly discussed later in this section.

(vi) Provides significant information into the distribution of data around the mean, approximating normality.

a. The mean ± one standard deviation contains approximately 68 per cent of the measurements in the series.

b. The mean ± two standard deviations that contain approximately 95 per cent of the measurements in the series.

c. The mean ± three standard deviations that contain approximately 99.7 per cent of the measurements in the series.

The meteorological department often uses standard deviations to help classify abnormal climatic conditions. The chart below describes the abnormality of a data value and the standard deviations that distance it from the mean. The probabilities in the third column assume that data is normally distributed.

Standard Deviations Away From Mean	Abnormality	Probability of Occurrence
beyond –3 sd	extremely subnormal	0.15%
–3 to –2 sd	greatly subnormal	2.35%
–2 to –1 sd	subnormal	13.5%
–1 to +1 sd	normal	68.0%
+1 to +2 sd	above normal	13.5%
+2 to +3 sd	greatly above normal	2.35%
beyond +3 sd	extremely above normal	0.15%

Root mean square anomaly/Root mean square

Root mean square anomaly is:

- Also known as root mean square deviation.
- Very similar to standard deviation, except used for large sample sizes (i.e., divisor is n instead of n – 1) (Devore).

NOTES

- RMSA calculated using the formula: $S_{N-1} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_1 - \bar{x})^2}$, where xbar is the mean, x_i is each data value, and n is the number of observations.
- The term $x_1 - \bar{x}$ is an anomaly, which is discussed in another section.
- Provides similar information into the dispersion of data as the standard deviation.
- Often used as a measurement of error.
- More commonly used than the standard deviation function, in the statistical analysis of climate data because climate-related datasets are generally quite large in size, in terms of number of data points.

Root mean square

- Is calculated using the following formula:

$$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_1^2} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N}}$$

- Unlike the RMSA or standard deviation, the mean is not removed from the calculation.
- It is acceptable for use only when dealing with large sample datasets (Devore).

4.4.2 Chi-square Test

A chi-square test (also chi-squared test or χ^2 test) is any statistical hypothesis test where the sampling distribution of the test statistic is a chi-square distribution. Here, the null hypothesis is true, which means that the sampling distribution (if the null hypothesis is true) can be made to approximate a chi-square distribution as closely as desired by making the sample size large enough.

- Pearson's chi-square test is also known as the chi-square goodness-of-fit test or chi-square test for independence. When mentioned without any modifiers or without other precluding context, this test is usually understood (for an exact test used in place of χ^2, see Fisher's exact test).
- Yates' chi-square test, is also known as Yates' correction for continuity
- Mantel–Haenszel chi-square test.
- Linear-by-linear association chi-square test.
- The portmanteau test in time-series analysis, testing for the presence of autocorrelation

One case where the distribution of the test statistic is an exact chi-square distribution is the test where the variance of a normally-distributed population has a given value. This value is based on a sample variance. Such a test is not common in practice because values of variances to test against are seldom known exactly.

NOTES

Chi-square is a statistical test which is commonly used to compare observed data with the data that is expected to emerge, 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. It would be possible to find out whether the deviations (differences between observed and expected) are a result of chance, or are due to other factors. 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.

The formula for calculating chi-square (χ^2) is:

$$\chi^2 = (o - e)^2/e$$

That is, chi-square is the sum of the squared difference between observed (o) and the expected (e) data (or the deviation, d), divided by the expected data in all possible categories.

- Chi-square requires that you use numerical values, not percentages or ratios.
- Chi-square should not be calculated if the expected value, in any category, is less than 5.

Calculating chi-square

	Green	Yellow
Observed (o)	639	241
Expected (e)	660	220
Deviation (o – *e)*	–21	21
Deviation^2 (d^2)	441	441
d^2/e	0.668	2
$\chi^2 = d^2/e = 2.668$	.	.

Chi-Square Distribution

Degrees of Freedom *(df)*	Probability (*p*)										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Nonsignificant								Significant		

NOTES

4.4.3 Correlation

Correlation is one of the most popular and widely used instruments of statistics. It is a single number that defines the depth of the relationship between two variables. The following example shows how this statistic is computed.

Example of Correlation

This example attempts to understand the relationship between two variables, height (in inches) and self esteem. A hypothesis may state that one's height affects one's self esteem. For this some information is collected on twenty individuals (since it is known that the average height differs for males and females, so to simplify this example we use males). The measurement of height is in inches. The measurement of self esteem is on the basis of the average of 10 (1-to-5) rating items (where higher scores mean higher self esteem). The data for 20 cases is listed below:

Person	Height	Self Esteem
1	68	4.1
2	71	4.6
3	62	3.8
4	75	4.4
5	58	3.2
6	60	3.1
7	67	3.8
8	68	4.1
9	71	4.3
10	69	3.7
11	68	3.5
12	67	3.2
13	63	3.7
14	62	3.3
15	60	3.4
16	63	4.0
17	65	4.1
18	67	3.8
19	63	3.4
20	61	3.6

Figure 4.1 is a quick look histogram for each variable:

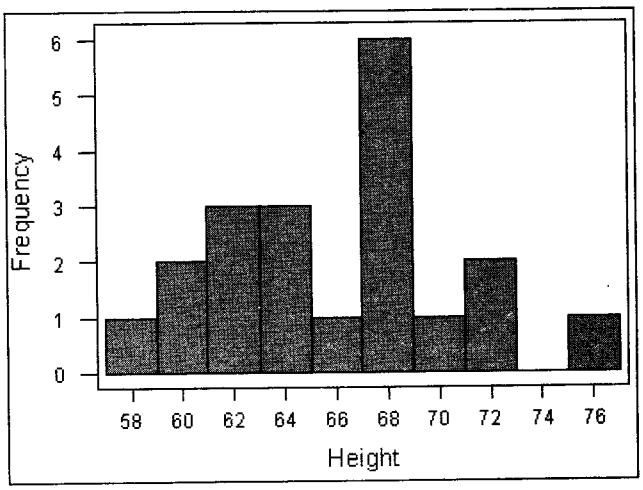

Fig. 4.1(a) Height and Frequency Histogram

NOTES

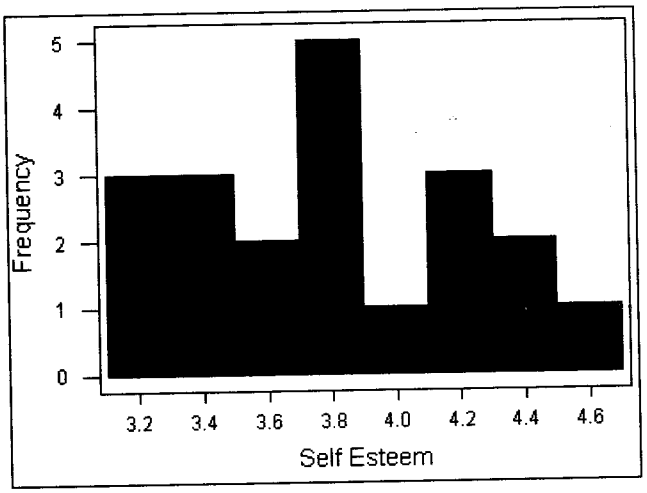

Fig. 4.1(b) Frequency and Self Esteem Histogram

And, here are the descriptive statistics:

Variable	Mean	StDev	Variance	Sum	Minimum	Maximum	Range
Height	65.4	4.40574	19.4105	1308	58	75	17
Self Esteem	3.755	0.426090	0.181553	75.1	3.1	4.6	1.5

Finally, Figure 4.2 is a simple bivariate (i.e., two-variable) plot:

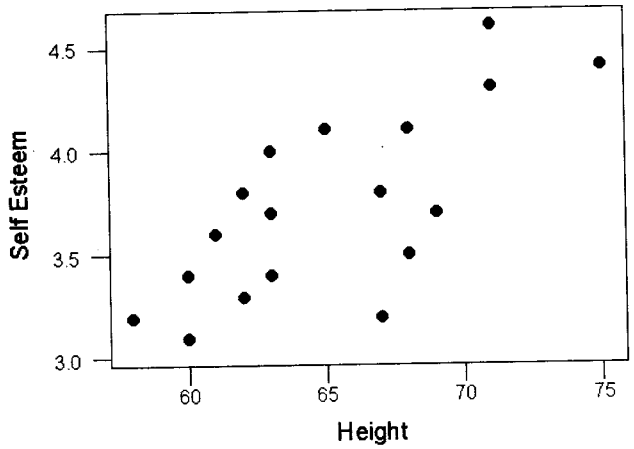

Fig. 4.2 Simple Bivariate Plot

In the bivariate plot, it should be immediately evident that there is a positive relationship between the variables. This is because if a single straight line were to emerge by joining the dots, it would be a positively inclined slope or move up from left to right. Since the correlation is nothing more than a quantitative estimate of the relationship, a positive correlation would be expected.

In this context, what does a positive relationship signify? It means that, generally, higher scores on one variable are likely to pair with higher scores on the other. Likewise, lower scores on one variable are likely pair with lower scores on the other. One needs to visually confirm that this is generally true in the above case.

NOTES

Calculation of correlation

In the following example, we prepare to calculate the correlation value. The formula o compute correlation is as follows:

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where,

N = number of pairs of scores
Σxy = sum of the products of paired scores
Σx = sum of x scores
Σy = sum of y scores
Σx^2 = sum of squared x scores
Σy^2 = sum of squared y scores

The symbol r stands for the correlation. As per mathematics, it turns out that r will always be between –1.0 and +1.0. if the correlation is negative, we have a negative relationship; if correlation is positive, the relationship is positive. To understand the relation of the formula to real data, one can use the formula to compute the correlation. Let us look at the data needed for the formula. The original data with the other necessary columns is as follows:

Person	Height (x)	Self Esteem (y)	x*y	x*x	y*y
1	68	4.1	278.8	4624	16.81
2	71	4.6	326.6	5041	21.16
3	62	3.8	235.6	3844	14.44
4	75	4.4	330	5625	19.36
5	58	3.2	185.6	3364	10.24
6	60	3.1	186	3600	9.61
7	67	3.8	254.6	4489	14.44
8	68	4.1	278.8	4624	16.81
9	71	4.3	305.3	5041	18.49
10	69	3.7	255.3	4761	13.69
11	68	3.5	238	4624	12.25
12	67	3.2	214.4	4489	10.24
13	63	3.7	233.1	3969	13.69
14	62	3.3	204.6	3844	10.89
15	60	3.4	204	3600	11.56
16	63	4	252	3969	16
17	65	4.1	266.5	4225	16.81
18	67	3.8	254.6	4489	14.44
19	63	3.4	214.2	3969	11.56
20	61	3.6	219.6	3721	12.96
Sum =	1308	75.1	4937.6	85912	285.45

The first three columns are the same as in the table above. The next three columns are simple computations based on the height and self esteem data. The bottom row consists of the sum of each column. This is all the information that is needed to compute the correlation. Here are the values from the bottom row of the table (where N is 20 people), as they are related to the symbols in the formula:

NOTES

$N = 20$
$\Sigma xy = 4937.6$
$\Sigma x = 1308$
$\Sigma y = 75.1$
$\Sigma x^2 = 85912$
$\Sigma y^2 = 285.45$

Now, when we plug these values into the formula given above, we get the following (Here, it is tediously shown one step at a time):

$$r = \frac{20(4937.6) - (1308)(75.1)}{\sqrt{[20(85912) - (1308*1308)][20(285.45) - (75.1*75.1)]}}$$

$$r = \frac{98752 - 98230.8}{\sqrt{[1718240 - 1710864][5709 - 5640.01]}}$$

$$r = \frac{521.2}{\sqrt{[7376][68.99]}}$$

$$r = \frac{521.2}{\sqrt{508870.2}}$$

$$r = \frac{521.2}{713.3514}$$

$$r = 0.73$$

So, the correlation for our twenty cases is 0.73, which is a fairly positive relationship. It is likely there is a relationship between height and self esteem, at least in this made up data.

Testing the significance of a correlation

Once a correlation has been computed, one can determine the probability that the observed correlation occurred by chance. That is, a significance test can be conducted. Most it is required to determine the probability of the correlation being real and not a chance occurrence. In this case, the mutually exclusive hypothesis is being tested:

Null Hypothesis:	r = 0
Alternative Hypothesis:	r <> 0

The easiest way to test this hypothesis is to find a statistics book that has a table of critical values of r. Most introductory statistics texts would have a table like this. As in all hypothesis testing, you need to first determine the significance level. Here, we use the common significance level of alpha = 0.05. This means conducting a test where the odds are that the correlation, as a chance occurrence, is no more than 5 out of 100. Before one looks at the critical value in a table, one also has to compute the degrees of freedom or df. The df is simply equal to N–2

NOTES

or, in this example, is 20 – 2 = 18. Finally, one has to decide whether between a one-tailed or two-tailed test. In this example, since there is no strong prior theory to suggest whether the relationship between height and self esteem would be positive or negative, one can opt for a two-tailed test. With this information, the significance level (alpha = 0.05), degrees of freedom (df = 18) and type of test (two-tailed), one can now test the significance of the correlation that is found. When one looks up this value in the handy little table at the back of my statistics book, it is found that the critical value is 0.4438. This means that if the correlation is greater than +4438 or less than – 0.4438 (keeping in mind that this is a two-tailed test). It can be concluded that the odds are less than 5 out of 100 and that this is a chance occurrence. Since a correlation of 0.73 is actually quite a bit higher, it can be concluded that it is not a chance finding and that the correlation is 'statistically significant' (given the parameters of the test). The null hypothesis can be rejected and the alternative may be accepted.

The correlation matrix

So far, we have seen how to compute a correlation between two variables. In most studies, the number of variables is more than two. Let us take an instance of a study of 10 interval-level variables, where one wants to estimate a relationship between all of them (i.e., between all possible pairs of variables). In this instance, there are 45 unique correlations to estimate. We could do the above computations 45 times to obtain the correlations. Or we could use just about any statistics program to automatically compute all 45, with a simple click of the mouse.

Below, a simple statistics program is used to generate random data for 10 variables with 20 cases (i.e., persons) for each variable. Then, the program was to compute the correlations among these variables. The result is as follows:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1.000									
C2	0.274	1.000								
C3	-0.134	-0.269	1.000							
C4	0.201	-0.153	0.075	1.000						
C5	-0.129	-0.166	0.278	-0.011	1.000					
C6	-0.095	0.280	-0.348	-0.378	-0.009	1.000				
C7	0.171	-0.122	0.288	0.086	0.193	0.002	1.000			
C8	0.219	0.242	-0.380	-0.227	-0.551	0.324	-0.082	1.000		
C9	0.518	0.238	0.002	0.082	-0.015	0.304	0.347	-0.013	1.000	
C10	0.299	0.568	0.165	-0.122	-0.106	-0.169	0.243	0.014	0.352	1.000

This type of table is called a correlation matrix. It lists the variable names (C1 – C10) down the first column and across the first row. The diagonal of a correlation matrix (i.e., the numbers that go from the upper left corner to the lower

NOTES

right) always consists of ones. That is because these are the correlations between each variable and itself (and a variable is always perfectly correlated with itself). This statistical program focuses only on the lower section of the correlation matrix. Every correlation matrix comprises of two triangles that are the values below and to the left of the diagonal (lower triangle) and above and to the right of the diagonal (upper triangle). There is no reason to print both triangles because the two triangles of a correlation matrix are always mirror images of each other (the correlation of variable x with variable y is always equal to the correlation of variable y with variable x). When a matrix has this mirror-image quality above and below the diagonal we refer to it as a symmetric matrix. A correlation matrix is always symmetric.

To locate the correlation for any pair of variables, find the value in the table for the row and column intersection for those two variables. For instance, to find the correlation between variables C5 and C2, we look for row C2 and column C5 (in this case it is blank because it falls in the upper triangle area) and row C5 and column C2. In the second case, we find that the correlation is –0.156.

So how did we know that there are 45 unique correlations when we have 10 variables? There is a handy formula that indicates the number of pairs (e.g., correlations) there are for any number of variables:

$$\frac{N*(N-1)}{2}$$

Where N is the number of variables. In the above example, we had 10 variables, so we know we have (10 * 9)/2 = 90/2 = 45 pairs.

Other correlations

The specific type of correlation that has been illustrated here is known as the Pearson Product Moment Correlation. It is appropriate when both variables are measured at an interval level. However, there are a wide variety of other types of correlations for other circumstances. For instance, if you have two ordinal variables, you could use the Spearman Rank Order Correlation (rho) or the Kendall Rank Order Correlation (tau). When one measure is a continuous interval level and the other is dichotomous (i.e., two-category) one can use the Point-Biserial Correlation. For other situations, consult the web-based statistics selection program.

Check Your Progress

6. Name the three most commonly used techniques to measure central tendency.
7. Why are the measures of dispersion important?
8. What is the other name for Pearson's chi-square test?
9. For what purpose is the chi-square test commonly used?
10. Which is one of the most common and also the most useful tool in statistics?

NOTES

4.5 SUMMARY

- Statistics are actually used in most of the communication researches that are undertaken in communication studies.
- Basic tools for statistical analysis are: standard deviation, F-statistics, etc.
- The three most common measures of central tendency are: the mean, the median and the mode.
- Three common measures of dispersion are: standard deviation, chi-square test and correlation.
- The measures of central tendency are not as useful for statistical analysis, as are the measures of dispersion of the values around central tendency.

4.6 KEY TERMS

- **Standard deviation:** A measure of the variation between individuals on a variable
- **F-statistics:** A method of statistical analysis which is frequently used to describe the presence of population structure
- **Standard error:** The standard deviation of the sampling distribution associated with the estimation method
- **Bias:** A term used to describe a tendency or preference towards a particular perspective, ideology or result
- **Method bias:** Difference between the (mean) test results, which is obtained from a number of laboratories using the same method and an accepted reference value
- **Laboratory bias:** Difference between the (mean) test result from a particular laboratory and the accepted reference value
- **Sample bias:** Difference between the mean of replicate test results of a sample and the ('true') value of the target population from which the sample was taken
- **t-test:** Any statistical hypothesis test in which the test statistic follows a student's t distribution, if the null hypothesis is true
- **Null hypothesis:** A hypothesis (within the most frequent context of statistical hypothesis testing) that might be falsified using a test of observed data
- **Mean:** An average of n numbers computed by adding some function of the numbers and dividing by some function of n
- **Median:** The numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half

- **Mode:** The value that occurs the most frequently in a data set or a probability distribution

4.7 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Statistics is the science of the collection, organization and interpretation of data.
2. Two common basic tools for statistical data analysis are standard deviation and F-statistics.
3. The common types of bias are: method bias, laboratory bias and sample bias.
4. Standard deviation can be defined as the square root of the variance (V).
5. The paired t-test is a better tool for comparison, than the normal *t*-test, when two data sets are not independent.
6. The three most commonly used techniques to measure central tendency are: mean, median and mode.
7. The measures of dispersion are important for describing the spread of data, or its variation around a central value.
8. Pearson's chi-square test is also known as the chi-square goodness-of-fit test or chi-square test for independence.
9. Chi-square is a statistical test which is commonly used to compare observed data with the data that is expected to emerge, according to a specific hypothesis.
10. Correlation is one of the most common and also the most useful tool in statistics.

4.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. How is statistics in communication research important for advertising and public relations?
2. What is the use of the Cochran's test?
3. Which tool is more useful for comparison, when the two data sets are not independent?
4. How is mean calculated?
5. Define the term range, in statistical analysis.

NOTES

Long-Answer Questions

1. Discuss the use of statistics in communication research.
2. List and define the measures of central tendency.
3. What are the different measures of dispersion and how are they used?
4. Write a note on the chi-square test.

4.9 FURTHER READING

Giesbrecht, Norman. 1996. *Strategies for Developing and Delivering Effective Introductory-Level Statistics and Methodology Courses.*

Oliver, John E and E. J. Arnold. 1981. *Climatology: Selected Applications.* p.45.

Sterling, Christopher, James K. Bracken and Susan M. Hill. 1998. *Mass Communications Research Resources: An Annotated Guide.* New Jersey: Lawrence Erlbaum Associates.

Snee, R.. *Statistical Thinking and its Contribution to Quality.* The American Statistician, No. 44, 1990, pp. 116–121.

UNIT 5 DATA PROCESSING

NOTES

Structure

5.0 Introduction
5.1 Unit Objectives
5.2 Data Processing
5.3 Data Analysis
5.4 Presenting and Interpreting Data
  5.4.1 Data Presentation
  5.4.2 Role of Graphics in Presenting Data
  5.4.3 Data Interpretation
5.5 How to Write a Research Proposal
5.6 Writing Research Report-Component and Style
5.7 Summary
5.8 Key Terms
5.9 Answers to 'Check Your Progress'
5.10 Questions and Exercises
5.11 Further Reading

5.0 INTRODUCTION

At its most basic level, every research pertains to data. Processing data, as a result, is the fundamental science of research. However, there is an art to this science and it requires an understanding that getting the most out of data involves much more than just crunching numbers and manipulating codes. It takes a strategic approach that anticipates the best form for the data, integrates data from different sources together and presents the data in an understandable and compelling format. This section discusses how data is processed, analyzed, presented and interpreted. It also explains how graphics are used in data presentation and how a research proposal and report are written.

5.1 UNIT OBJECTIVES

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

- Define the concept of data processing
- Analyse data
- Present and interpret data
- Use graphics in data presentation
- Outline the process of writing a research proposal

5.2 DATA PROCESSING

NOTES

The word data is commonly used to understand 'information.' Some times large amounts of information are required in a standard format. Data may comprise of letters, numbers, equations, dates, images and other material. Data processing is the activity that involves a number of structured procedures that make information meaningful. In information processing, data is obtained, collated, evaluated, processed, stored and output. This is a result of the response to queries or in the form of routine reports. Data processing is the activity where one or more sets of data are recorded or handled.

The Bureau of Labor Statistics (BLS), in their publication occupational outlook handbook, 2010–2011 edition, have discussed the topic of 'Data Entry and Information Processing Workers.' In the handbook, raw data is differentiated from processed information: the two are referred to as distinctly, with data that precedes information.

Britannica Concise Encyclopedia has a different outlook. It discusses both, raw data that is keyed into a computer system, as well as processed information that results as output from the computer.

In the language of computers, data processing may be termed as the use of a software application for working on some kind of input data, for creating an output. The outcome may range from a multimedia file to an image. or a text file. In general, keying in words cannot be termed as data entry. Similarly, making use of a word processor to write a story cannot be called data processing.

Data processing is basically the transmission of data from one source to another. Those who are engaged in data processing jobs, frequently have different titles such as, typist, transcriber, data specialist, word processor and keypunch technician. Data processing may be considered as the act of entering monthly receipts and expenses into a spreadsheet program and printing a report. There are two types of data processing: general and computing.

General: is a term that covers activities which are performed on specific data. These activities are aimed at filtering the necessary information in a required format. The format may be in the form of diagrams, reports, or tables.

Computing: is the activity of manipulation of input data with an application program to obtain desired output as an audio/video, graphic, numeric, or text data file. Actually, data are raw facts which can be processed into useful information. The process of converting facts to information is known as data processing. In practice, every process that occurs naturally may be regarded as an example of data processing systems. Information in the form of pressure, light, etc., are transformed into electrical signals, in the nervous system. This is similar to the senses of recognition like, touch, sound and vision. The interaction of non-living systems might also be regarded as rudimentary information processing systems.

The use of conventional terms like data processing and information systems restrict themselves to algorithmic derivations, logical deductions and statistical calculations that repeat consistently in overall commercial scenarios.

NOTES

Elements of data processing

Data needs first be converted into a machine readable format before it is processed by a computer. Once data is in digital format, various procedures can be applied on it to get useful information. It includes various processes like:

Data summarization

Primitive and derived data are summarized to extract evaluated information from it. This is done to achieve data that is more generic in value. Since a data warehouse has huge volumes of data, it needs to have a mechanism in place for extracting the required data from the excess junk. Data summarization gives a general view of different sets of data to consumers who require the data.

Data aggregation

Data aggregation is the process of achieving a value by means of calculating the aggregate of two or more key characteristics of data. Aggregation may be achieved by focusing on different values of data, pertaining to the same subject. The subjects may link commercial dealings and a de-normalized database to the real world and detailed data resource design within the common data architecture.

Data validation

Data validation is the process of ensuring that a program operates on clean, correct and useful data. *The technique* of validation provides for character checks. These checks ensure the presence of only expected characters in a field, for instance, a numeric field would only have numbers ranging from 0 to 9, the decimal point and probably a minus sign or commas.

Data tabulation

Data tabulation is a method of statistics that gives rise to a mutually dependent interrelationship between two tables of values. However, this does not determine a casual relationship between the two tables. This can be studied in a better way by taking an instance where, in a car factory, a two way tabulation may exhibit that cars that were made on Wednesday had more service problems than those that were built on Monday. Cross tabulation also finds its use in the analysis of the outcome of a consumer survey, which exhibits the division of preferences for certain advertisements. This preference depends on the part of the country of which a consumer is a resident.

Statistical analysis

Statistical analysis is the mathematics of the collection, organization and interpretation of numerical data, especially the analysis of population characteristics by inference

NOTES

from sampling. Statistical tools can use indicators like mean and standard deviation to summarize past data. In addition, future events can also be predicted through the use of frequency distribution functions. With the help of statistics, efficient experiments can be designed to cut down on the time involved in trial and error methods. Statistical analysis is also beneficial to double-blind tests for polls, intelligence and aptitude tests and medical, biological and industrial experiments. Though the reliability factor varies, every outcome is instrumental in predicting the future performance.

5.3 DATA ANALYSIS

When data is gathered from a scientific or an engineering sphere, terms like data processing and information systems have a wide range. A more dedicated concept of data analysis is typically used in such cases. Here, the focus lies on the highly-specialized and highly-accurate algorithmic derivations and statistical calculations. These deviations are not common occurrences in the typical general business environment. The highlight of this context is the free usage of data analysis packages like DAP, GRETL or PSPP. The cultural divergence is indicated by special numeric representations that play an active role in the processing of data. More specifically, data processing's measurements are typically represented by integers or by fixed-point or binary-coded decimal figures. On the other hand, a large amount of measurements that pertain to data analysis are often represented by floating-point representation of rational numbers. Data analysis can be both, quantitative as well as qualitative.

Quantitative data analysis: is the process of presenting and interpreting numerical data

Qualitative data analysis: is the process of interpreting data that is collected during the course of qualitative research.

Process of data analysis: includes four main processes, i.e., data cleaning, initial data analysis, main data analysis and most importantly, final stage of data analysis.

(i) Data cleaning

Data cleaning is a vital procedure which comprises of inspection of data. In this process, errors within the data are necessarily, preferably and possibly corrected. Data can be filtered during the entry stage. Once this is accomplished, it is important to avoid subjective decision-making. The guiding principle provided by Adèr (ref) is: during subsequent manipulation of the data, information should always be cumulatively retrievable. In other words, there should always be room for the possibility to undo any alterations in the data set. Hence, it is emphasized to guard information at every phase of data cleaning. Every bit of information should be recorded and kept intact (i.e., when altering variables, both, the original and the latest values should be saved, in a duplicate dataset or under a different variable

name). It is also important to exercise care and clarity in documenting all alterations to the data set, for instance in a syntax or a log.

(ii) Initial data analysis

NOTES

The fundamentally prominent difference between the initial and the main data analysis phases is that during the initial phase, any analysis aimed at answering the original research question is avoided. The following questions guide the initial data analysis phase:

a. Quality of data

It is always advisable to make a quality check of the data in the initial stage. There are different ways to measure the quality of data, they are, different types of analyses: frequency counts, descriptive statistics (mean, standard deviation, median), normality (skewness, kurtosis, frequency histograms, normal probability plots) and associations (correlations, scatter plots).

b. Quality of measurements

An early check should be made on the quality of the instruments used for measurement purpose. This is due to the fact that in the initial stage, it is not the key area of focus or the research question of the study. The structure of measurement instruments should match with that reported in the literature.

(iii) Main data analysis

The main analysis comprises of answering research questions along with other correlated and significant analyses. These are vital for writing a first draft of the research report.

(iv) Exploratory and confirmatory approaches

The main analysis may comprise of an approach that tends to explore or one that tends to confirm. The type of approach is confirmed before the collection of data. There is no clarity of hypothesis before the analysis of data. A confirmatory analysis tests the clarity of data.

It is important to exercise care while interpreting exploratory analyses. At the time of testing multiple models at the same time, there is a high probability of at least one of them being significant. The reason for this may be a type 1 error. At the time of testing multiple models, the level of significance needs to be appropriately adjusted, for instance, a Bonferroni correction. Additionally, the same dataset should not be subjected to both, an exploratory analysis and a confirmatory analysis simultaneously. An exploratory analysis is useful in finding ideas for a theory, but not for testing the theory. When a model is found exploratory in a dataset, then following up that analysis with a confirmatory analysis in the same dataset could simply mean that the results of the confirmatory analysis are due to the same type 1 error that resulted in the exploratory model in the first place. In this case, the confirmatory analysis will provide less information in comparison to the original exploratory analysis.

NOTES

(v) Final stage of data analysis

The final stage begins with the documentation of the findings of the initial data analysis. Necessary, preferable and possible corrective actions are taken for documenting the same. Additionally, it is advisable for specifying the initial plan for analysis of data in a more elaborated way. For this purpose, a number of decisions pertaining to the analysis of data, may and must be taken.:

- In the case of non normals: variables should be transformed; variables should be categorized (ordinal/dichotomous); the technique of analysis should be adapted to.
- If the data is not available: the unavailable data should be challenged or ignored; the technique of challenging should be selected.
- In the presence of outliers: the feasibility of using robust analysis techniques may be determined.
- If the items do not fit the scale: the question to be asked is whether should one adapt to the instrument of measurement by ignoring items, or make sure that they are comparable with other (uses of the) measurement instrument(s)?
- Where the subgroups are very small in size: should the hypothesis pertaining to inter-group variations be discarded and replaced with small sample techniques, like exact tests or bootstrapping?
- If the process of randomization exhibits defects: would it be advisable for one to compute propensity scores and make them covariates in the main analyses?

Check Your Progress

1. What is data processing?
2. What is data?
3. Define data validation.

5.4 PRESENTING AND INTERPRETING DATA

5.4.1 Data Presentation

Easy-to-understand tables and graphics will greatly enhance the readability of the written research report. As a general rule, all tables and figures should contain:

(i) Identification number corresponding to the list of tables and the list of figures

(ii) A title that conveys the content of the table or figure, also corresponding to the list of tables and the list of figures

(iii) Appropriate column labels and row labels for tables and figure legends to define essential parts of the figure.

One may represent data that is separate, or in groups as: Pie charts, frequency histograms, frequency polygons, ogives and box plots, for example, survey results of the ages of students in the adult basic education maths classes are shown in Table 5.1.

Table 5.1 Frequency Data

Age Interval(yrs)	Frequency
15-19	13
20-24	15
25-29	20
30-34	10
35-39	8
40-44	4

This data can be used to create a frequency histogram as shown in Figure 5.1.

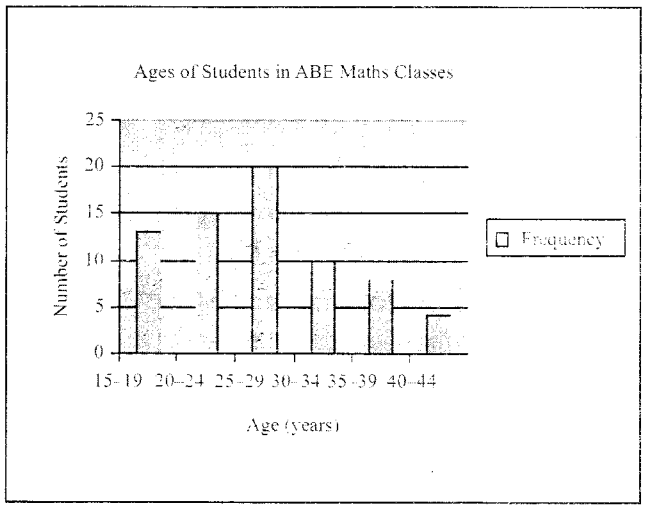

Fig. 5.1 Frequency Histogram

5.4.2 Role of Graphics in Presenting Data

A graph is a diagram in which different valves of the variables are plotted on a graph paper. A graph causes a movement or a change in the variable over a period of time. Data can be presented attractively with the help of diagrams. Diagrams are frequently used for publicity. However, they are not of much significance in statistical analysis. This enhances the efficiency of graphical presentation, which is more result oriented. Graphic presentation renders the data more effective and gives it more meaning.

NOTES

According to A. L. Boddington, 'The wandering of a line is more powerful in its effect on the mind than a tabulated statement; it shows what is happening and what is likely to take place, just as quickly as the eye is capable of working.'

Advantages of graphs

When statistics are displayed in the form of graphs, they facilitate many processes in economics. The important uses of graphs are as follows:

- Data is presented effectively and attractively: Graphs enhance the attractiveness and efficiency of presenting data. Graphs present data in a form that is easy for even an ordinary man to understand, in a better way. A graph may be conveniently described as a picture that is worth a thousand words.
- Data is made simple and easy to understand: Graphs present complex data in an easy-to-understand way. Consequently, graphs eradicate the complexity involved in statistics.
- Important for comparing data: Graphs are useful for comparing statistical data. To portray the investment in multiple ventures, graphs can be used. Through graphs, it is very easy to calibrate the difference between different investments.
- Useful for interpreting: Graphs are also capable for interpreting an outcome. They reduce the amount of labour and the time involved.
- Easy to mentally retain facts: Graphs make it easy to retain the facts mentally for a long time.
- Support forecasting: Graphs make it possible to forecast the tendencies that are likely to occur in the near future.
- Universal utility: In recent times, graphs find their use in every area of work, such as, trade, economics, government departments, advertisement, etc.
- Capacity to inform and entertain: Graphs provide information as well as entertainment. No obstacle can prevent the in-depth analysis of information through the use of graphs.
- Help to transmit information: Graphs are useful in the process of transmission as well as information of facts.
- Training not required: One does not need any special training to interpret the facts that are presented through graphs.

Limitations

Following are the main drawbacks/ limitations of graphs.

(i) **Limited application:** The use of graphical representation is higher for a common man than that for an expert. For experts, the use of graphical representation is restricted.

NOTES

(ii) **Lack of accuracy:** Graphs are not able to measure the magnitude of data accurately. They are only able to exhibit the fluctuations in them.

(iii) **Subjectivity:** Graphs display subjectivity of character. Different persons interpret them differently.

(iv) **Misleading conclusions:** A layman who lacks knowledge or is unable to understand graphs well, is likely to derive incorrect results from graphs.

5.4.3 Data Interpretation

The process of applying statistical methods for the analysis of facts that are particularly observed or assumed, from a specific study, is referred to as data interpretation. This is a very popular activity that occurs within the educational sector. They are project in the form of questions in tests that help to measure the understanding of the student, of a subject in question. Data interpretation is commonly used in schools, colleges, universities and other domains of higher education. Data interpretation serves as a means to qualify students on the basis of their understanding of the subject, in various entrance exams in educational institutes.

It is very essential for data to be interpreted correctly to excel in these tests. For students who plan to take up finance and mathematics as their subjects of study, it holds higher importance. Generally, an interpretation question is made up of a chart or a graph. These questions also comprise of data or sets of data which are to be analysed by students for achieving results. At the time of solving an interpretation question, it is very important to interpret the graph or chart correctly. In case numbers are part of the data, one needs to find out their meaning. Following this step, it is required to draw a data set that displays the concerned graph or chart. The data set may be analysed and the outcome should be based on the analysis. Regular practice is the key to be successful in interpretation questions.

5.5 HOW TO WRITE A RESEARCH PROPOSAL

A research proposal presents a concept that one wishes to pursue. An ideal research proposal assumes that one has already planned a project and efforts and time have gone into its planning, information gathering, reading and organization of ideas. Thus, a research proposal can be considered to be a second step that follows the activity of selecting a broad subject.

The actual proposal indicates the analytical questions that one needs to address. A good research question is the basis of a good research paper. This can be compared to a tree whose growth and strength depends on the quality of the root system.

An analytical question can have various forms and it is not necessary for one form to be better than others. All analytical questions have one common factor. This factor is their lack of description. That is, an analytical question moves beyond

NOTES

the 'what' and explores the 'how,' and the 'why.' A good analytical research paper will use 'the what' as part of the answer to 'the why.' But it is obligatory to address a question beyond the 'what.' For example, a descriptive paper would ask, 'What was Gorbachev's policies of economic perestroika?' An analytical paper would ask, 'Why did Gorbachev's economic perestroika policies fail?' As part of the paper, one would review the policies but this review would call for probing beyond these policies.

A research proposal comprises of the following specific sections:

(i) The question to be answered. The fact that the paper is aimed at explaining something should be kept in mind.

(ii) A short review of the literature which should comprise of the major works on the topic and define the subject of the argument.

If a theory is being applied to an issue, it can be reviewed before its use. The purpose is to highlight an awareness that pertains to the issue. This comprises of the evidence which was used, the theories that were applied and the arguments that were made.

(i) A statement pertaining to the text of the argument/explanation.

(ii) A statement that determines the difference between explanation/ argument from those put across by other authors. It determines whether the explanation is similar or different.

If a theory is being applied, one should be able to explain the theory and the reason for its selection and also the events that the specific theory intends to discuss.

General information on paper

A good research question is the basis of a good research paper. The questions should be well-defined. If there are no questions, it is vital to define them before the research begins. The paper should focus on analysis and not on being descriptive. Therefore, the organization of the paper should focus on a question. Papers that do not address a research question are liable to be returned without being graded.

A research paper should be a semester-long project since it will be graded in this manner. In other words, it is relevant to consider the amount of effort expended at the time of evaluating the quality of the paper.

Some general guidelines and instructions that may be helpful are as follows:

(i) The process of writing a research paper has many steps. The most common mistake made by researchers is the lack of time that is allowed. This system has a minimum of four distinct phases, these are, research (gathering materials), reading the materials, synthesis of the materials and writing the paper. One should start early, most people underestimate how long the first three steps take. These three steps also directly influence the quality of the paper: they are its foundation. To write a paper of god quality, one needs a strong foundation.

(ii) In terms of the number of citations, one should shoot for around 10-15 different sources, in some combination of books and academic journals. It is necessary to look at 2-3 sources for every one that is actually used, so in the course of researching your question you will look at about 30-45 separate items. Required readings for the class do NOT 'count' towards your total; however, non-assigned readings from a required text may be used. The point is to attain maximum exposure by using outside sources.

(iii) There should be four distinct parts of the paper. The first part, the introduction, should spell out clearly for the reader what the paper is about and what you will do. This part should indicate the question what are you exploring, the importance of the topic, what the parts of the paper will be, and a summary of hypothesis or findings.

The second part of research paper is the literature review where the major works on decided topic are reviewed and they indicate what the arguments are.

The third part is the body, or researches which have been have conducted. And the fourth part is the conclusion, summarizing what are finding and what answer is to the question need to have been posed. For organizational purposes, in the body of the paper one should use sections to divide the paper.

Sections to organize paper

(i) One has three options for notes to the paper: notes in the text, notes at the bottom of the page and notes at the end of the paper. However, in all the cases, the format should be right.

(ii) The paper should have a title page, indicating the title of the paper, the course and the name of the researcher.

(iii) A bibliography in the correct format should be part of the paper.

(iv) The paper should be typed with double spaces and regular margins. It should be spell-checked and tested for correct grammar.

(v) Numbering of pages is a must, with the second page following the title page. This means that page number 1 is the title page, but it is not numbered. This makes page 2 as the first page of the text.

(vi) The font size of the writing should be normal (12 points) and of regular print. It should not be in bold format.

5.6 WRITING RESEARCH REPORT-COMPONENT AND STYLE

The outcome of the market research must be conveyed to the management in an effective way. The outcome of the market research, when presented to the management, usually comprises of a report that is formally written as well as orally presented.

NOTES

Both, the written and the oral presentation are very significant. The primary reason for this is that in general, the outcome of marketing research is not very easy to define (after completion of the study and a final decision, there is very little physical evidence that indicates the amount of resources, such as time and effort that went into the project). The written report is actually a documentation of the project.

On the other hand, the documented report and the oral presentation are specifically the only features of the study to which marketing executives have access. Thus, as a result, the whole evaluation of the research project completely depends on the manner in which the communication of information takes place.

Another factor is that the supplier of the market research is completely responsible for both, the documented research report and the oral presentation. The efficiency and utility of the information that is communicated, has a critical role to play in deciding to repeat the use of that particular supplier in the future.

There is a difference in the writing style of every individual. There is not really one right style for a report. However, there are some basic rules which are common for writing any research report clearly. The making of a research report comprises of many other activities in addition to writing; to be more specific, writing is actually the final step of the preparation process. It is important to understand the outcome of the research project before one begins the process of writing. An in-depth thought should be given to the subject matter that the report is to address. Consequently, there are three steps that are involved in the preparation of a research report: understanding, organizing and writing. Following are the general guidelines that need to be adhered to, for any report or research paper:

- **Consider the audience:** The information that finally results from the study is very important for the managers/researchers, as this information is used by them for decision-making. Hence, they need to understand the report; the report should not be too technically inclined and use of much jargon should be avoided. This is a difficult task, most of the time. However, a researcher should ideally have a strong command on the statistical methods used by him.

Qualitative research also has its own problems. The terms used by behavioral sciences are different. One does not come across most of them in daily conversations. Instances of this are: cognitive dissonance, evoked set, perception, needs versus wants, self-actualization, etc. Noticeably, these examples are extreme cases; most of the words, phrases and terminologies that are used a very precise way by behavioral scientists are also present in everyday speech. However, in regular speech, these are used with lesser precision or difference. This also gives rise to opportunities for misrepresentations.

- **Concise, but precise:** In a way, a written report should be self-sufficient such that it should be able to stand by itself and should not depend on any additional clarification. Secondly, the report should be brief and to

NOTES

the point with specific focus on the critical elements of the project. It must not include issues that are not important. Researchers who are not experienced are highly tempted to find opportunities to convey all that they have done, to collect information for completing the research. This is an attempt to make sure that the audience realize and appreciate the time, effort and intellectual hindrances, encountered by the researchers to complete the research. It is more important for the researcher to realize that his/her level of contribution is measured by the ability to solve the marketing problem and not by the amount of efforts involved in the research methodology.

- **Understand the results and drawing conclusions:** Those who go through the report expect to see conclusions that can be interpreted in it. Hence, the researcher should be able to understand the results and interpret them. Besides reiterating the facts, the researcher should also be able to list the implications involved. For instance, if the researcher is comparing a client's product with that of a competitor and reports that 60 percent of respondents preferred brand A to brand B, then this is something that describes the results and does not interpret them. This kind of a comparison does not answer the 'So what?' question.

The outline suggested to write the research report, should include Title Page, Summary of findings, Table of contents, List of tables and List of figures etc. Introduction should include background to the research problem followed by Objectives and then hypotheses. The second step should comprise of methodology-process of data collection, sample and sampling method to be used and description of samples.

Findings include results, interpretation and conclusions. The most important part of the report is its summary of results. This is because the majority of the people from the management and others who are recipients of this report will only focus on its summary. For this reason, the summary should be present immediately after the title page, or is should be bound separately and presented along with the report.

The introduction should comprise of a description of the aim of the research along with an elaborated account of the research problem. This should be followed by a specification of the broad aim of the research. Following that, automatically the broad aim of the research can be specified and translated into a number of specific objectives. Furthermore, this section should also contain the hypotheses that are to be tested in the research.

After describing the samples, the key section of the findings should be presented in a way to target the achievement of all objectives of the study and testing of the hypotheses. As already discussed earlier, it is essential to interpret the main findings well and draw inferences, wherever possible.

NOTES

CHECK YOUR PROGRESS

4. What is meant by a graph?
5. What is the advantage of using graphs in interpretation?
6. How is data interpretation defined?

5.7 SUMMARY

- Data is another word for information. It generally means large amounts of information.
- Data processing is the function of identifying, acquiring, checking and reforming and formatting data into a presentable and understandable form.
- Data analysis is the identification, collection and storing of raw data in an organized way so that useful information can be extracted from it.
- Presentation of data is the basis of many experiments and assignments which depend on the ability to create and format data.
- Data interpretation is defined as use of statistical tools to analyze data that is found from a research.
- A research proposal is an attempt to convince others about the ability and capacity of a research project that is being carried out.
- A research report is based on and written after a research is successfully conducted and completed.

5.8 KEY TERMS

- **Data processing:** The process of converting raw data into meaningful information
- **Data aggregation:** The process of collecting and summarizing information for statistical analysis
- **Data validation:** The process of evaluating the correctness and accuracy of data
- **Data tabulation**: The activity of arranging data in a matrix form or in the form of tables
- **Data analysis:** The process of inspecting, cleaning, transforming and modeling data with the goal of highlighting useful information
- **Ogive:** A distribution curve in which the frequencies are cumulative

- **Box plot:** A box format which graphically depicts groups of numerical data through their five-number summaries
- **Histogram:** A bar chart representing a frequency distribution
- **Graphics:** Drawings which are in accordance to the rule of mathematics
- **Research proposal:** A written presentation of a proposed research

NOTES

5.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Data processing is the activity that involves a number of structured procedures that make information meaningful.
2. Data are raw facts which can be processed into useful information.
3. Data validation is the process of ensuring that a program operates on clean, correct and useful data.
4. A graph is a diagram in which different valves of the variables are plotted on a graph paper.
5. In the process of interpretation, graphs help to interpret the conclusion and save time as well as labour.
6. Data interpretation can be defined as 'the application of statistical procedures to analyze specific, observed or assumed, facts from a particular study'.

5.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Why is data validation important?
2. What is a histogram?
3. Why is data tabulated?
4. What is the main activity of preparing a research report?
5. Why are research proposals required?

Long-Answer Questions

1. Explain the process of data processing in detail.
2. What are the ways of presenting and interpreting data?
3. Discuss the significance of writing a research proposal.
4. Write a note on the importance of a research report.

5.11 FURTHER READING

NOTES

Adèr, H.J., Gideon Mellenbergh and David J. Hand. *Advising on Research Methods: A Consultant's Companion.* 2008. pp. 333–356.

Dunn, O. J. and V. A. Clark. 2001. *Basic Statistics.* New York: John Wiley and Sons.

Gingras, François-Pierre. 'Writing and Style Guide for University Papers and Assignments', available on www.socialsciences.uottawa.ca/guide-en.asp, access date 06.08.2003.